Activity Report 2011

Section New Results

Edition: 2012-03-23
<table>
<thead>
<tr>
<th>Team Name</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>AOSTE Project-Team</td>
<td>4</td>
</tr>
<tr>
<td>GALAAD Project-Team</td>
<td>9</td>
</tr>
<tr>
<td>GEOMETRICA Project-Team</td>
<td>18</td>
</tr>
<tr>
<td>MARELLE Project-Team</td>
<td>27</td>
</tr>
<tr>
<td>APICS Project-Team</td>
<td>30</td>
</tr>
<tr>
<td>NACHOS Project-Team</td>
<td>39</td>
</tr>
<tr>
<td>OPALE Project-Team</td>
<td>44</td>
</tr>
<tr>
<td>PUMAS Team</td>
<td>53</td>
</tr>
<tr>
<td>SMASH Project-Team</td>
<td>58</td>
</tr>
<tr>
<td>TOSCA Project-Team</td>
<td>61</td>
</tr>
<tr>
<td>TROPICS Project-Team</td>
<td>65</td>
</tr>
<tr>
<td>ABS Project-Team</td>
<td>68</td>
</tr>
<tr>
<td>ASCLEPIOS Project-Team</td>
<td>71</td>
</tr>
<tr>
<td>ATHENA Project-Team</td>
<td>86</td>
</tr>
<tr>
<td>BIOCORE Project-Team</td>
<td>98</td>
</tr>
<tr>
<td>DEMAR Project-Team</td>
<td>106</td>
</tr>
<tr>
<td>MODEMIC Team</td>
<td>123</td>
</tr>
<tr>
<td>NEUROMATHCOMP Project-Team</td>
<td>131</td>
</tr>
<tr>
<td>VIRTUAL PLANTS Project-Team</td>
<td>138</td>
</tr>
<tr>
<td>FOCUS Project-Team</td>
<td>150</td>
</tr>
<tr>
<td>INDES Project-Team</td>
<td>155</td>
</tr>
<tr>
<td>MAESTRO Project-Team</td>
<td>160</td>
</tr>
<tr>
<td>MASCOTTE Project-Team</td>
<td>171</td>
</tr>
<tr>
<td>OASIS Project-Team</td>
<td>184</td>
</tr>
<tr>
<td>PLANETE Project-Team</td>
<td>188</td>
</tr>
<tr>
<td>ARIANA Project-Team</td>
<td>203</td>
</tr>
<tr>
<td>AROBAS Project-Team</td>
<td>220</td>
</tr>
<tr>
<td>AXIS Project-Team</td>
<td>232</td>
</tr>
<tr>
<td>COPRIN Project-Team</td>
<td>240</td>
</tr>
<tr>
<td>EDELWEISS Project-Team</td>
<td>246</td>
</tr>
<tr>
<td>GRAPHIK Project-Team</td>
<td>256</td>
</tr>
<tr>
<td>PULSAR Project-Team</td>
<td>260</td>
</tr>
<tr>
<td>REVES Project-Team</td>
<td>287</td>
</tr>
<tr>
<td>ZENITH Team</td>
<td>300</td>
</tr>
</tbody>
</table>
6. New Results

6.1. Logical time in Model-Based embedded design

Participants: Charles André, Frédéric Mallet, Julien Deantoni, Robert de Simone, Marie-Agnès Peraldi Frati, Régis Gascon, Calin Gliția, Kelly Garces Pernett, Benoît Ferrero, Nicolas Chleq, Arda Göknil.

The foundational basis of our approach to modeling and analysis of embedded systems using logical time and logical clock specification constraints (CCSL) is recalled in 3.2, and was surveyed in [2]. This year we conducted a number of works exploiting this approach and promoting its introduction to various application domains.

Charles André presented the general approach in an invited lecture at the French Summer School on Real-Time, in Brest [21].

The HDR manuscript of Frédéric Mallet, where the MARTE Time Model is deeply considered, also in relation with other standards such as AADL, was published in book format [39].

In the article [19] we showed how CCSL observers could be encoded in the synchronous language Esterel, using crucial features of simultaneity, and how otherwise simultaneity could be obtained in simulation. This work was also presented internally as deliverable of the FUI Lambda project (see 8.2.3).

We drew a definite link with our activities on Process Network analysis (see 6.3), by showing how the CCSL primitives could be used to provide the loose timing semantic constraints of exiting PN models such as SDF (Synchronous Data-Flow domain of UC Berkeley’s Ptolemy), and its Multi-Dimensional extension (MD-SDF). This resulted in a journal publication [38]. Existing static schedules can then be obtained by analysis with K-Passa 5.2, or simulated using TimeSquare 5.1 (with an ASAP strategy).

In a collaboration with researchers at East China Normal University (ECNU Shanghai), we showed how CCSL constraints could be translated towards the PROMELA language implemented in the SPIN model-checker, which once again raises the issues of faithfully modeling simultaneity. This work resulted in a communication at the ICECCS conference [33]. Following this work one of our co-author, Yin Ling, earned a one-year scholarship from the Chinese government to visit us as part of her PhD.

The usage of CCSL expressions in the role of predicate property formulas, and their comparison with the more classical temporal formalisms such as PSL (Property Specification Language), was investigated in [24]. A longer internal report version can be found at [42].

In [23] we tackle the issue of recovering global information from multiple execution traces living in distinct logical time bases, with polychronous constraints relating them. The use for efficient debug of embedded systems from distributed traces is exemplified on a case study of terrestrial robot. This work was conducted in the framework of the ANR RT-Simex project, see 8.2.1.

A case study in modeling with logical time and CCSL, from requirements to implementation, based on an automotive spark ignition system, is provided in [31]. We worked more generally on the introduction of our approach to existing formalism in the automotive domain, such as EAST-ADL2 and AutoSar, as part of our contribution to the new ITEA2 Timmo2U project. Premises of this effort are described in [32].

The use of CCSL constraints in general requirement engineering was also studied and demonstrated in a conference article, jointly with colleagues at ECNU Shanghai, presented at APSEC’2011 [22].

The use and modeling of priorities amongst timed events (i.e., logical clock ticks), which has strong impacts on efficient logical clock based simulations and scheduling (as the choice of next event), is still a topic of ongoing work. Several advanced considerations are to be found as part of Jean-François Le Tallec PhD thesis, to be defended in January 2012 [16].
6.2. Model-Based approaches to SoC design

Participants: Charles André, Robert de Simone, Benoît Ferrero, Carlos Gomez Cardenas, Jean-François Le Tallec.

The main practical goal of this work was to combine in a sensible way the various formalisms SystemC, IP-XACT, UML MARTE, and UPF (for power consumption representation) (see 4.1 for further descriptions). There were true motivations for this: SystemC is a de-facto standard for SoC simulation at various levels, but currently lacks any sort of formal description so that systems can be analysed, reasoned about for correctness and optimized (and it becomes even more so with newer draft standard evolutions). IP-XACT was introduced as an ADL to ease composition and assembly of IP components (written in SystemC or not), but again it currently fails short of its goal, and in particular does not allow standard decoration of model attributes in prominent non-functional domains such as timing/performance and low-power/energy consumption. These could be provided with the help of dedicated features in UML MARTE, and aligned on the UPF standard for power management modeling.

While the intended design flow would take the UML MARTE and UPF to IP-XACT to SystemC direction, it was important to extract IP-XACT and MARTE structural representation from existing SystemC programs, both to populate the flow with existing legacy models, and to explore better the requirements for complete and consistent modeling towards IP block assembly. This work was conducted in Jean-François Le Tallec PhD, to be defended in January 2012 [16]. Together with Benoît Ferrero he defined and realized a software tool named SCiPX (SystemC to IP-XACT translator), originally based on the PinaVM tool by VERIMAG and the DoxyGen syntactic analyzer.

SCiPX is available in prototype version from our site http://www-sop.inria.fr/aoste/index.php?page=software/scipx. It can be combined with the former software transformation modules IPXACT2Marte and Marte2IPXACT developed previously. These results were partly supported by the ID/TL-M contract with ST Microelectronics (see 7.1), and the ANR HeLP project (see 8.2.2), and were presented in [26], [25].

As part of his PhD thesis, Carlos Gomez Cardenas described a subset of UPF standard as a metamodel inside UML MARTE. He also considered compatibility and interconnections with the industrial environments AcePlorer (by Docea Power), and Synopsys Virtualizer (formerly CoWare), provided to us in the context of the CIM PACA tool farm 8.1.1. Preliminary results were presented in [36]. This work was also presented during internal meetings of the ANR HeLP project, and coordinated with work conducted in the team of Michel Auguin at CNRS UMR LEAT (also in Sophia-Antipolis).

6.3. Process Network analysis


This year we comforted the type of analysis on regular static scheduling and routing in dedicated process network models such as studied in the successive PhD thesis of Julien Boucaron, Jean-Vivien Millo, and Anthony Coadou, and recently surveyed in [5]. This resulted mainly in further implementation upgrades of our K-Passa tool (see 5.2), performed first by Anthony Coadou (before he left on a postdoc internship), then continued by Jean-Vivien Millo (on a return postdoc position with us).

In a work mostly conducted while member of the Alchemy EPI in Saclay, but which draws a clear link to our past and present activities in the subject, Sid-Ahmed-Ali Touati studied efficient heuristics to the general problem of one-dimensional periodic task scheduling under storage requirements, using a modeling framework akin to Process Networks. This resulted in a journal article accepted once the author had become attached to the Aoste EPI [17].

6.4. Correct and efficient implementation of polychronous formalisms

Participants: Thomas Carle, Manel Djemal, Virginia Papailiopoulos, Dumitru Potop Butucaru, Robert de Simone, Yves Sorel.
Existing analysis techniques for synchronous and polychronous languages, such as clock calculi, are meant to extract relations of simultaneity (time inclusion) and exclusiveness (time exclusion) between the various computations and communications. This approach is well-suited when targeting sequential processors. For distributed or multi-threaded implementations, further independence relations are needed to express potential concurrency. This resulted in a general theory of endochronous systems, meant to support this additional analysis [11].

Last year we completed a first prototype tool implementation for weak endochrony checking. This was completed this year in two directions:

- connecting our tool with Signal as input language, and interface it in practice to the Polychrony/SME environment developed by the Espresso EPI;
- Improving algorithmic complexity and internal data representaion, so that our tool can now handle reasonable size Signal programs.

This work was of course conducted in collaboration with Espresso members. Experimental results were presented at the ESLsyn 2011 conference [30]. We are currently expanding the framework in order to take modes/states into account in the program specifications. Effective generation of multi-threaded GALS wrappers for Signal programs is also under way.

We worked at extending the AAA methodology for polychronous processes by providing a better integration of clock analysis in the various phases of the implementation process (allocation, scheduling, pipelining, etc.). We also considered a wider range of implementation targets (time-triggered, MPSoC). We defined a dedicated software pipelining algorithm to match conditional scheduling/reservation tables such as used in SynDEx, with the goal of improving throughput with the same duration of individual computation cycles (as is the goal of any pipelining techniques). The originality here is to make logical clocks of polychronous systems act as triggers for the predicated executions as used in traditional software pipelining. First results have been presented during the Synchron 2011 workshop and in a research report [41].

Further work on time-triggered systems was submitted inside the FUI Parsec 8.2.4 and P 8.2.5 projects, including real-time implementation methods for the IMA/ARINC 653 avionics platforms. In particular we conducted experiments to replace the scheduling policy of the second-level scheduler (L1 in the standard) from dynamic priority-driven to dynamic Time Division time triggered (TT-IMA). Preliminary results are under way, and were informally presented at the yearly Synchron seminar.

An important emerging trend in target MPSoC platforms is that on-chip networks are progressively introduced to cope with the bottleneck of inter processor communications. Correct implementation of polychronous systems in this context thus relies on efficient routing of data in such networks, and ultimately may assume that on-chip NoC routers may be programmed in one way or another to behave predictably according to the global application distributed on the cores. We started a collaboration on this topic with the "Embedded Systems-on-Chips" department of the LIP6 laboratory, one of the main site of expertise for SoC/NoC design and Hardware/software codesign. This collaboration first materialized with the co-supervision of M. Djemal’s PhD thesis. A generic MPSoC architecture is being defined, which includes a 2D mesh network-on-chip with programmable routers, on which static routing schedules such as synthesized by our tools may be implemented and run.

### 6.5. Uniprocessor Real-time Scheduling

**Participants:** Laurent George, Mohamed Marouf, Daniel de Rauglaudre, Yves Sorel.

#### 6.5.1. Strict periodic harmonic tasks

This year, we focused our work on scheduling of strict periodic tasks to the particular case of harmonic tasks [28]. After transforming the scheduling problem into a bin-packing problem, we performed a schedulability analysis and proposed schedulability conditions in each sub-case of harmonic tasks: we proposed a necessary and sufficient condition in the case where all tasks periods are distinct, and we proposed a sufficient condition in the case where some tasks have the same period. Finally, we proposed a scheduling algorithm based on the bin-packing problem resolution.
6.5.2. Combination of strict periodic and sporadic tasks

Non-preemptive strict periodic tasks are harder to schedule than preemptive ones. One can hope to extend schedulability results when combining non-preemptive strict periodic tasks with preemptive sporadic one.

We proposed in [27] a schedulability analysis for a combination of strict periodic and sporadic tasks. We considered all tasks with fixed priorities, where the highest priorities are given to strict periodic tasks and the lower priorities are given to sporadic tasks. First, we scheduled strict periodic tasks using our former scheduling algorithm. Then, we computed the critical instants which maximize the response time of a sporadic task. We proved that the critical instants are contained in the permanent phase of strict periodic tasks, and are given by the start times of strict periodic jobs in a hyper-period. We also proved that we can reduce critical instants by eliminating some of them. Then, we gave the analytic expression of the computing time $W_i(t)$ at any time $t$ necessary for the execution of a task $\tau_i$ taking in consideration all the tasks with higher priorities. That allows the computation of the response time $r_i$ by solving the equation $W_i(t) = r_i$. Therefore, for a sporadic task, if its response time $r_i$ is less of equal to its deadline for all critical instants, then this latter task is schedulable, else it is not schedulable. We proceed similarly for all sporadic tasks to prove that a tasks set is schedulable or not.

6.5.3. Exact cost of RTOS

It is important to determine the exact cost of the real-time operating system (RTOS) when preemptive scheduling is used for better processor utilization compared to non preemptive scheduling [43]. Indeed, in this case it is possible to trust the schedulability conditions when they include that cost, and also to avoid waste resources. This year we developed a generic RTOS modelled with Petri nets and we determined its exact cost on an ARM9 processor. We used Petri nets on the one hand to choose through simulations the best structure of that scheduler, and on the other hand to verify non blocking properties. In order to obtain its exact cost the scheduler was programmed in assembly language, and coded such as it is deterministic, i.e. its cost does not depend on alternative statements but only on the number of tasks which is known a priori. Using this RTOS we experimented simple task sets on the ARM9 processor for which we were able to include the exact RTOS cost in the schedulability conditions.

6.5.4. Formal proofs of real-time scheduling theorems

Scheduling involves numerous models and theorems, sometimes dated of several decades, but never formally proved. We made a formal proof in Coq (proof assistant developed at Inria) to check a classical theorem giving a schedulability condition for a set of real-time strictly periodic tasks (about 1500 lines of Coq). This work was published in a paper accepted for publication in the conference JFLA 2012.

A second proof is actually being carried now, dealing with response time of a set of fixed priority real-time preemptive tasks. The theorem states that the worst case of this response time occurs when all tasks start simultaneously. A step in the original argument by Jane W. S. Liu [55] involves the proof of a function whose fixpoint computes the response time of the first instance of the least priority task. This specific step is now formally proved in Coq (3500 lines of Coq), and we are now working on the completion of the full theorem.

6.6. Multiprocessor Real-time Scheduling

Participants: Laurent George, Maxence Guesdon, Mohamed Marouf, Falou Ndoye, Simon Nivault, Yves Sorel, Cécile Stentzel.

6.6.1. Partitioned scheduling with exact RTOS cost

In the case of partitioned scheduling we propose a greedy heuristic to solve the real-time scheduling problem of periodic preemptive tasks on a multiprocessor architecture while taking into account the exact RTOS cost. This is achieved by combining an allocation heuristic, of “best fit” type, and a schedulability condition based on the operation $\oplus$ which takes into account the exact RTOS cost [43]. The allocation heuristic minimizes the makespan (total execution time of the tasks executed on the multiprocessor taking into account inter-processor communication costs). A first version of this work was presented in [29].
6.6.2. Semi-partitioned scheduling

In [18] we study two cases of semi-partitioned scheduling of sporadic tasks with constrained deadlines on homogeneous multi-processor: (i) the case where the Worst Case Execution Time (WCET) of a job can be portioned, each portion being executed on a dedicated processor, according to a static pattern of migration and using for solving the local assignment problem a linear programming approach; (ii) the case where the jobs of a task are released on a processor, 1 time out of p, where p is an integer less than or equal to the number of processors, according to a Round-Robin migration pattern. The first approach has been investigated in the state-of-the-art by migrating a job at its local deadline, computed from the deadline of the task it belongs to.

6.6.3. Fault tolerance on electric vehicles

We consider applications composed of a real-time task set running on the distributed heterogeneous architecture of the CyCab (electric vehicle developed in the IMARA team-project) based on dsPICs processors, MPC555 micro-controllers, and an embedded PC all together connected through CAN (Controller Area Network) buses. For hardware reasons we suppose that only dsPICs and CAN buses can fail. Our goal is to find a fault-tolerant software solution to tolerate such failures while the applications satisfy the real-time constraints. Because extra hardware for error detection is expensive in such electric vehicle, we proposed a software error detection based on watchdogs. We solved separately two different problems: buses and dsPICs fault-tolerance. In both cases we use active redundancy policies. For buses fault-tolerance, we assume that all processors are reliable, and all but one bus can fail. The same data is sent through all the CAN buses. If a CAN bus fails then the data is sent by the other CAN buses. For processors fault-tolerance, we assume that all communication media are reliable and at least one processor can fail. The first step consists in performing active redundancy for all the tasks of the application. A task and their redundant tasks are assigned to different processors. If processor fails then the data which are not sent by tasks running on that faulty processor, are actually sent by the redundant tasks. All the tasks with their redundant counterparts are scheduled according to the schedulability analysis proposed in [28].

6.6.4. Scicos/SynDEx gateway and code generation for multi-core

This work was carried out in the Openprod project (see 8.3.2.2). The gateway between Scicos and SynDEx has been updated to deal with the last Scicos data structures and the last version of SynDEx. Besides, this gateway has been improved and partially rewritten to support as much Scicos blocks as possible. We use the gateway to automatically produce from a control model specified and simulated in Scicos a real-time executable running on a multi-core platform. The latter platform is described according to the shared memory model defined last year. In order to generate real-time executable code we had to develop a new SynDEx executive kernel based on Windows-RTX which supports shared memory communications and multi-core parallel execution. That executive kernel is used with the macro-code generated by SynDEx to produce the real-time executable code.

6.6.5. SynDEx updates

We continued the developments of future version 8 of SynDEx which will feature a new software architecture to allow better functionality evolutions and maintenance. On the other hand in the COTROS ADT ("Génération de code temps réel distribué optimisé et sûr"), we achieved the new automatic code generator for the current version 7 of SynDEx. This generator creates intra and inter-processor synchronizations according to well defined rules, checks deadlock absence in inter-processor synchronizations, manages efficiently buffers and semaphores (declaration, naming, etc.).
6. New Results

6.1. Algebraic Algorithms for Geometric Computing

6.1.1. An Algebraic Approach to Continuous Collision Detection for Ellipsoids

**Participant:** Bernard Mourrain.

In [28], we present algebraic conditions for characterizing three configurations of two ellipsoids in $\mathbb{R}^3$ that are the most relevant to collision detection: separation, external touching and overlapping. These conditions are given in terms of explicit formulae expressed by the subresultant sequence of the characteristic polynomial of the two ellipsoids and its derivative. For any two ellipsoids, the signs of these formulae can easily be evaluated to classify their configuration. Furthermore, based on these algebraic conditions, an efficient method is developed for continuous collision detection for two moving ellipsoids under arbitrary motion.

This is a joint work with Xiaohong Jia, Yi-King Choi and Wenping Wang from the university of Hong Kong.

6.1.2. On Continued Fraction Expansion of Real Roots of Polynomial Systems, Complexity and Condition Numbers

**Participants:** Angelos Mantzaflaris, Bernard Mourrain.

In [29], we elaborate on a correspondence between the coefficients of a multivariate polynomial represented in the Bernstein basis and in a tensor-monomial basis, which leads to homography representations of polynomial functions, that use only integer arithmetic (in contrast to Bernstein basis) and are feasible over unbounded regions. Then, we study an algorithm to split this representation and we obtain a subdivision scheme for the domain of multivariate polynomial functions. This implies a new algorithm for real root isolation, MCF, that generalizes the Continued Fraction (CF) algorithm of univariate polynomials. A partial extension of Vincent’s Theorem for multivariate polynomials is presented, which allows us to prove the termination of the algorithm. Bounding functions, projection and preconditioning are employed to speed up the scheme. The resulting isolation boxes have optimized rational coordinates, corresponding to the first terms of the continued fraction expansion of the real roots. Finally, we present new complexity bounds for a simplified version of the algorithm in the bit complexity model, and also bounds in the real RAM model for a family of subdivision algorithms in terms of the real condition number of the system. Examples computed with our C++ implementation illustrate the practical aspects of our method.

This is a joint work with E. Tsigaridas, from the Department of Computer Science, University of Aarhus.

6.1.3. Matrix-based representations of rational hypersurfaces

**Participants:** Laurent Busé, Nicolas Botbol.

This ongoing work is related to matrix-based representations of rational hypersurfaces whose theoretical foundations has been recently developed by our team and several other authors in the context of the implicitization problem. Being given a parameterized curve or hypersurface, this method consists in building a matrix whose entries are typically linear forms in the variables of the ambient space and such that the ideal generated by the maximal minors of this matrix provide a good approximation of the original curve or surface.

We aim to study and determine the geometric informations that are contained in a representation matrix. In particular, we are currently adressing the two following questions:

1) understand the extraneous components that are added by taking the initial Fitting ideal of a representation matrix, with respect to the original curve or surface. Indeed, these extraneous components appear because of the good behavior of Fitting ideals under change of bases. Therefore, one can expect that these extraneous components yields some geometric properties of a curve or surface as a member of a certain family. 2) examine the extraction of singularities from a representation matrix, similarly to the recent results on what is called "singular factors" for the case of rational curves.
6.1.4. The surface/surface intersection problem by means of matrix based representations
Participants: Laurent Busé, Thang Luu Ba.
Evaluating the intersection of two rational parameterized algebraic surfaces is an important problem in solid modeling. In this work, we made use of some generalized matrix based representations of parameterized surfaces in order to represent the intersection curve of two such surfaces as the zero set of a matrix determinant. As a consequence, we extended to a dramatically larger class of rational parameterized surfaces the applicability of a general approach to the surface/surface intersection problem due to J. Canny and D. Manocha. In this way, we obtained compact and efficient representations of intersection curves allowing to reduce some geometric operations on such curves to matrix operations using results from linear algebra.
See the preprint version at http://hal.inria.fr/inria-00620947/en/

6.2. Symbolic-Numeric Analysis

6.2.1. A Subdivision Method for Computing Nearest Gcd with Certification
Participants: André Galligo, Bernard Mourrain.
A new subdivision method for computing the nearest univariate gcd is described and analyzed in [24]. It is based on an exclusion test and an inclusion test. The exclusion test in a cell exploits Taylor expansion of the polynomial at the center of the cell. The inclusion test uses Smale's alpha-theorems to certify the existence and unicity of a solution in a cell. Under the condition of simple roots for the distance minimization problem, we analyze the complexity of the algorithm in terms of a condition number, which is the inverse of the distance to the set of degenerate systems. We report on some experimentation on representative examples to illustrate the behavior of the algorithm.
This is a joint work with Guillaume Chèze and Jean-Claude Yakoubsohn (University Paul-Sabatier, Toulouse).

6.2.2. An Adapted Version of the Bentley-Ottmann Algorithm for Invariants of Plane Curves Singularities
Participant: Bernard Mourrain.
In [34], we report on an adapted version of the Bentley-Ottmann algorithm for computing all the intersection points among the edges of the projection of a three-dimensional graph. This graph is given as a set of vertices together with their space Euclidean coordinates, and a set of edges connecting them. More precisely, the three-dimensional graph represents the approximation of a closed and smooth implicitly defined space algebraic curve, that allows us a simplified treatment of the events encountered in the Bentley-Ottmann algorithm. As applications, we use the adapted algorithm to compute invariants for each singularity of a plane complex algebraic curve, i.e. the Alexander polynomial, the Milnor number, the delta-invariant, etc.
This is a joint work with Madalina Hodorog and Joseph Schicho, from RICAM, Linz, Austria.

6.2.3. Virtual Roots of a Real Polynomial and Fractional Derivatives
Participant: André Galligo.
After the works of Gonzales-Vega, Lombardi, Mahé, and Coste, Lajous, Lombardi, Roy, we consider the virtual roots of a univariate polynomial $f$ with real coefficients. Using fractional derivatives, we associate to $f$ a bivariate polynomial $P_f(x,t)$ depending on the choice of an origin $a$, then two type of plan curves we call the FDcurve and stem of $f$. We show, in the generic case, how to locate the virtual roots of $f$ on the Budan table and on each of these curves. The paper [32] is illustrated with examples and pictures computed with the computer algebra system Maple. It is a joint work with Daniel Bembe.

6.2.4. Computing monodromy via continuation methods on random Riemann surfaces
Participant: André Galligo.
In [25], we consider a Riemann surface $X$ defined by a polynomial $f(x, y)$ of degree $d$, whose coefficients are chosen randomly. Hence, we can suppose that $X$ is smooth, that the discriminant $\delta(x)$ of $f$ has $d(d-1)$ simple roots $\Delta$ and that $\delta(0) \neq 0$, i.e. the corresponding fiber has $d$ distinct points $\{y_1, \ldots, y_d\}$. When we lift a loop $0 \in \gamma \subset C - \Delta$ by a continuation method, we get $d$ paths in $X$ connecting $\{y_1, \ldots, y_d\}$, hence defining a permutation of that set. This is called monodromy.

Here we present experimentations in Maple to get statistics on the distribution of transpositions corresponding to loops around each point of $\Delta$. Multiplying families of “neighbor” transpositions, we construct permutations and the subgroups of the symmetric group they generate. This allows us to establish and study experimentally two conjectures on the distribution of these transpositions and on transitivity of the generated subgroups.

Assuming that these two conjectures are true, we develop tools allowing fast probabilistic algorithms for absolute multivariate polynomial factorization, under the hypothesis that the factors behave like random polynomials whose coefficients follow uniform distributions. It is a joint work with Adrien Poteaux (University of Lille).

6.3. Algebraic representations for geometric modeling

6.3.1. Multihomogeneous Polynomial Decomposition using Moment Matrices

Participants: Alessandra Bernardi, Jérôme Brachat, Bernard Mourrain.

In [33], we address the important problem of tensor decomposition which can be seen as a generalisation of Singular Value Decomposition for matrices. We consider general multilinear and multihomogeneous tensors. We show how to reduce the problem to a truncated moment matrix problem and we give a new criterion for flat extension of Quasi-Hankel matrices. We connect this criterion to the commutation characterization of border bases. A new algorithm is described: it applies for general multihomogeneous tensors, extending the approach of J.J. Sylvester on binary forms. An example illustrates the algebraic operations involved in this approach and how the decomposition can be recovered from eigenvector computation.

This is a joint work with Pierre Comon (I3S, CNRS).

6.3.2. On the variety parametrizing completely decomposable polynomials.

Participant: Alessandra Bernardi.

The purpose of the paper [15] is to relate the variety parameterizing completely decomposable homogeneous polynomials of degree $d$ in $n + 1$ variables on an algebraically closed field, called $\text{Split}_d(\mathbb{P}^n)$, with the Grassmannian of $n - 1$ dimensional projective subspaces of $\mathbb{P}^{n+d-1}$. We compute the dimension of some secant varieties to $\text{Split}_d(\mathbb{P}^n)$ and find a counterexample to a conjecture that wanted its dimension related to the one of the secant variety to $G(n-1, n + d - 1)$. Moreover by using an invariant embedding of the Veronese variety into the Plücker space, then we are able to compute the intersection of $G(n-1, n + d - 1)$ with $\text{Split}_d(\mathbb{P}^n)$, some of its secant variety, the tangential variety and the second osculating space to the Veronese variety.

This is a joint work with Enrique Arrondo (Universidad Complutense de Madrid, Spain)

6.3.3. Computing symmetric rank for symmetric tensors.

Participant: Alessandra Bernardi.

In [21] we consider the problem of determining the symmetric tensor rank for symmetric tensors with an algebraic geometry approach. We give algorithms for computing the symmetric rank for $2 \times \ldots \times 2$ tensors and for tensors of small border rank. From a geometric point of view, we describe the symmetric rank strata for some secant varieties of Veronese varieties.

This is a joint work with Alessandro Gimigliano and Monica Idà (Università di Bologna, Italy).

6.3.4. Higher secant varieties of $\mathbb{P}^n \times \mathbb{P}^m$ embedded in bi-degree $(1, d)$.

Participant: Alessandra Bernardi.
Let $X^{(n,m)}$ denote the Segre-Veronese embedding of $\mathbb{P}^n \times \mathbb{P}^m$ via the sections of the sheaf $\mathcal{O}(1, d)$. In [20] we study the dimensions of higher secant varieties of $X^{(n,m)}$ and we prove that there is no defective $s^{th}$ secant variety, except possibly for $n$ values of $s$. Moreover when $\binom{m+d}{d}$ is multiple of $(m + n + 1)$, the $s^{th}$ secant variety of $X^{(n,m)}$ has the expected dimension for every $s$.

This is a joint work with Enrico Carlini (Politecnico di Torino, Italy, Maria Virginia Catalisano (Università di Genova, Italy).

### 6.3.5. On the $X$-rank with respect to linear projections of projective varieties.

**Participant:** Alessandra Bernardi.

In [17] we improve the known bound for the $X$-rank $R_X(P)$ of an element $P \in \mathbb{P}^N$ in the case in which $X \subset \mathbb{P}^n$ is a projective variety obtained as a linear projection from a general $v$-dimensional subspace $V \subset \mathbb{P}^{n+v}$. Then, if $X \subset \mathbb{P}^n$ is a curve obtained from a projection of a rational normal curve $C \subset \mathbb{P}^{n+1}$ from a point $O \subset \mathbb{P}^{n+1}$, we are able to describe the precise value of the $X$-rank for those points $P \in \mathbb{P}^n$ such that $R_X(P) \leq R_C(O) - 1$ and to improve the general result. Moreover we give a stratification, via the $X$-rank, of the osculating spaces to projective cuspidal projective curves $X$. Finally we give a description and a new bound of the $X$-rank of subspaces both in the general case and with respect to integral non-degenerate projective curves.

This is a joint work with Edoardo Ballico (Università di Trento, Italy).

### 6.3.6. Decomposition of homogeneous polynomials with low rank.

**Participant:** Alessandra Bernardi.

Let $F$ be a homogeneous polynomial of degree $d$ in $m + 1$ variables defined over an algebraically closed field of characteristic zero and suppose that $F$ belongs to the $s$-th secant varieties of the standard Veronese variety $X_{m,d} \subset \mathbb{P}(\mathbb{P}^{m+d})$ but that its minimal decomposition as a sum of $d$-th powers of linear forms $M_1, \ldots, M_r$ is $F = M_1^r + \cdots + M_r^r$ with $r > s$. In [16] we show that if $s + r \leq 2d + 1$ then such a decomposition of $F$ can be split in two parts: one of them is made by linear forms that can be written using only two variables, the other part is uniquely determined once one has fixed the first part. We also obtain a uniqueness theorem for the minimal decomposition of $F$ if the rank is at most $d$ and a mild condition is satisfied.

This is a joint work with Edoardo Ballico (Università di Trento, Italy).

### 6.3.7. On the $X$-rank with respect to linearly normal curves.

**Participant:** Alessandra Bernardi.

In [18] we study the $X$-rank of points with respect to smooth linearly normal curves $X \subset \mathbb{P}^n$ of genus $g$ and degree $n + g$.

We prove that, for such a curve $X$, under certain circumstances, the $X$-rank of a general point of $X$-border rank equal to $s$ is less or equal than $n + 1 - s$.

In the particular case of $g = 2$ we give a complete description of the $X$-rank if $n = 3, 4$; while if $n \geq 5$ we study the $X$-rank of points belonging to the tangential variety of $X$.

This is a joint work with Edoardo Ballico (Università di Trento, Italy).

### 6.3.8. Symmetric tensor rank with a tangent vector: a generic uniqueness theorem

**Participant:** Alessandra Bernardi.

Let $X_{m,d} \subset \mathbb{P}^N$, $N := \binom{m+d}{m} - 1$, be the order $d$ Veronese embedding of $\mathbb{P}^m$. Let $\tau(X_{m,d}) \subset \mathbb{P}^N$, be the tangent developable of $X_{m,d}$. For each integer $t \geq 2$ let $\tau(X_{m,d}, t) \subset \mathbb{P}^N$, be the join of $\tau(X_{m,d})$ and $t - 2$ copies of $X_{m,d}$. In [19] we prove that if $m \geq 2, d \geq 7$ and $t \leq 1 + \left\lfloor \frac{m+d-2}{(m+1)} \right\rfloor$, then for a general $P \in \tau(X_{m,d}, t)$ there are uniquely determined $P_1, \ldots, P_t \in X_{m,d}$ and a unique tangent vector $\nu$ of $X_{m,d}$ such that $P$ is in the linear span of $\nu \cup \{P_1, \ldots, P_t\}$, i.e. a degree $d$ linear form $f$ (a symmetric tensor $T$ of order $d$) associated to $P$ may be written as
\[ f = L_{t-1} L_t + \sum_{i=1}^{t-2} L_i, \quad (T = v_{t-1}^{d-1} v_t + \sum_{i=1}^{t-2} v_i^d) \]

with \( L_i \) linear forms on \( \mathbb{P}^m \) (\( v_i \) vectors over a vector field of dimension \( m + 1 \) respectively), \( 1 \leq i \leq t \), that are uniquely determined (up to a constant).

This is a joint work with Edoardo Ballico (Università di Trento, Italy).

6.3.9. Parametrization of computational domain in isogeometric analysis: methods and comparison
Participants: André Galligo, Bernard Mourrain.

Parameterization of computational domain plays an important role in isogeometric analysis as mesh generation in finite element analysis. In this paper, we investigate this problem in the 2D case, i.e., how to parametrize the computational domains by planar B-spline surface from the given CAD objects (four boundary planar B-spline curves). Firstly, two kinds of sufficient conditions for injective B-spline parameterization are derived with respect to the control points. Then we show how to find good parameterization of computational domain by solving a constraint optimization problem, in which the constraint condition is the injectivity sufficient conditions of planar B-spline parametrization, and the optimization term is the minimization of quadratic energy functions related to the first and second derivatives of planar B-spline parameterization. By using this method, the resulted parameterization has no self-intersections, and the isoparametric net has good uniformity and orthogonality. After introducing a posteriori error estimation for isogeometric analysis, we propose \( r \)-refinement method to optimize the parameterization by repositioning the inner control points such that the estimated error is minimized. Several examples are tested on isogeometric heat conduction problem to show the effectiveness of the proposed methods and the impact of the parameterization on the quality of the approximation solution. Comparison examples with known exact solutions are also presented. This joint work with Régis Duvigneau (EPI OPALE) and Gang Xu (Hangzhou Dianzi University, China) is published in [31].

6.3.10. Variational Harmonic Method for Parameterization of Computational Domain in 2D Isogeometric Analysis
Participants: André Galligo, Bernard Mourrain.

In isogeometric analysis, parameterization of computational domain has great effects as mesh generation in finite element analysis. In this paper, based on the concept of harmonic map from the computational domain to parametric domain, a variational approach is proposed to construct the parameterization of computational domain for 2D isogeometric analysis. Different from the previous elliptic mesh generation method in finite element analysis, the proposed method focuses on isogeometric version, and converts the elliptic PDE into a nonlinear optimization problem. A regular term is integrated into the optimization formulation to achieve more uniform grid near convex (concave) parts of the boundary. Several examples are presented to show the efficiency of the proposed method.

This joint work with Régis Duvigneau (EPI OPALE) and Gang Xu (Hangzhou Dianzi University, China) is published in [36].

6.3.11. Warp-based Helical Implicit Primitives
Participant: Evelyne Hubert.

Implicit modeling with skeleton-based primitives has been limited up to now to planar skeletons elements, since no closed-form solution was found for convolution along more complex curves. We show that warping techniques can be adapted to efficiently generate convolution-like implicit primitives of varying radius along helices, a useful 3D skeleton found in a number of natural shapes. Depending on a single parameter of the helix, we warp it onto an arc of circle or onto a line segment. For those latter skeletons closed form convolutions are known for entire families of kernels. The new warps introduced preserve the circular shape of the normal cross section to the primitive.
This is joint work with Cédric Zanni and Marie-Paule Cani from the project-team EVASION (INRIA Grenoble Rhône-Alpes / LJK Laboratoire Jean Kuntzmann) which is published in [37].

6.4. National Initiatives

6.4.1. PlantScan3D

PlantScan3D is an ARC between coordinated by the EPI Virtual Plants (UMR DAP, INRIA-CIRAD, Montpellier), with the EPI Galaad (INRIA, Méditerranée) and Evasion (INRIA Rhône-Alpes, Grenoble).

A close collaboration between specialists in plant structures modelling, algebraic geometry, and 3D computer graphic is required to address plant structure reconstruction from laser scanned point clouds. Indeed it is required to take into account efficiently knowledge from topology and geometry to allow mapping and reconstruction of data despite noise, occlusions, and thinness of structure. The objective of the project is to provide as output a compact geometrical model that model smoothly branching point of tubular structure and organs (like leaves). At the end, this model should make it possible an interactive visualisation and automatize different measurement operators needed by biological partners.


6.4.2. GEOLMI

GEOLMI - Geometry and Algebra of Linear Matrix Inequalities with Systems Control Applications - is an ANR project working on topics related to the Geometry of determinantal varieties, positive polynomials, computational algebraic geometry, semidefinite programming and systems control applications.

The partners are LAAS-CNRS, Univ. de Toulouse (coordinator), LJK-CNRS, Univ. Joseph Fourier de Grenoble; INRIA Sophia Antipolis Méditerranée; LIP6-CNRS Univ. Pierre et Marie Curie; Univ. de Pau et des Pays de l’Adour; IRMAR-CNRS, Univ. de Rennes.


6.4.3. ANEMOS

ANEMOS - Advanced Numeric for ELMs : Modeling and Optimized Schemes - is an ANR project devoted to the numerical modelling study of such ELM control methods as Resonant Magnetic Perturbations (RMPs) and pellet ELM pacing both foreseen in ITER. The goals of the project are to improve understanding of the related physics and propose possible new strategies to improve effectiveness of ELM control techniques. The study of spline spaces for isogeometric finite element methods is proposed in this context.

The partners are IRFM, CEA, Cadarache; JAD, University of Nice - Sophia Antipolis; INRIA, Bacchus; Maison de la Simulation CEA-CNRS-INRIA-University of Orsay- University of Versailles St Quentin.

6.5. European Initiatives

6.5.1. FP7 Projects

6.5.1.1. TERRIFIC

Title: Towards Enhanced Integration of Design and Production in the Factory of the Future through Isogeometric Technologies
Type: COOPERATION (ICT)
Defi: PPP FoF: Digital factories: Manufacturing design and product lifecycle manage
Instrument: Specific Targeted Research Project (STREP)
Duration: September 2011 - August 2014
Coordinator: SINTEF, Oslo (Norway)
Abstract: The project aims at significant improvement of the interoperability of computational tools for the design, analysis and optimization of functional products. An isogeometric approach is applied for selected manufacturing application areas (cars, trains, aircraft) and for computer-aided machining. Computer Aided Design and numerical simulation algorithms are vital technologies in modern product development, yet they are today far from being seamlessly integrated. Their interoperability is severely disturbed by inconsistencies in the mathematical approaches used. Efficient feedback from analysis to CAD and iterative refinement of the analysis model is a feature of isogeometric analysis, and would be an essential improvement for computer-based design optimization and virtual product development. Our vision is to provide and disseminate tangible evidence of the performance of the isogeometric approach in comparison to traditional ones in four important application areas as well as addressing interoperability and other issues that necessarily arise in a large-scale industrial introduction of isogeometry.

6.5.1.2. EXCITING

Title: Exact geometry simulation for optimized design of vehicles and vessels
Type: FP7-CP-SST-2007-RTD-1-218536, COOPERATION (TRANSPORTS)
Instrument: Specific Targeted Research Project (STREP)
Duration: October 2008 - September 2011
Coordinator: Jozef Kepler universitet, Linz (Austria)
Others partners:
SINTEF, Oslo (Norway); Siemens AG (Germany); National Technical University of Athens (Greece); Hellenic Register of Shipping (Greece); University of Technology, Munich (Germany); INRIA Méditerranée (France); VA Tech Hydro (Austria); Det Norske Veritas AS (Norway).
See also: http://exciting-project.eu/

Abstract: This project focuses on computational tools for the optimized design of functional free-form surfaces. Specific applications are ship hulls and propellers in naval engineering and car components, frames, and turbochargers in the automotive and railway transportation industries. The objective is to base the corresponding computational tools on the same exact representation of the geometry. This should lead to huge benefits for the entire chain of design, simulation, optimization, and life cycle management, including a new class of computational tools for fluid dynamics and solid mechanics, simulations for vehicles and vessels based. This seamless integration of CAD and FEM will have direct applications in product design, simulation and optimization of core components of vehicles and vessels.

6.5.1.3. SAGA

Title: ShApe, Geometry and Algebra, 2008-2012
Type: FP7-PEOPLE-2007-1-1-ITN.
Instrument: Initial Training Network (ITN)
Duration: November 2008 - October 2012
Coordinator: SINTEF (Norway)
Others partners: University of Oslo (Norway); Johannes Kepler Universitaet Linz (Austria); Universidad de Cantabria, Santander (Spain); Vilniaus Universitetas (Lithuany); National and Kapodistrian University of Athens (Greece); INRIA Méditerranée (France); GraphiTech (Italy); Kongsberg SIM GmbH (Austria); Missler Software (France);

See also: http://saga-network.eu/.

Abstract: The project aims at promoting the interaction between Geometric Modeling and Real Algebraic Geometry and, in general, at strengthening interdisciplinary and inter-sectorial research and development concerning CAD/CAM. Its objective is also to train a new generation of researchers familiar with both academic and industry viewpoints, while supporting the cooperation among the partners and with other interested collaborators in Europe.

6.5.1.4. DECONSTRUCT

Title: Decomposition of Structured Tensors, Algorithms and Characterization.
Type: PEOPLE (FP7-PEOPLE-2009-IEF)
Instrument: Marie Curie Intra-European Fellowships for Career Development (IEF)
Duration: November 2010 - November 2012
Coordinator: INRIA (France)

Others partners: No.

See also: http://www-sop.inria.fr/teams/galaad/joomla/index.php/international-collaborations-147/172-deconstruct.html

Abstract: Tensors play a wide role in numerous application areas as Signal Processing for Telecommunications, Arithmetic Complexity or Data Analysis. In some applications tensors may be completely symmetric, or symmetric only in some modes, or may not be symmetric. In most of these applications, the decomposition of a tensor into a sum of rank-1 terms is relevant, since tensors of interest have a reduced rank. Most of them are structured i.e. they are either symmetric or enjoy some index-invariance. Lastly, they are often real, which raises open problems concerning the existence and calculation of the decompositions. These issues build the basic bricks of the research program we propose. The classes of tensors described above have a geometric translation in terms of classical algebraic varieties: Segre, Veronese, Segre-Veronese varieties and Grassmannians and their secant varieties. A complete description of equations for those secant varieties and their dimensions is still not known (only dimensions of secant varieties to Veronesean are classified), although they have been studied by algebraic and differential geometers and algebraists for a long period up to now. The aim of this research project is:

- to attack both the description of the ideal of those secant varieties and their dimensions, starting from low dimensions and low degrees,
- to propose algorithms able to compute the rank of structured tensors.

6.6. International Initiatives

6.6.1. Visits of International Scientists

Annie Cuyt and Wen-Shin Lee (University of Antwerpen, Belgium) visited from January 24th to February 2nd to initiate a collaboration on the topic of shape from moments.

George Labahn (University of Waterloo, Canada) visited February 14th-18th and October 3rd-8th to collaborate with Evelyne Hubert on scaling invariants and their application to symmetry reduction of dynamical system (with parameters).

Mark Hickman (University of Canterbury, New Zealand) visited from March to June, as part of his sabbatical year, to collaborate with Evelyne Hubert on the topic of integral and moment invariants and their applications in computer vision.
Nelly Villamizar (University of Oslo) visited us from March 15 to May 15, to collaborate with B. Mourrain on splines spaces, for her secondement in the context of the ITN Marie-Curie SAGA.

Nguyen Tuan Thien (JKU, LINZ) visited us from March to May, to collaborate with B. Mourrain and A. Galligo on parameterization problems in isogeometric analysis, for his secondement in the context of the ITN Marie-Curie SAGA.

6.6.2. Participation In International Programs

6.6.2.1. PAI STAR South Corea collaboration

Participants: Laurent Busé, André Galligo, Evelyne Hubert, Angelos Mantzaflaris, Bernard Mourrain.

The objective of this collaboration is to conduct research in algebraic techniques for solving geometric modeling problems. More specially, we are interested in developing efficient and robust methods to solve non-linear constraints which appear in geometric computation. These methods will be used in applications such as shape design and reconstruction for solving interpolation or approximation problems. A typical area in which we will apply our methods is ship design. Experimentation and validation will lead to open source software implementation.

Collaborators from Seoul National University: Tae-Wan Kim, Sharma Rajiv, Hur Seok, Yeong-hwa Seo.

Tae-Wan Kimm visited INRIA-GALAAD from April 17 to April 23.
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. 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.](image)

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.
Figure 3. Two examples of anisotropic meshes produced by our algorithm.

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 $\hat{M}$. The algorithm uses only simple numerical operations. We show that the size of the sample is $O(\epsilon^{-k})$ and that $\hat{M}$ is a good triangulation of $M$. More specifically, we show that $M$ and $\hat{M}$ 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.

In collaboration with Jérôme Dedecker and Bertrand Michel

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 $W_2^2$ 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 ].

Figure 5. Reconstruction obtained on a data set with heterogeneous dimensions.

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.

![Hyperbolic Delaunay triangulation and Voronoi diagram in the Poicaré plane.](image)

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 $\alpha$-shapes of a Voronoi filament model realization. It concerns a sample of 200000 particles in a periodic box of $50 \, h^3\, \text{Mpc}$ size with 8 Voronoi cells. From top left to bottom right:
$\alpha = 0.5 \times 10^4, 1.0 \times 10^4, 2 \times 10^4, \text{and} 4.0 \times 10^4$. 
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.
MARELLE Project-Team

6. New Results

6.1. Type theory and formalization of mathematics

6.1.1. Foundational aspects of mechanized proofs

Participants: José Grimm, Loïc Pottier.

We attempt to prove all theorems in the “Theory of Sets” of Bourbaki. The first chapter describes Formal Mathematics, and we show that it can be interpreted in the Coq language, thanks to a bunch of axioms introduced by Carlos Simpson (CNRS, Nice), modulo some modifications. This work that was started in 2009, when J. Grimm was in the Apics project-team. A new formulation of this work using ssreflect has proved more efficient than the initial formulation relying on standard Coq.

The second chapter of Bourbaki covers the theory of sets, per se. It defines ordered pairs, correspondences, unions, intersections and products of a family of sets, as well as equivalence relations. The work of formalizing this chapter comprises 15000 lines of Coq script and is described in a technical report and a paper for the journal of formal reasoning published in 2010.

The third chapter of Bourbaki covers the theory of ordered sets, well-ordered sets, equipotent sets, cardinals, natural integers, and infinite sets; its implementation in Coq is described in [21]. This chapter is longer (22000 lines of code), and there are more exercises (18000 lines of code for about half of the exercises currently implemented).

We also looked at the univalent foundation proposed by V. Voevodsky to provide a new model for equality in type theory and simplified the proof that he proposed to derive extensionality from the univalence axiom.

6.1.2. Group theory (Character theory)

Participants: Georges Gonthier [Microsoft Research], Laurence Rideau, Laurent Théry.

We participate in the collaborative research agreement “Mathematical Components” with Microsoft Research. This project aims at evaluating the applicability of a new approach to mathematical proofs called “small-scale reflection”, especially in the domain of finite group theory [4].

This year, we have initiated the formalisation of the second book of the proof of Feit-Thompson’s theorem. The basic properties of character theories are now covered. This lets us formalised the first 4 chapters of the second book, “Character theory for the Odd Order Theorem” by Peterfalvi.

6.1.3. Proofs in geometry

Participants: Tuan Minh Pham, Yves Bertot.

The work on elementary (synthetic) geometry has been completed. A publication on the topic has also been presented at a conference [19]. This work was also the main content of Tuan Minh Pham’s thesis which was defended in November [5].

6.1.4. Towards constructive algebraic topology

Participants: Laurence Rideau, Maxime Dénès, Yves Bertot.

We have participated in the formalization of a complete chain of computation from an image (as a bitmap) to the corresponding Betti numbers and homology groups. In particular, we improved the formalization of “incidence simplicial matrices” in ssreflect. This work was described in conference article [17].

6.1.5. Computing with polynomials and matrices

Participants: Maxime Dénès, Yves Bertot.
The libraries of the project "Mathematical Components" propose a rather complete formalisation of polynomials and matrices. Unfortunately, these objects cannot be used directly for computing.

We have continued our study of executable algorithms to compute with matrices and polynomials inside Coq. In collaboration with other members of the European project Formath, we have looked at implementation of Strassen-Winograd and Karatsuba for fast matrix multiplication and other algorithms for various kinds of matrix normal forms: Smith normal form, Frobenius, and Jordan normal forms. This work is described in an article that has been submitted for publication.

6.1.6. Regularity of interval matrices
Participants: Guillaume Cano, Yves Bertot.

As part of our work on the regularity of interval matrices, we still needed to formalize the Perron-Frobenius theorem. This year we concentrated on an important lemma for this formalization, the Bolzano-Weierstrass theorem, which requires a usable formalization of general topology, in particular the concept of compact.

6.1.7. Type-based termination
Participants: Jorge Luis Sacchini, Benjamin Grégoire.

The work on this topic has been completed and is described in Jorge-Luis Sacchini’s Ph.D thesis, which was defended in June 2011 [6].

6.1.8. Native compilation of terms with primitive structures
Participants: Mathieu Boespflug [McGill University, Canada], Maxime Dénès, Benjamin Grégoire.

We kept working on the integration of the native compiler of the Ocaml language into a scheme for the efficient reduction of terms in the calculus of inductive constructions. This work is described in a publication at the conference CPP11 in Taiwan [14].

6.2. Proving tools

6.2.1. Connecting an SMT prover and Coq
Participants: Michaël Armand, Germain Faure [project-team Typical], Benjamin Grégoire, Chantal Keller [project-team Typical], Laurent Théry.

Our previous work on integrating SAT technology has been used as a basis to obtain SMT automation within Coq. We are now capable of replaying traces produced by the SMT prover VERIT that deal with conjunctive normal forms, congruence closures, and linear arithmetic. We are actively working on adding quantified formulae. This work is supported by the ANR Decert project. A preliminary version [10] of this work has been presented at the workshop PSATTT’11, a full version [9] at the conference CPP11. The generic exchange proof format [13] for SMT has been presented at the workshop PXTP’11.

6.2.2. Geometric Algebras and Automatic Theorem Proving
Participants: Laurent Fuchs [Université de Poitiers], Laurent Théry.

We have completed our work on Grassman-Cayley algebras. This has been published in the post-proceedings of the ADG’10 conference. We are now working on the natural continuation of this work: Clifford’s algebras. We have very encouraging preliminary results.

6.2.3. Taylor models in Coq
Participants: Erik Martin-Dorel [project-team Arénaire], Ioana Paşca [project-team Arénaire], Micaela Mayero [Université Paris XIII], Laurence Rideau, Laurent Théry.
Taylor models are a very effective way to approximate real functions with polynomials. We have started a formalisation of these models in the Coq prover. In a first step, we have concentrated our efforts in having a computational version of these models within Coq using native computations, certified floating point and interval arithmetics. Since our first evaluations show that they behave well computationally, we are now working on completing this work with the corresponding correctness proofs. This work is supported by the ANR Tamadi.

6.2.4. Tactics on polynomial equalities: nsatz

Participant: Loïc Pottier.

We started describing in the Coq programming language an efficient algorithm to compute Gröbner bases, similar to the one written in ocaml for the nsatz tactic. We hope to prove it correct and to use it for proofs by reflexion in commutative algebra.

6.2.5. D-Modules

Participant: Loïc Pottier.

We studied normalization of non-commutative polynomials ad exponentials in the Weyl algebra. The normal forms we found are similar with the one described found by Blasiak and Flajolet for graph models.

6.3. Formal study of cryptography

6.3.1. Certicrypt

Participants: Gilles Barthe, Benjamin Grégoire, Sylvain Heraud, Santiago Zanella.

CertiCrypt is a general framework to certify the security of cryptographic primitives in the Coq proof assistant. We completed a machine-checked proof of the security of OAEP (a widely public-key encryption scheme based on trapdoor permutations) against adaptive chosen ciphertext attacks under the assumption that the underlying permutation is partial-domain one-way. This work has been described in a publication at the conference CT-RSA 2011 in San Francisco [12].

6.3.1.1. Easycrypt

Participants: Gilles Barthe [IMDEA], Benjamin Grégoire, Sylvain Heraud, Anne Pacalet, Santiago Zanella.

Based on our experience with Certicrypt, we started last year the development of the tool Easycrypt. The goal of this work is to provide a friendly tool easily usable by cryptographers without knowledge of formal proof assistants. The idea is to use the techniques formally proved in Certicrypt and to call SMT-provers instead of using Coq. We have applied Easycrypt on a variety of academic examples and one bigger example: the proof of IND-CCA security of the Cramer-Shoup cryptosystem. The drawback of this tool is that it provide less guarantees than Certicrypt for the correctness of the proof. To fill this gap we are now able to generate Coq files (based on Certicrypt) allowing to check the validity of Easycrypt proofs. This work has been described in a publication at the conference CRYPTO 2011 in Santa Barbara and has obtained the best paper Award [11].
6. New Results

6.1. Tralics: a LaTeX to XML Translator

Participant: José Grimm.

The major use of Tralics remains the production of the RaWeb (Scientific Annex to the Annual Activity Report of Inria). The software is described in [65], [68], [67], [66]. Other applications of Tralics consist in putting scientific papers on the Web; for instance Cedram (http://www.cedram.org, Centre de diffusion de revues académiques mathématiques), that publishes the Journal de théorie des nombres de Bordeaux, uses Tralics for the abstracts and plans to translate full papers. Tralics is also used by Zentralblatt for converting comments, reviews and abstracts. The Software has been presented at the DML2010 conference [69]. Tralics has been used for the HTML+MathML documentation of the open TURNS software http://trac.openturns.org/.

6.2. Inverse problems for elliptic operators


6.2.1. Boundary value problems for Laplace equation in 3-D

Solving overdetermined Cauchy problems for the Laplace equation on a spherical layer (in 3-D) in order to process incomplete experimental data is a necessary ingredient of the team’s approach to inverse source problems, in particular for applications to EEG since the latter involves propagating the initial conditions from the boundary to the center of the domain where the singularities (i.e., the sources) are sought after. Here, the domain is typically made of several homogeneous layers of different conductivities.

Such problems offer an opportunity to state and solve extremal problems for harmonic fields for which an analog of the Toeplitz operator approach to bounded extremal problems [45] has been obtained in [2]. Still, a best approximation on the subset of a general vector field generated by a harmonic gradient under a $L^2$ norm constraint on the complementary subset can be computed by an inverse spectral equation for some Toeplitz operator. Constructive and numerical aspects of the procedure (harmonic 3-D projection, Kelvin and Riesz transformation, spherical harmonics) and encouraging results have been obtained on numerically simulated data.

Issues of robust interpolation on the sphere from incomplete pointwise data are also under study in order to improve numerical accuracy of our reconstruction schemes. Spherical harmonics, Slepian bases and related special functions are of special interest (thesis of A.-M. Nicu), while splines, spherical wavelets, cubature techniques should be considered as well.

It turns out that Slepian functions are eigenfunctions of truncated Toeplitz operators in the complex plane (the framework of 2-D problems). These properties will be used in order to quantify the robustness properties of our resolution schemes for $L^2$ bounded extremal problems [45], and to establish error estimates.
The analogous problem in $L^p$, $p \neq 2$, is quite important to get tighter control on pointwise approximation. However, it is considerably more difficult. In a collaborative effort with the university of Orléans, within the framework of the ANR project AHPI, we set ourselves the goal of understanding the case $p = \infty$ better. Namely, connections between the BMO $^2$ distance of a bounded vector field on the sphere to a BMO harmonic gradient and the spectral properties of a “big” Hankel-like operator acting on $L^2$ harmonic gradients and valued in its orthogonal space are currently being investigated. We obtained a generalization of the Hodge decomposition (we call it the Hardy-Hodge decomposition) for $\mathbb{R}^n$-valued vector fields on $\mathbb{R}^{n-1}$ (resp. $S^{n-1}$) which stands analog to the Hardy direct sum decomposition in dimension 1. In this decomposition, the analytic part becomes the trace of the gradient of a harmonic function in the half-space (resp. the ball) whose BMO-norm on parallel hyperplanes (resp. concentric spheres) is uniformly bounded. The two main difficulties facing a generalization of Nehari’s theorem are the absence of a constructive derivation of Coifman-Rochberg weak factorizations and Wolff’s phenomenon that a harmonic gradient (even $C^1$-smooth) is not determined by its values on a set of positive measure on $\mathbb{R}^n$ (resp. $S^n$). This last point is the only obstacle to establish uniqueness and constancy-of-modulus properties. We shall concentrate on these two items in the future.

The above issue is also interesting in $L^p$, $1 < p < \infty$, where it leads to analyze particular solutions to the the $p$-Laplacian on the sphere. This aspect is not pursued in depth at the moment.

6.2.2. Sources recovery in 3-D domains, application to MEEG and geophysics

The problem of sources recovery can be handled in 3-D balls by using best rational approximation on 2-D cross sections (disks) from traces of the boundary data on the corresponding circles (see section 4.1).

In 3-D, functional or clinical active regions in the cortex are often represented by pointwise sources that have to be localized from measurements on the scalp of a potential satisfying a Laplace equation (EEG, electroencephalography). In the work [7] it was shown how to proceed via best rational approximation on a sequence of 2-D disks cut along the inner sphere, for the case where there are at most 2 sources. A long-haul research on the behaviour of poles of best rational approximants of fixed degree to functions with branch points was completed this year [19], which shows that the technique carries over to finitely many sources. In this connection, a dedicated software “FindSources3D” (see section 5.7) has been developed, in collaboration with the team Athena.

Further, it appears that in the rational approximation step of these schemes, multiple poles possess a nice behaviour with respect to the branched singularities (see figure 4). This is due to the very basic physical assumptions on the model (for EEG data, one should consider triple poles). Though numerically observed, there is no mathematical justification so far why these multiple poles have such strong accumulation properties, which remains an intriguing observation. This is the topic of [30].

Also, magnetic data from MEG (magneto-encephalography) will soon become available, which should enhance the accuracy of source recovery algorithms.

This approach also appears to be interesting for geophysical issues, namely discretizing the gravitational potential by means of pointwise masses. This is one topic of A.-M. Nicu’s PhD thesis.

Magnetic sources localization from observations of the field away from the support of the magnetization is an issue under investigation in a joint effort with the Math. department of Vanderbilt University and the Earth Sciences department at MIT. The goal is to recover the magnetic properties of rock samples (meteorites) from fine measurements extremely close to the sample that can nowadays be obtained using SQUIDs (supraconducting coil devices).

The magnetization operator is the Riesz potential of the divergence of the magnetization. When the latter has bounded variation, we already described the kernel of this operator (the so-called silent magnetizations or silent source distributions) in terms of measures whose balayage on the boundary of the sample vanishes. This however, is not so very effective, computationally.
The case of a thin slab (the magnetization is then modelled as a vector field on a portion of the plane) has proved more amenable. We have shown that that silent sources from above or below can be characterized via the Hardy-Hodge decomposition mentioned in section 6.2.1. The smoothness assumptions have been weakened considerably to accommodate magnetizations that may be any distribution with compact support, more generally any finite sum of partial derivatives of any order of $L^p$ or $BMO$ functions. Silent unidirectional and bi-directional magnetizations demonstrably reduce this way to certain divergence free tangential vector fields. In particular no nonzero compactly supported unidirectional magnetization exists. In the $L^2$ setting, equivalent magnetizations of minimal $L^2$-norm can be computed using the Hardy-Hodge decomposition (which is orthogonal in this case), and an uncertainty principle relating the support of a magnetization and the support of its minimum-norm equivalent magnetization has been obtained. A paper is being written on these results.

Meanwhile, the severe ill-posedness of the reconstruction challenges discrete Fourier methods, one of the main problems being the truncation of the observations outside the range of the SQUID measurements. A next step will be to develop the extrapolation techniques initiated by the project team, using bounded extremal problems, in an attempt to overcome this issue.

6.2.3. Boundary value problems for 2-D conductivity equations, application to plasma control

In collaboration with the CMI-LATP (University Marseille I) and in the framework of the ANR AHPI, the team considers 2-D diffusion processes with variable conductivity. In particular its complexified version, the so-called conjugate or real Beltrami equation, was investigated. In the case of a smooth domain, and for Lipschitz conductivity, we analyzed the Dirichlet problem for solutions in Sobolev and then in Hardy classes $\mathcal{C}$. Their traces merely lie in $L^p (1 < p < \infty)$ of the boundary, a space which is suitable for identification from pointwise measurements. Again these traces turn out to be dense on strict subsets of the boundary. This allows us to state Cauchy problems as bounded extremal issues in $L^p$ classes of generalized analytic functions, in a reminiscent manner of what was done for analytic functions as discussed in section 3.1.1.

This year we generalized the construction to finitely connected Dini-smooth domains and $W^{1,q}$-smooth conductivities, with $q > 2$ [43]. The case of an annular geometry is the relevant one for the application to plasma shaping mentioned below [17]. The application that initially motivated this work came from free boundary problems in plasma confinement (in tokamaks) for thermonuclear fusion. This work was initiated in
collaboration with the Laboratoire J. Dieudonné (University of Nice) and is now the topic of a collaboration with two teams of physicists from the CEA-IRFM (Cadarache).

In the transversal section of a tokamak (which is a disk if the vessel is idealized into a torus), the so-called poloidal flux is subject to some conductivity equation outside the plasma volume for some simple explicit smooth conductivity function, while the boundary of the plasma (in the Tore Supra tokamak) is a level line of this flux [53]. Related magnetic measurements are available on the chamber, which furnish incomplete boundary data from which one wants to recover the inner (plasma) boundary. This free boundary problem (of Bernoulli type) can be handled through the solutions of a family of bounded extremal problems in generalized Hardy classes of solutions to real Beltrami equations, in the annular framework. Such approximation problems also allow us to approach a somewhat dual extrapolation issue, raised by colleagues from the CEA for the purpose of numerical simulation. It consists in recovering magnetic quantities on the outer boundary (the chamber) from an initial guess of what the inner boundary (plasma) is.

In the particular case at hand, the conductivity is $1/x$ and the domain is an annulus embedded in the right half-plane. We obtained a basis of solutions (exponentials times Legendre functions) upon separating variables in toroidal coordinates. This may be viewed as a generalization to the annulus of the Bessel type expansions derived in [21] for simply connected geometries. This provides a computational setting to solve the extremal problems mentioned before, and was the topic of the PhD thesis of Y. Fischer [17], [22]. In the most recent tokamaks, like Jet or ITER, an interesting feature of the level curves of the poloidal flux is the occurrence of a cusp (a saddle point of the poloidal flux, called an X point), and it is desirable to shape the plasma according to a level line passing through this X point for physical reasons related to the efficiency of the energy transfer. We established well-posedness of the Dirichlet problem in weighted $L^p$ classes for harmonic measure on piecewise smooth domains without cusps, thereby laying ground for such a study. This issue is next in line, now that the present approach has been validated numerically on Tore Supra data.

On the half-plane, the conductivity $1/x$ is severely unbounded but the analysis of this test case is quite important for the convergence of extrapolation algorithms to recover magnetic quantities on the chamber. Additive decompositions into Hardy solutions inside the outer boundary and outside the inner boundary, with controlled vanishing on the imaginary axis, have been obtained as part of the PhD work of S. Chaabi. The latter developed this year a multiplicative parameterization of Hardy-smooth solutions by holomorphic functions for the conjugate Beltrami equation, a result which is both subtler and weaker than the classical Stoilow factorization for solutions to the complex Beltrami equation. This factorization is of considerable numerical interest in situations where conductivity is little or not known.

6.3. Rational Approximation for fitting Non Negative EPT densities

**Participants:** Martine Olivi, Bernard Hanzon [Univ. Cork], Conor Sexton [Univ. Cork], Fabien Seyfert.

We explored this year a new application field for our rational approximation methods. We studied the problem of fitting a probability density function from a large set of financial data. The class of density function that we considered is that of non negative EPT (Exponential-Trigonometric-Polynomials) functions that seems to provide a very relevant framework for probabilistic calculations. Parseval’s theorem implies that approximating the rational transform is equivalent to approximating the density itself. During his visit, Conor Sexton (a PHD student of Bernard Hanzon) adapted and ran the RARL2 software on this problem. The results were encouraging albeit the major problem of imposing positivity is still under study.

6.4. Orthogonal rational functions and non-stationary stochastic processes

**Participants:** Laurent Baratchart, Stanislas Kupin [Univ. Bordeaux 1].
The theory of orthogonal polynomials on the unit circle is a most classical piece of analysis which is still the object of intensive studies. The asymptotic behaviour of orthogonal polynomials is of special interest for many issues pertaining to approximation theory and to spectral theory of differential operators. Its connection with prediction theory of stationary stochastic processes has long been known \[64\]. Namely, the \(n\)-th orthonormal polynomial with respect to the spectral measure of the process yields the optimal regression coefficients of a linear one-step ahead predictor from the \((n-1)\)-st last values, in the sense of minimum variance of the error. Likewise, the (inverse of) the dominant coefficient of the polynomial gives the prediction error. In particular, asymptotics for the dominant coefficient determine the asymptotically optimal prediction error from the past as time goes large.

As compared to orthogonal polynomials, orthogonal rational functions have not been much considered up to now. They were apparently introduced by Dzrbasjan but the first systematic exposition seems to be the monograph by Bultheel et al. \[57\] where the emphasis is more on the algebraic side of the theory. In fact, the asymptotic analysis of orthogonal rational functions is still in its infancy.

We recently developed an analog of the Kolmogorov-Krein-Szegö theorem \[18\] for orthogonal rational functions which is first of its kind in that it allows for the poles of these functions to approach the unit circle, generalizing previously known results for compactly supported singular set. Dwelling on this asymptotic analysis of orthogonal rational functions, we developed a prediction theory for certain, possibly nonstationary stochastic processes that we call Blaschke varying processes. \[44\]. These are characterized by a spectral calculus where time shift corresponds to multiplication by an elementary Blaschke product (that may depend on the time instant considered). This class of processes contains the familiar Gaussian stationary processes, but it contains many more that exhibit a much more varied behaviour. For instance, the process may be asymptotically deterministic along certain subsequences and nondeterministic along others. The optimal predictor is constructed from the spectral measure via orthogonal rational functions, and its asymptotic behaviour is characterized by the above-mentioned generalization of the Kolmogorov-Krein-Szegö theorem. In the same vein, we also developed prediction theory for another class of nonstationary processes, the so-called Cauchy-processes, that may be characterized as stationary processes feeding in turn a sequence of varying filters of degree 1. Their covariance matrices can be characterized via Nevanlinna-Pick interpolation. The issue of characterizing covariance sequences of Blaschke processes is still open. Their identification raises the problem of constructing optimal Schur rational approximants to a given Schur function.

### 6.5. Rational and meromorphic approximation

**Participants:** Laurent Baratchart, Herbert Stahl [TFH Berlin], Maxim Yattselev.

We demonstrated in a recent past, under mild smoothness assumptions, the possibility of convergent rational interpolation to Cauchy integrals of complex measures on analytic Jordan arcs and their strong asymptotics \[14\], \[13\]. Subsequently, we started investigating the case of Cauchy integrals on so-called symmetric contours for the logarithmic potential. These correspond to functions with more than two branched singularities, like those arising in the slicing method for source recovery in a sphere when there is more than one source (see section 6.2.2). Recently we obtained weak asymptotics in this case, and dwelled on them to elucidate the asymptotic of poles of best \(L^2\) meromorphic approximants of given degree to a function with branched singularities on a curve encompassing them. Namely, the counting measure of the poles converges weak-star, when the degree goes large, to the Green equilibrium distribution of the set with minimum Green capacity inside the curve, outside of which the function is single-valued. The technical core of this contribution is an existence and uniqueness result, along with a differential characterization, of the compact of minimum weighted capacity outside of which the function is single-valued \[19\]. This teams up with results from \[63\] to produce the results.

We presently study strong asymptotics, limiting ourselves at present to a threefold geometry, and to the case of Padé approximants (interpolation at a single point with high order). The result is that uniform convergence can only take place if the weights of the branches of the threefold with respect to the equilibrium distribution are rational. If they are rationally dependent, a spurious pole clusters to certain curves within the domain of analyticity, and if they are rationally independent, exactly one pole exhibits chaotic behaviour in the complex...
plane. Moreover, we have shown that the chaotic situation is generic, in a measure theoretic sense, with respect to the location of branchpoints. This generalizes and sharpens results of Suetin for Cauchy integrals on disconnected pieces of a smooth symmetric contour. It is the first time that a branched contour is analyzed with respect to general densities. A paper is being written to report these results.

6.6. Circuit realisations of filter responses: determination of canonical forms and exhaustive computations of constrained realisations

Participants: Smain Amari [Royal Military College, Kingston, Canada], Jean Charles Faugère [EPI SALSA, INRIA Rocquencourt], Giuseppe Macchiarella [Politecnico di Milano, Milan, Italy], Uwe Rosenberg [Design and Project Engineering, Osterholz-Scharmbeck, Germany], Matteo Oldoni [Politecnico di Milano, Milan, Italy], Fabien Seyfert.

We pursued our work on circuit realisations of filter responses with mixed type (inductive or capacitive) coupling elements and constrained topologies [1]. We now focus on the use of resonating couplings in the design of asymmetric filter’s characteristics without the use cross-coupling in order to simplify the practical implementation. In parallel, efforts are being payed to improve the synthesis method for higher order filters, having in mind application to díplexers with high number of symmetrically located transmission zeros.

6.7. Synthesis of compact multiplexers and de-embedding of multiplexers

Participants: Martine Olivi, Fabien Seyfert, Stéphane Bila [Xlim, Limoges, France], Hussein Ezzedin [Xlim, Limoges, France], Damien Pacaud [Thales Alenia Space, Toulouse, France], Giuseppe Macchiarella [Politecnico di Milano, Milan, Italy], Matteo Oldoni [Politecnico di Milano, Milan, Italy].

6.7.1. Synthesis of compact multiplexers

Our work on the synthesis problem for diplexers has continued this year. Based on the polynomial structure highlighted last year [39], a synthesis algorithm was devised and allowed the effective synthesis of multiplexer characteristics (see figure 5). As opposed to other synthesis algorithms [74] the latter only involves polynomial computations on the sub-filters of the overall multiplexer, allowing its application to synthesis problems relative to devices with a numerous number of ports. The latter is based on the recursive solving of an extended Nevanlinna-Pick problem, first introduced in [62]. The convergence of this fixed point procedure is under study as well as its extension to the synthesis of general multiplexer.

Figure 5. Diplexer’s characteristic computed with an algorithm based on the recursive solving of extended Nevanlinna-Pick interpolation problems.
6.7.2. De-embedding of multiplexers

While we presented our work on the de-embedding diplexers at [26] we extended the latter to general multiplexers. The multiplexer de-embedding problem we study is the following. Let $S$ be the measured scattering matrix of a multiplexer composed of a $N$-port junction with response $T$ and $N-1$ filtering devices with responses $F_1, \cdots, F_{N-1}$ as depicted by Figure 6.

![Figure 6. Multiplexer made of a junction T and filtering devices $F_1, F_2, \cdots, F_N$.](image)

The de-embedding question is the following: given $S$ and $T$, is it possible to derive the $F_k$’s? We derived some rather important and surprising characteristics of this general problem:

- When stating the de-embedding problem in terms of chain rather that scattering matrices, the latter becomes linear,
- For a generic junction the de-embedding problem has a unique solution for $N = 2$ and $N > 3$ while it is degenerated for $N = 3$,
- Degeneracy depends on the junction’s response and can be explained by the occurrence of filtering responses that hide behind the junction. These are responses $F_1, \cdots, F_{N-1}$ such that when chained to the junction, the response $T$ remains unchanged.

In the light of these theoretical results the tractability of the practical problem is currently under investigation. For example, it appears that in practice the $T$ responses can not be considered as generic as they are often built by chaining several 3-junctions together. Moreover, while theoretically uniqueness of the de-embedding solution seems to be guaranteed for large values of $N$, the sensitivity of the latter with respect to measurement errors grows dramatically with $N$. Therefore approaches that entail additional hypotheses, such as rationality of the responses $F_k$ are being studied. This work is pursued in collaboration with Thales Alenia Space and the Politecnico di Milano.

6.8. Averaging in control

**Participants:** Bernard Bonnard, Jean-Baptiste Pomet, Jana Nemcova.

A reference paper on the construction and properties of an “average control system”, has been submitted [27]; it is based on Alex Bombrun’s doctoral work [54] (defended in 2007). It connects properties of convergence of solutions of highly oscillating control systems to those of an average control system, when the frequency of oscillation goes high. Likewise, it details (on a time-interval that goes to infinity) the properties of solutions
of a conservative system with small controls in relation to those of an average system as the magnitude of control goes to zero. It also gives many properties of this average control system that has “more controls” than the original system, and yields, when this number of new controls is maximal, a Finsler metric on the state manifold. It is however difficult to compute explicitly and is never twice differentiable.

In [33], we study into details this average system arising from low-thrust orbital transfer, in the restricted “meridian” co-planar case, and prove that its trajectories for minimum time never leave the “elliptic domain” where averaging is valid. This gives some ground to using it as a limit to describe transfer from an elliptic orbit to another.

More exploration on this average system and the corresponding Finsler metric is planned.

6.9. Optimal transport

Participants: Bernard Bonnard, Ahed Hindawi, Jean-Baptiste Pomet, Ludovic Rifford.

In [23], we provide results on existence and continuity of the optimal transport map between absolutely continuous measures (with respect to Lebesgue measure) for a point-to-point cost coming from controllable linear-quadratic optimal control. This is the simplest case of a cost coming from a system with drift; it bears some interest because the cost can be explicitly computed. A. Hindawi’s PhD also aims at more general systems. It will be defended in 2012.

6.10. Detection of the instability of amplifiers

Participants: Laurent Baratchart, Sylvain Chevillard, Martine Olivi, Jean-Baptiste Pomet, Fabien Seyfert.

A new collaboration with the CNES and the University of Bilbao began this year. The goal is to help the development of amplifiers, in particular to detect instability at an early stage of the design.

Currently, Electrical Engineers from the University of Bilbao, under contract with CNES (the French Space Agency), use heuristics to diagnose instability before the circuit is physically implemented. We intend to set up a rigorously founded algorithm instead, based on properties of transfer function of such amplifiers which belong to particular classes of analytic functions.

We completed the first stage of this collaboration, in that we now have a formal definition of stability within these classes: a stable function is one which, when connected in parallel with a large resistor, yields a $H^\infty$ transfer function (in this context transfer functions might have infinitely many poles). We also obtained a characterization of transfer functions that may actually be realized in terms of elementary electric components (i.e., resistors, self, capacities, LEDs, and transmission lines): they are the rational functions in the variable and the exponentials thereof.

6.11. Tools for numerically guaranteed computations

Participant: Sylvain Chevillard.

This work has been performed in collaboration with members of the teams Arénaire in Lyon and Caramel in Nancy. The overall and long-term goal is to enhance the quality of numerical computations. Several aspects are studied:

- A first topic is the development of software code for the multiprecision evaluation of elementary and special functions. Developing such codes is a long and error-prone task. It is hence relevant to automatically generate such codes whenever possible. A first step has been to design an algorithm that automatically generates multiprecision code for the evaluation of constant expressions with an a priori guaranteed error [25]. This is usually necessary for the evaluation, e.g., of the first terms of a Taylor series.
• Another topic consists in the design of algorithms that allow developers of double precision mathematical libraries (so-called libm) to certify their library. In the process of developing a libm, one usually replaces the function $f$ to be evaluated by a good polynomial approximation $p$. In order to certify the quality of the library, it is then necessary to give a rigorous mathematical proof that the relative error $\varepsilon = (p - f)/f$ between $p$ and $f$ is bounded by a small constant. This turns out to be equivalent to the problem of computing a sharp yet certified upper bound of the supremum norm of $\varepsilon$. An efficient algorithm has been designed for this purpose [20] (this work is the publication of a work initially begun in the Arénaire team and continued in the Caramel team).

• Finally, a more general endeavor is to develop a tool that helps developers of libms in their task. This is performed by the software Sollya, which has originally been developed in the Arénaire team, in collaboration with C. Lauter and M. Joldeș. A new release has been performed this year [32].
NACHOS Project-Team

6. New Results

6.1. Discontinuous Galerkin methods for Maxwell’s equations

6.1.1. DGTD-\textit{P}_p method based on hierarchical polynomial interpolation

**Participants:** Loula Fezoui, Joseph Charles, Stéphane Lanteri.

The DGTD (Discontinuous Galerkin Time Domain) method originally proposed by the team for the solution of the time domain Maxwell’s equations \[14\] relies on an arbitrary high order polynomial interpolation of the component of the electromagnetic field, and its computer implementation makes use of nodal (Lagrange) basis expansions on simplicial elements. The resulting method is often denoted by DGTD-\textit{P}_p, where \( p \) refers to the interpolation degree that can be defined locally i.e. at the element level. In view of the design of a \( hp \)-adaptive DGTD method, i.e. a solution strategy allowing an automatic adaptation of the interpolation degree \( p \) and the discretization step \( h \), we now investigate alternative polynomial interpolation and in particular those which lead to hierarchical or/and orthogonal basis expansions. Such basis expansions on simplicial elements have been extensively studied in the context of continuous finite element formulations (e.g. \[46\]) and have thus been designed with global conformity requirements (i.e. \( H_1 \), \( H(\text{rot}) \) or \( \text{div} \)) whose role in the context of a discontinuous Galerkin formulation has to be clarified. This represents one of the objectives of this study.

![Figure 3. Scattering of a plane wave by a business aircraft geometry computed by DGTD-\textit{P}_1 method on a tetrahedral mesh. Contour lines of the amplitude of the electric field.](image)

6.1.2. DGTD-\textit{P}_p\textit{Q}_k method on multi-element meshes

**Participants:** Clément Durochat, Stéphane Lanteri, Claire Scheid, Mark Loriot [Distene, Pôle Teratec, Bruyères-le-Chatel].
In this work, we study a multi-element DGTD method formulated on a hybrid mesh which combines a structured (orthogonal) quadrangulation of the regular zones of the computational domain with an unstructured triangulation for the discretization of the irregularly shaped objects. The general objective is to enhance the flexibility and the efficiency of DGTD methods for large-scale time domain electromagnetic wave propagation problems with regards to the discretization process of complex propagation scenes. As a first step, we have designed and analyzed a DGTD-\(\mathbb{P}_p\mathbb{Q}_k\) method formulated on conforming hybrid quadrangular/triangular meshes for the solution of the 2D Maxwell’s equations.

![Figure 4. Scattering of a plane wave by a perfectly conducting cylinder computed by a DGTD-\(\mathbb{P}_2\mathbb{Q}_4\) method on a hybrid triangular-quadrangular mesh.](image)

**6.1.3. DGTD-\(\mathbb{P}_p\) method for dispersive materials**  
**Participants:** Claire Scheid, Maciej Klemm [Communication Systems & Networks Laboratory, Centre for Communications Research, University of Bristol, UK], Stéphane Lanteri.

This work is undertaken in the context of a collaboration with the Communication Systems & Networks Laboratory, Centre for Communications Research, University of Bristol (UK). This laboratory is studying imaging modalities based on microwaves with applications to dynamic imaging of the brain activity (Dynamic Microwave Imaging) on one hand, and to cancerology (imaging of breast tumors) on the other hand. The design of imaging systems for these applications is extensively based on computer simulation, in particular to assess the performances of the antenna arrays which are at the heart of these systems. In practice, one has to model the propagation of electromagnetic waves emitted from complex sources and which propagate and interact with biological tissues. In relation with these issues, we study the extension of the DGTD-\(\mathbb{P}_p\) method originally proposed in [14] to the numerical treatment of electromagnetic wave propagation in dispersive media. We consider an approach based on an auxiliary differential equation modeling the time evolution of the electric polarization for a dispersive medium of Debye type (other dispersive media will be considered subsequently). This work comprises both theoretical aspects (stability and convergence analysis) of the resulting DGTD-\(\mathbb{P}_p\) method for the time domain Maxwell equations for dispersive media, and application aspects [35].

**6.1.4. DGFD-\(\mathbb{P}_p\) method for the frequency domain Maxwell equations**  
**Participants:** Victorita Dolean, Mohamed El Bouajaji, Stéphane Lanteri, Ronan Perrussel [Laplace Laboratory, INP/ENSEEIHT/UPS, Toulouse].
For certain types of problems, a time harmonic evolution can be assumed leading to the formulation of the frequency domain Maxwell equations, and solving these equations may be more efficient than considering the time domain variant. We are studying a high order Discontinuous Galerkin Frequency Domain (DGFD-$P_p$) method formulated on triangular meshes for solving the 2D time harmonic Maxwell equations [16]. This work is undertaken in the context of the ANR MAXWELL project whose objective is the development of an ultra wideband georadar system for imaging the subsurface. In this context, the DGFD-$P_p$ method that we have proposed is used as the forward solver in an inversion process for the electric permittivity [17].

6.1.5. **Hybridized DGFD-$P_p$ method**

**Participants:** Stéphane Lanteri, Liang Li, Ronan Perrussel [Laplace Laboratory, INP/ENSEEIHT/UPS, Toulouse].

One major drawback of DG methods is their intrinsic cost due to the very large number of globally coupled degrees of freedom as compared to classical high order conforming finite element methods. Different attempts have been made in the recent past to improve this situation and one promising strategy has been recently proposed by Cockburn *et al.* [40] in the form of so-called hybridizable DG formulations. The distinctive feature of these methods is that the only globally coupled degrees of freedom are those of an approximation of the solution defined only on the boundaries of the elements. This work is concerned with the study of such Hybridizable Discontinuous Galerkin (HDG) methods for the solution of the system of Maxwell equations in the time domain when the time integration relies on an implicit scheme, or in the frequency domain. As a first step, HDGTD and HDGFD [33] methods have been developed for the solution of the 2D propagation problems.

6.1.6. **Exact transparent condition in a DGFD-$P_p$ method**

**Participants:** Mohamed El Bouajaji, Nabil Gmati [ENIT-LAMSIN, Tunisia], Stéphane Lanteri, Jamil Salhi [ENIT-LAMSIN, Tunisia].

In the numerical treatment of propagation problems theoretically posed in unbounded domains, an artificial boundary is introduced on which an absorbing condition is imposed. For the frequency domain Maxwell equations, one generally use the Silver-Müller condition which is a first order approximation of the exact radiation condition. Then, the accuracy of the numerical treatment greatly depends on the position of the artificial boundary with regards to the scattering object. In this work, we have conducted a preliminary study aiming at improving this situation by using an exact transparent condition in place of the Silver-Müller condition. Promising results have been obtained in the 2D case and call for an extension of this work to the more challenging 3D case.

6.2. **Discontinuous Galerkin methods for the elastodynamic equations**

6.2.1. **DGTD-$P_p$ method for the elastodynamic equations**

**Participants:** Nathalie Glinsky, Fabien Peyrusse.

We continue developing high order non-dissipative discontinuous Galerkin methods on simplicial meshes for the numerical solution of the first order hyperbolic linear system of elastodynamic equations. These methods share some ingredients of the DGTD-$P_p$ methods developed by the team for the time domain Maxwell equations among which, the use of nodal polynomial (Lagrange type) basis functions, a second order leap-frog time integration scheme and a centered scheme for the evaluation of the numerical flux at the interface between neighboring elements. Recent results concern two particular points.

The first novelty is the extension of the DGTD-$P_p$ method initially introduced in [5] to the numerical treatment of viscoelastic attenuation. For this, the velocity-stress first order system is completed by additional equations for the anelastic functions describing the strain history of the material. These additional equations result from the rheological model of the generalized Maxwell body and permit the incorporation of realistic attenuation properties of viscoelastic material accounting for the behaviour of elastic solids and viscous fluids. In practice, one needs to add $3L$ additional equations in 2D and $6L$ in 3D, where $L$ is the number of relaxation mechanisms of the generalized Maxwell body. This method has been implemented in 2D and validated thanks to comparisons with a FDTD method.
The second contribution is concerned with the numerical assessment of site effects especially topographic effects. The study of measurements and experimental records proved that seismic waves can be amplified at some particular locations of a topography. Numerical simulations are exploited here to understand further and explain this phenomenon. The DGTD-$\mathbb{P}_p$ method has been applied to a realistic topography of Rognes area (where the Provence earthquake occurred in 1909) to model the observed amplification and the associated frequency. Moreover, the results obtained on several homogeneous and heterogeneous configurations prove the influence of the medium in-depth geometry on the amplifications measures at the surface [26], [25].

6.3. Time integration strategies and resolution algorithms

6.3.1. Hybrid explicit-implicit DGTD-$\mathbb{P}_p$ method

Participants: Stéphane Descombes, Stéphane Lanteri, Ludovic Moya.

Existing numerical methods for the solution of the time domain Maxwell equations often rely on explicit time integration schemes and are therefore constrained by a stability condition that can be very restrictive on highly refined meshes. An implicit time integration scheme is a natural way to obtain a time domain method which is unconditionally stable. Starting from the explicit, non-dissipative, DGTD-$\mathbb{P}_p$ method introduced in [14], we have proposed the use of a Crank-Nicolson scheme in place of the explicit leap-frog scheme adopted in this method [4]. As a result, we obtain an unconditionally stable, non-dissipative, implicit DGTD-$\mathbb{P}_p$ method, but at the expense of the inversion of a global linear system at each time step, thus obliterating one of the attractive features of discontinuous Galerkin formulations. A more viable approach for 3D simulations consists in applying an implicit time integration scheme locally i.e. in the refined regions of the mesh, while preserving an explicit time scheme in the complementary part, resulting in an hybrid explicit-implicit (or locally implicit) time integration strategy. Such an approach, combining a leap-frog scheme and a Crank-Nicolson scheme, has been studied numerically in [6], showing promising results which have motivated further investigations on theoretical issues (especially, convergence in the ODE and PDE senses) [28].

6.3.2. Explicit local time stepping DGTD-$\mathbb{P}_p$ method

Participants: Joseph Charles, Julien Diaz [MAGIQUE-3D project-team, INRIA Bordeaux - Sud-Ouest], Stéphane Descombes, Stéphane Lanteri.

We have initiated this year a collaboration with the MAGIQUE-3D project-team aiming at the design of local time stepping strategies inspired from [41] for the time integration of the system of ordinary differential equations resulting from the discretization of the time domain Maxwell equations in first order form by a DGTD-$\mathbb{P}_p$ method. A numerical study in one- and two-space dimensions is underway.

6.3.3. Optimized Schwarz algorithms for the frequency domain Maxwell equations

Participants: Victorita Dolean, Mohamed El Bouajaji, Martin Gander [Mathematics Section, University of Geneva], Stéphane Lanteri, Ronan Perrussel [Laplace Laboratory, INP/ENSEEIHT/UPS, Toulouse].

We continued with the design of optimized Schwarz algorithms for the solution of the frequency domain Maxwell equations. In particular, we have analyzed a family of methods adapted to the case of conductive media [21]. Besides, we have also proposed discrete variants of these algorithms in the framework of a high order discontinuous Galerkin discretization method formulated on unstructured triangular meshes for teh solution of the 2D time harmonic Maxwell equations.

6.4. High performance computing

6.4.1. High order DGTD-$\mathbb{P}_p$ method on hybrid CPU/GPU parallel systems

Participants: Tristan Cabel, Stéphane Lanteri.
Modern massively parallel computing platforms most often take the form of hybrid shared memory/distributed memory heterogeneous systems combining multi-core processing units with accelerator cards. In particular, graphical processing units (GPU) are increasingly adopted in these systems because they offer the potential for a very high floating point performance at a low purchase cost. DG methods are particularly appealing for exploiting the processing capabilities of a GPU because they involve local linear algebra operations (mainly matrix/matrix products) on relatively dense matrices whose size is directly related to the approximation order of the physical quantities within each mesh element. We have initiated this year a technological development project aiming at the adaptation to hybrid CPU/GPU parallel systems of a high order DGTD-$P_p$ method for the numerical solution of the 3D Maxwell equations.
6. New Results

6.1. Mathematical analysis and control of macroscopic traffic flow models

6.1.1. Vehicular traffic

Participants: Maria Laura Delle Monache, Paola Goatin, Mauro Garavello [Piedmont University, Italy].

Concerning road traffic, the research activity during 2011 focused on the mathematical analysis of traffic flow models on road networks or subject to unilateral constraints. In particular, [34] is devoted to a hyperbolic 2nd order model for traffic flow with local flux constraint. We describe two admissible Riemann solvers and we construct ad hoc finite volume numerical schemes to compute these solutions. The paper [59] is devoted to the study of a traffic flow model on a network composed by an arbitrary number of incoming and outgoing arcs connected together by a node with a buffer. We define the solution to the Riemann problem at the node and we prove existence and well posedness of solutions to the Cauchy problem. Finally, a general traffic flow model with phase-transition is proposed and described in [28].

M.L. Delle Monache just started her doctoral thesis in the same topic. More precisely she will study hyperbolic models of traffic flow and associated optimization problems.

6.1.2. Crowd motion

Participants: Nora Aïssiouene, Régis Duvigneau, Nader El Khatib, Jihed Joobeur, Paola Goatin, Massimiliano D. Rosini [ICM, Warsaw University, Poland].

Concerning pedestrian motion modeling, we are interested in the optimization of facilities design, in order to maximize pedestrian flow and avoid or limit accidents due to panic situations. To this aim, we are now studying a macroscopic model for crowd movements consisting in a scalar conservation law accounting for mass conservation coupled with an Eikonal equation giving the flux direction depending on the density distribution. From the theoretical point of view, and as a first step, we are studying the problem in one space dimension (for applications, this case corresponds to a crowd moving in a corridor). In collaboration with M. Rosini (supported by the project CROM3, funded by the PHC Polonium 2011), we have established entropy conditions to select physically relevant solutions, and we have constructed explicit solutions for some simple initial data (these results are presented in [54]). We are now studying existence and uniqueness of solutions of the corresponding initial boundary value problem. From the numerical point of view, we are implementing the model in two space dimensions on triangular meshes on the Num3sis platform. This was partly done by N. El-Khatib (postdoc at INRIA from January to August 2011), and will be completed soon by Nora Aïssiouene. This will provide a performing numerical tool to solve the related optimization problems arising in the optimization of facilities design, such as the position and size of an obstacle in front of (before) a building exit in order to maximize the outflow through the door and avoid or limit over-compression. Moreover, jointly with the PULSAR team, we have supervised J. Joobeur’s internship, which was devoted to pedestrian data collection from real-word video recordings (Turin metro station). The density data will serve to validate the model.

The above researches were partially funded by the ERC Starting Grant "TRAM3 - Traffic management by macroscopic models".

6.2. Optimum design in fluid dynamics and its couplings

In computational sciences for physics and engineering, Computational Fluid Dynamics (CFD) are playing one of the major roles in the scientific community to foster innovative developments of numerical methodologies. Very naturally, our expertise in compressible CFD has led us to give our research on numerical strategies for optimum design a particular, but not exclusive focus on fluids.
6.2.1. Cooperation and competition in multidisciplinary optimization

Participants: Étienne Baratchart [ENSEIBB MATMÉCA], Jean-Antoine Désidéri, Régis Duvigneau, Adrien Zerbinati.

The framework of our research aims to contribute to numerical strategies for PDE-constrained multiobjective optimization, with a particular emphasis on CPU-demanding computational applications in which the different criteria to be minimized (or reduced) originate from different physical disciplines that share the same set of design variables. These disciplines are often fluids, as a primary focus, coupled with some other discipline, such as structural mechanics.

Our approach to competitive optimization is based on a particular construction of Nash games, relying on a split of territory in the assignment of individual strategies. A methodology has been proposed for the treatment of two-discipline optimization problems in which one discipline, the primary discipline, is preponderant, or fragile. Then, it is recommended to identify, in a first step, the optimum of this discipline alone using the whole set of design variables. Then, an orthogonal basis is constructed based on the evaluation at convergence of the Hessian matrix of the primary criterion and constraint gradients. This basis is used to split the working design space into two supplementary subspaces to be assigned, in a second step, to two virtual players in competition in an adapted Nash game, devised to reduce a secondary criterion while causing the least degradation to the first. The formulation has been proved to potentially provide a set of Nash equilibrium solutions originating from the original single-discipline optimum point by smooth continuation, thus introducing competition gradually. This approach has been demonstrated over a testcase of aero-structural aircraft wing shape optimization, in which the eigen-split-based optimization reveals clearly superior [33].

While the two-discipline method is currently being applied to various complex physical multiobjective situations (see in particular 6.2.2, 6.2.6, 6.2.7, 6.2.8), the method has been extended to situations involving more than two objectives when the initial point is Pareto-optimal. Then, a particular convex combination of the criteria is locally stationary, and the two-discipline strategy can be applied using this combination as preponderant criterion, and a particular other criterion as secondary one. Whence, the proposed split of territory produces a continuum of Nash equilibrium points tangent to the Pareto set. This theoretical result has been illustrated in the context of a simpler numerical experiment by E. Baratchart during his internship [53], see Fig. 3.

![Figure 3. Combination of cooperative and competitive optimization algorithms: in red the Pareto set, in blue MGDA steps directed to the Pareto set, in green steps by Nash games with split of territory tangent to the Pareto set.](image-url)
Our approach to cooperative optimization is based on a result of convex analysis established for a general unconstrained multiobjective problem in which all the gradients are assumed to be known. The theorem [58] states that in the convex hull of the gradients, there exists a unique vector of minimal norm, \( \omega \); if it is nonzero, the vector \( \omega \) is a descent direction common to all criteria; otherwise, the current design point is Pareto-optimal. This result led us to generalize the classical steepest-descent algorithm by using the vector \( \omega \) as search direction. We refer to the new algorithm as the multiple-gradient descent algorithm (MGDA). The MGDA yields to a point on the Pareto set, at which a competitive optimization phase can possibly be launched on the basis of the local eigenstructure of the different Hessian matrices. This general formulation fosters several connected studies detailed in 6.2.3.

6.2.2. Virtual games for coupling global to local shape optimization

Participant: Régis Duvigneau.

In several engineering problems, the system to optimize is characterized by some parameters that define global shape properties, while remaining parameters define local shape modifications. Of course, these two sets of parameters do not play the same role and have not the same impact on the cost functional value. Therefore, we are studying how to construct an efficient optimization strategy that takes benefit of this global / local splitting of parameters.

A typical aerodynamic shape optimization problem has been studied, that consists of a lift-constrained drag minimization for a transonic wing, whose sections are defined by two B-Spline curves whereas global shape characteristics are defined by five parameters (span, roottip length ratio, angle of attack, twist angle, sweep angle). It has been found that the naive simultaneous optimization of all parameters failed, due to the multimodality of the problem. Alternatively, the use of a virtual game strategy, based on a splitting between the local and global parameters, yields a satisfactory result for a moderate cost [47].

6.2.3. Multiple-Gradient Descent Algorithm (MGDA)

Participants: Jean-Antoine Désidéri, Régis Duvigneau, Adrien Zerbinati.

6.2.3.1. Basic experiments and validation

In multi-objective optimization, the knowledge of the Pareto set provides valuable information on the reachable optimal performance. A number of evolutionary strategies (PAES, NSGA-II, etc), have been proposed in the literature and proved to be successful to identify the Pareto set. However, these derivative-free algorithms are very demanding in terms of computational time. Today, in many areas of computational sciences, codes are developed that include the calculation of the gradient, cautiously validated and calibrated.

In [50], MGDA has been tested over a number of classical multiobjective-optimization testcases, and found successful to converge to Pareto-optimal solutions in situations of either convex or concave Pareto sets. Additionally, MGDA and PAES [61] were found to have complementary merits, making a hybrid method promising.

6.2.3.2. Metamodel-supported CFD optimization by MGDA

Using MGDA in a multi objective optimization problem requires the evaluation of a large number of points with regard to criteria, and their gradients. In the particular case of a CFD problems, each point evaluation is very costly since it involves a flow computation, possibly the solution of an adjoint-equation. To alleviate this difficulty, we have proposed to construct metamodels of the functionals of interest (lift, drag, etc) and to calculate approximate gradients by local finite differences. These metamodels are updated throughout the convergence process to the evaluation of the new design points by the high-fidelity model, here the 3D compressible Euler equations.

This variant of MGDA has been tested successfully over a problem of external aerodynamic optimum-shape design of an aircraft wing consisting of reducing wave-drag, and augmenting lift. After only a few cycles of database updates, the Pareto front visibly forms, and this result is achieved at a very moderate computational cost. This variant is currently being tested and extended to an internal flow optimization problem related to an automobile air-conditioning system and governed by the Navier-Stokes equations. This more difficult problem has been proposed by Renault within the OMD2 ANR project.
6.2.3.3. MGDA in functional setting

One aspect of the theoretical result concerning the minimal-norm element $\omega$ is that, regardless the possibly-functional setting of the problem in case of a distributed system, the descent-direction $\omega$ is identified in the standard $n$-dimensional vector space $\mathbb{R}^n$ ($n$: the number of objective functions).

This observation has led to examine the application of MGDA in the functional setting of domain-decomposition methods (DDM) in which a functional criterion and a functional control can be defined at each interface independently permitting to formulate the DDM problem as a multi-objective optimization. On-going research in this area is related to the necessary preconditioning, or normalization procedure, of the gradients.

6.2.4. Flow control

Participants: Régis Duvigneau, Jérémie Labroquère.

Shape optimization methods are not efficient to improve the performance of fluid systems, when the flow is characterized by a strong unsteadiness related to a massive flow separation. This is typically the case for the flow around an automotive body or a wing in stall condition. To overcome this difficulty, flow control strategies are developed, that aim at manipulating vortex dynamics by introducing some active actuators, such as periodic blowing/suction jets. In this context, the choice of the control parameters (location, amplitude, frequency) is critical and not straightforward. Therefore, a numerical study is conducted to i) improve the understanding of controlled flows ii) develop a methodology to determine optimal control parameters by coupling the controlled flow simulation with optimization algorithms. Two research axes have been considered:

- the solution of the unsteady sensitivity equations derived from the state equations, to exhibit the dependency of the flow dynamics with respect to the control;
- the optimization of control parameters using a statistical metamodel-based strategy. First results show the efficiency of such an approach for laminar flow problems [31], [44].

6.2.5. Optimum shape design in aerodynamics by the adjoint method

Participants: Manuel Bompard, Sébastien Bourasseau, Jean-Antoine Désideri, Jacques Peter [Research Engineer, ONERA/DSNA].

At ONERA, compressible flow simulations governed by the Euler or Navier Stokes (RANS) equations are conducted with the software elsA [57] that admits both structured and unstructured-grid formulations. Local aerodynamic optimizations are made with a version that includes the calculation of the shape gradient via the solution of an adjoint equation. The discrete adjoint is calculated formally step-by-step to include the various derivative terms involved, and is being enhanced gradually to account for more complex models. In particular, for RANS computations, this gradient today includes the differentiation of the turbulence model.

6.2.5.1. Metamodels including derivative information

In this context, to alleviate the cost of an optimum-shape design in aerodynamics, M. Bompard in his thesis [26], has examined how metamodels, firstly based on functional values only, could be used to determine shortcuts in the convergence process. Second, when the gradient w.r.t. the design parameters is known, the gradients of functionals of interest, that is, most commonly, aerodynamic coefficients, are calculated. Thus, these derivative informations can also used to construct more elaborate metamodels. Such constructions have also been studied systematically and used efficiently in global optimizations [37]; in particular co-Kriging and Support-Vector Regression, for which a technique to adjust automatically the free parameters has been proposed based on a simplification of the leave-one-out test.

6.2.5.2. Parameterization-free local optimization

When the derivatives of the functionals w.r.t. the volume geometry, $dJ/dX$, have been calculated, it is also possible to calculate the gradient w.r.t. surface coordinates, $dJ/DS$. Since the surface deformation steers the entire mesh movement, often through analytical dependencies, M. Bompard [26] has also examined how could $dJ/DS$ be used directly in a local aerodynamic optimization. However, it is well-known that the distribution of $dJ/DS$ is very irregular, and its usage in the optimization loop necessitates that adequate smoothing procedures be elaborated. Partial success was achieved in this area, still subject to research.
6.2.6. Aero-structural optimization

Participants: Gérald Carrier [Research Engineer, ONERA/DAAP], Jean-Antoine Désideri, Imane Ghazlane.

In industry, aircraft wings are designed by accounting for several multidisciplinary couplings. Certainly of greatest importance is the coupling, or concurrency, between aerodynamic optimization and structural design. At ONERA, in the former thesis of M. Marcelet, the aerodynamic gradient has been extended to account for (the main terms of) static fluid-structure interaction, commonly referred to as the “aeroelastic gradient”.

In her thesis, I. Ghazlane has extended M. Marcelet’s work to take into account, in the aeroelastic gradient, the terms originating from the differentiation of the wing-structural model. In this development, the wing structure is treated as an equivalent Euler-Bernoulli beam. These formal extensions have been validated by an extensive experimentation. Additionally, special post-processing procedures are applied to evaluate accurately the various physical contributions to drag. As a result, the numerical tools necessary to conduct a very realistic aircraft wing optimization are now set up and are being exploited [ 38 ]. It is also envisaged to conduct a two-objective optimization (drag and mass reduction) via a Nash game using our optimization platform FAMOSA.

6.2.7. Sonic boom reduction

Participants: Gérald Carrier [Research Engineer, ONERA/DAAP], Jean-Antoine Désideri, Andrea Minelli, Itham Salah El Din [Research Engineer, ONERA/DAAP].

When an aircraft flies at supersonic speed, it generates at ground level an N-shaped shock structure which can cause serious environmental damage (“sonic boom”). Thus a problem of interest in aerodynamic optimization is to design such an aircraft to reduce the intensity of the sonic boom while maintaining the aerodynamic performance (drag minimization under lift constraint). Andrea Minelli’s aimed at contributing to this two-discipline optimization problem. In the first part of his work, an inverse problem has been formulated and solved for “shaped sonic boom” and found in excellent agreement with the George-Seebass-Darden theory [ 60 ] for the calculation of the Whitham function corresponding to the lowest-boom (axisymmetric) shape. The method is currently being extended to account for more general geometries. Besides, aero-acoustic optimizations have been realized successfully by coupling the aerodynamic optimizer (based on Euler calculations by the elsA sofware) with the sonic-boom computation in a Nash game formulation. These experiments, conducted with our optimization platform FAMOSA, have demonstrated that starting from the shape optimized aerodynamically, one could retrieve smoothly a shape corresponding to nearly-optimal sonic-boom reduction.

6.2.8. Helicopter rotor blade optimization in both situations of hovering and forward flight

Participants: Michel Costes [Research Engineer, ONERA/DAAP], Jean-Antoine Désideri, Arnaud Le Pape [Research Engineer, ONERA/DAAP], Enric Roca Leon.

E. Roca Leon has recently started at ONERA a CIFRE thesis supported by EUROCOPTER, Marignane. This thesis follows the doctoral thesis of A. Dumont in which the adjoint-equation approach was used to optimize a rotor blade in hovering flight. The goal of this new thesis is to solve a two-objective optimization problem in which the hovering-flight criterion is considered preponderant, but a new criterion that takes into account the forward-flight situation is also introduced, concurrently. The thesis work includes the set up of a hierarchy of models from low to high fidelity, in order to calibrate appropriate functional criteria. Secondly, our Nash game approach to competitive optimization will be implemented, using our optimization platform FAMOSA, and comparisons with the results by A. Dumont will be made.

6.2.9. Optimum design in naval hydrodynamics

Participants: Régis Duvigneau, Louis Blanchard.

Naval hydrodynamics field has recently shown a growing interest for optimum design methods. The computational context is especially complex because it implies unsteady two-phase turbulent flows, with possibly very high Reynolds number (up to $10^9$). The use of automated design optimization methods for such problems requires new developments to take into account the large CPU time necessary for each simulation and the specificity of the geometries considered.
In collaboration with GALAAD Project-Team, some developments have been initiated on the geometrical modeling of hull shapes by parametric surfaces. The objective is to be able to modify existing hull shapes by controlling a small number of parameters, that are meaningful for naval architects. Two testcases are considered: the bow shape for trawler ships (see Fig. 4) and the whole hull shape for canoes, in collaboration with the Fédération francaise de Canoe-Kayak.

Figure 4. Initial shape (left) and deformed shape to generate a bow (right) for the trawler ship, to reach two line targets (in red).

6.3. Optimum design in structural mechanics

6.3.1. Shape Optimization in Multidisciplinary Non-Linear Mechanics

Participants: Aalae Benki, Jean-Antoine Désidéri, Abderrahmane Habbal.

In collaboration with the ArcelorMittal’s Center for Research in Automotive and Applications, we study the multidisciplinary shape and parameter design of highly non linear mechanical 2D and 3D structures. We have developed methods adapted to the approximation of Pareto Fronts such as Normal Boundary Intersection NBI and Normalized Normal Constraint Method NNCM. Due to the time consuming cost evaluation, the use of cheap to evaluate surrogate models is mandatory. We have studied the consistency of the approach NBI or NNCM plus surrogates, which turned out to be successful for a broad panel of standard mathematical benchmarks. The application of this approach for the case of beverage cans which undergo elastoplastic deformation under high pressure is ongoing.

6.3.2. Optimization of Addendum Surfaces in Stamping

Participants: Fatima Zahra Oujebbour, Jean-Antoine Désidéri, Abderrahmane Habbal.

Within the OASIS Consortium (ArcelorMittal, ErDF, INRIA, UTC, EURODECISION, ESILV, NECS, Delta-CAD, SCILAB-DIGITEO), Opale Project leads the Optimization task. Our aim is to develop decentralized decision-making algorithms dedicated to find efficient solutions (Pareto optimal) in a complex multidisciplinary framework (forming, stamping, welding non-linear processes, spring-back, vibration, in-function linear processes, crash and fatigue non linear and non differentiable processes) for several (between three and five) criteria. An important difficulty when trying to identify the Pareto Front, even when using adapted methods such the Normal Boundary Intersection, is that the criteria involved (thanks to the high nonlinearity in the mechanical models) exhibit many local optima. So one must use global optimization methods. We have studied the hybrid approach Simulated Annealing with Simultaneous Perturbation SASP for a suite of mathematical test-cases. To envisage the application of our method to the complex CPU time consuming stamping process, we lead an intermediate phase dedicated to the validation of the SASP method for the minimization of the spring-back that follows the stamping of a metal sheet, the design variable being the thickness distribution.
6.4. Application of shape and topology design to biology and medicine

6.4.1. Mathematical modeling of dorsal closure DC

Participants: Abderrahmane Habbal, Luis Almeida [University of Nice-Sophia Antipolis], Patrizia Bagnerini [Genova University], Fanny Serman [University of Nice-Sophia Antipolis], Stéphane Noselli [University of Nice-Sophia Antipolis], Glenn Edwards [Duke University].

A mathematical model for simulation of actin cable contraction, during wound closure for Drosophila embryo, which contains an extra term in addition to the curvature flow is developed. The basic mathematical model introduced and validated in [27] is extended in order to include the non-homogeneous wound healing or non-homogeneous dorsal closure [52].

6.5. Particular applications of simulation methods

6.5.1. Analysis of a two-level parameterization optimization for antenna design

Participants: Benoît Chaigne [Doctoral student, 2007-2010], Jean-Antoine Désideri.

Similar to the discretization of ordinary or partial differential equations, the numerical approximation of the solution of an optimization problem is possibly subject to numerical stiffness. In the framework of parametric shape optimization, hierarchical representations of the shape can be used for preconditioning, following the idea of Multigrid (MG) methods. By analogy with the Poisson equation, which is the typical example for linear MG methods, we have addressed a parametric shape inverse problem. The ideal cycle of a two-level algorithm can be defined and adapted to shape optimization problems that require appropriate transfer operators. With the help of a symbolic calculus software we have shown that the efficiency of an optimization MG-like strategy is ensured by a small dimension-independent convergence rate. Numerical examples are worked out and corroborate the theoretical results. Applications to antenna design have been realized. Finally, some connections with the direct and inverse Broyden-Fletcher-Goldfarb-Shanno preconditioning methods have been shown [29].

6.5.2. Mesh qualification

Participants: Jean-Antoine Désideri, Maxime Nguyen, Jacques Peter [Research Engineer, ONERA/DSNA].

M. Nguyen Dinh is conducting a CIFRE thesis at ONERA supported by AIRBUS France. The thesis topic is the qualification of CFD simulations by anisotropic mesh adaption. Methods for refining the 2D or 3D structured mesh by node movement have been examined closely. Secondly, it is investigated how could the local information on the functional gradient $\|dJ/dX\|$ be exploited in a multi-block mesh context. This raises particular questions related to conservation at the interfaces.

6.5.3. Hybrid meshes

Participants: Sébastien Bourasseau, Jean-Antoine Désideri, Jacques Peter [Research Engineer, ONERA/DSNA], Pierre Trontin [Research Engineer, ONERA/DSNA].

S. Bourasseau has started a CIFRE thesis at ONERA supported by SNECMA. The thesis is on mesh adaption in the context of hybrid meshes, that is, made of both structured and unstructured regions. Again, the aim is to exploit at best the function gradient provided by the adjoint-equation approach. Preliminary experiments have been conducted on geometries of stator blade yielding the sensitivities to global shape parameters.

6.5.4. Nash game approach to image processing

Participants: Abderrahmane Habbal, Rajae Aboulaich [Mohamed V University of Rabat], Maher Moakher [University of Tunis], Moez Kallel [University of Tunis], Anis Theljani [University of Tunis].
We have started in 2011 to study the application of game modeling to image processing problems. We propose an original game theory approach to simultaneously restore and segment noisy images [56]. We define two players: one is restoration, with the image intensity as strategy, and the other is segmentation with contours as strategy. Cost functions are the classical relevant ones for restoration and segmentation, respectively. The two players play a static game with complete information, and we consider as solution to the game the so-called Nash Equilibrium. For the computation of this equilibrium we present an iterative method with relaxation. The results of numerical experiments performed on some real images show the relevance and efficiency of the proposed algorithm. Based on a similar idea, we formulated well known data completion (Cauchy) problems for Laplace equation as Nash games [55] and obtained results of existence, uniqueness and stability of a Nash equilibrium which turns out to be the Cauchy solution when the Cauchy data are compatible. With A. Theljani, we study the extension of the Nash data completion approach to nonlinear parabolic equations with application to image inpainting.

### 6.6. Isogeometric analysis and design

**Participants:** Louis Blanchard, Régis Duvigneau, Bernard Mourrain [Galaad Project-Team], Gang Xu [Galaad Project-Team].

Design optimization stands at the crossroad of different scientific fields (and related software): Computer-Aided Design (CAD), Computational Fluid Dynamics (CFD) or Computational Structural Dynamics (CSM), parametric optimization. However, these different fields are usually not based on the same geometrical representations. CAD software relies on Splines or NURBS representations, CFD and CSM software uses grid-based geometric descriptions (structured or unstructured), optimization algorithms handle specific shape parameters. Therefore, in conventional approaches, several information transfers occur during the design phase, yielding approximations that can significantly deteriorate the overall efficiency of the design optimization procedure. Moreover, software coupling is often cumbersome in this context.

The isogeometric approach proposes to definitely overcome this difficulty by using CAD standards as a unique representation for all disciplines. The isogeometric analysis consists in developing methods that use NURBS representations for all design tasks:

- the geometry is defined by NURBS surfaces;
- the computation domain is defined by NURBS volumes instead of meshes;
- the solution fields are obtained by using a finite-element approach that uses NURBS basis functions;
- the optimizer controls directly NURBS control points.

Using such a unique data structure allows to compute the solution on the exact geometry (not a discretized geometry), obtain a more accurate solution (high-order approximation), reduce spurious numerical sources of noise that deteriorate convergence, avoid data transfers between the software. Moreover, NURBS representations are naturally hierarchical and allows to define multi-level algorithms for solvers as well as optimizers. In this context, some research axes have been developed in collaboration with GALAAD Project-Team:

- Methods for adaptive parameterization including a posteriori error estimate for elliptic problems [36], [35], [42];
- Numerical schemes based on Spline functions for 2D inviscid compressible flow simulations;
- Optimization methods for structural elasticity, based on shape-gradient concept, and fluid-structure interactions [48] (in collaboration with Technical University of Munich).

### 6.7. Resilient workflows for distributed multidiscipline optimization

**Participants:** Toan Nguyen, Laurentiu Trifan.
A distributed platform based on the YAWL workflow management system has been designed and implemented to deploy HPC applications on the Grid5000 network infrastructure. The goal is to provide a generic environment for the design of complex applications that require HPC resources for large-scale fault-tolerant applications, see Fig. 2 and [39].

The platform provides application-level fault-tolerance, i.e., resilience, in order to restart the workflow execution whenever abnormal behavior or system-level errors occur. This allows a variety of errors to be taken into account, ranging from execution time-outs to out-of-bounds parameter values to be managed, with the help of user intervention when necessary [40].

The error management procedure uses exception handlers in YAWL to trigger the appropriate corrective actions, which are defined by rules invoking the adequate compensating workflows. Once defined, this can be made transparent to the users [41].

An original scheme based on asymmetric checkpoints has been designed in order to reduce overhead in both checkpointing and application restarts. It minimizes the number of required checkpoints created based on default rules and user-specific needs.

The platform is currently developed in Java on Linux workstations and should be portable on Windows and MacOS, although this has not been tested yet.

Examples are deployed on the Grid5000 national network infrastructure using the OMD2 test-cases (e.g., vehicle air-conditioner pipe optimization). The goal is here to provide a demonstrator platform that deploys large-scale optimization applications involving several (typically over five) HPC clusters distributed on the Grid5000 network. The coarse-grain definition of the application is defined by a workflow that monitors the distributed execution of the parallel component codes on the various clusters, providing resilience capabilities in case of system and application errors, see Fig. 5.

![Figure 5. Application definition using YAWL.](image)
5. New Results

5.1. Numerical methods

5.1.1. Finite volume methods in curvilinear coordinates

Participants: Hervé Guillard, Boniface Nkonga, Afeintou Sangam.

Finite volume methods are specialized numerical techniques for the solving of divergence equations in strong conservation forms of the form

\[ \frac{\partial S}{\partial t} + \text{div} T = 0 \]  \hspace{1cm} (1)

where \( S \) is a scalar or a vector while \( T \) is a vector or a second-order tensor. Using textbook formulas for the expression of the divergence operator in curvilinear coordinates, the use of these coordinate systems instead of the Cartesian one can lead to a loss of the strong conservation form of the equations and introduce a source term in (1). Actually, this is unnecessary and one can show that whatever the system of curvilinear coordinate used, there exists a strong conservation law form of the system. However, when vector equations have to be considered (that is if \( S \) is a vector and \( T \) a tensor), it is necessary to extract from (1) scalar equations for the components of the vector \( S \) and this may destroy the strong conservation form of the equation. Following the work done in [12] where a general method (i.e that does not depend on the curvilinear system used) based on the projection of the discretized vector system have been designed, we have studied this year its application to cylindrical coordinates in the case where the geometry is a torus. This approach is robust and accurate for problems that take place for instance inside tokamak devices for magnetic confinement fusion or in toroidal plasmas occurring in stars and galaxies to take another examples. The method is now implemented in the PlaTo software.

5.1.2. Entropy preserving schemes for conservation laws

Participants: Christophe Berthon [University of Nantes], Bruno Dubroca [CEA/DAM/CESTA and University of Bordeaux 1], Afeintou Sangam.

Entropy preserving schemes for conservation laws

In collaboration with C. Berthon of University of Nantes, and B. Dubroca of CEA/DAM/CESTA and University of Bordeaux 1, we have established a new technique that proves discrete entropy inequalities of finite volume methods to approximate conservation laws. This technique is free of additional numerical models such as kinetic and relaxation schemes. Moreover, our results leads to a full class of entropy preserving schemes for general Euler equations [11]. This proposed thechnique has been successfully applied to two intermediates states scheme for 10-moments equations with laser source-term in context of Inertial Fusion Confinement. Moreover, the derived procedure is now extended to Saint-Venant model.

5.1.3. Mesh adaptation Methods

Participants: Anca Belme [Projet Tropics], Hubert Alcin [Projet Tropics], Alain Dervieux, Frédéric Alauzet [Projet Gamma, INRIA-Rocquencourt].

This activity results from a cooperation between Gamma, Tropics, Pumas, and Lemma company. See details in Tropics and Gamma activity reports. Its concerns Pumas’s subject through the current applications of mesh adaptation to flows with interfaces and the starting application of mesh adaptation to Large Eddy Simulation. It is also planned to use mesh adaptation for simplified plasma models in the context of ANEMOS ANR project.
5.1.4. Parallel CFD algorithms

**Participants:** Hubert Alcin [Tropics], Olivier Allain [Lemma], Anca Belme [Tropics], Marianna Braza [IMF-Toulouse], Alexandre Carabias [Tropics], Alain Dervieux, Bruno Koobus [Université Montpellier 2], Carine Moussaed [Université Montpellier 2], Hilde Ouvrard [IMF-Toulouse], Stephen Wornom [Lemma].

Pumas is associated to the ANR ECINADS project started in end of 2009, devoted to the design of new solution algorithms for unsteady compressible flows, adapted to scalable parallelism and to reverse (adjoint) Automatic Differentiation. See in the activity report of Tropics. The newer two-level deflation algorithm is currently applied to a simplified plasma model in the context of ANEMOS ANR.

5.2. Plasma physics

5.2.1. Analysis of the drift approximation

**Participants:** Hervé Guillard, Afeintou Sangam, Philippe Ghendrih [IRFM, CEA Cadarache], Yanick Sarazin [IRFM, CEA Cadarache], Patrick Tamin [IRFM, CEA Cadarache].

Drift approximation consider the slow evolution of the fields in the vicinity of a tokamak equilibrium. These models are typically used to study the micro-instabilities that are believed to be responsible of turbulent transport in tokamaks. Since the drift asymptotic uses a “slow” scaling of the velocity field, the resulting models are significantly different from the full MDH models. This is particularly true with respect to the computation of the electric field that is given by an Ohm’s law in MHD models whereas it is computed by a vorticity-like evolution equation in drift approximations. Drift asymptotic models are extremely interesting from a computational point of view since they save substantial CPU time and computer memory. However, the mathematical and numerical properties of these models are essentially unknown. We have begun a detailed study of the derivation of these models from two-fluid Braginskii-type models in order to establish the range of applicability of these asymptotic models, understand their mathematical properties and relations with the reduced MHD models and design appropriate numerical methods for their approximations.

5.2.2. Stabilized C1-Finite Element Method for MHD

**Participants:** Boniface Nkonga, Marie Martin.

Reduced MHD models are often used in plasma physics and therefore fast compressible waves are not taken into account. In the context of Elms instabilities investigations, full and extended MHD models are to be considered within the framework of high order finite element approximation. In order to obtain predictive simulations with reduced, full and extended MHD models, it is crucial to design numerical strategies that can face some difficulties related to the use of the classical Galerkin methods for convection dominated flows. We have developed a general VMS stabilization strategy for time dependent implicit scheme, which can be applied to MHD models in order to preserved the global accuracy of the initial Galerkin formulation and enforce physical properties as monotony and positivity. Higher order of continuity shape functions are important for accuracy and also help to obtain more robustness of the stabilization as there is no more singularities on elements edges. Numerical implementation and preliminary validations has been performed using C1-Bell shapes functions for triangular meshes. Order five accuracy of the theory is recovered for specific boundary condition. Improvement for general boundary condition is still an open issue. The next step is to combine this Bell shapes functions in the poloidal direction with B-splines functions to achieve accurate representation of complex torus as ITER.

5.2.3. Two fluid modelling of the Scrape-Off-Layer Plasma

**Participants:** Audrey Bonnemént, Hervé Guillard, Richard Pasquetti.
A two fluid physical model has been developed in close collaboration with researchers from IRFM. It is based on an hypothesis of stationary magnetic field and the electrostatic and electroneutrality assumptions. However the usual drift assumption, e.g. used in the CEA code TOKAM3D (thesis of P. Tamain), is not used. On the basis of the conservation equations of density, electron and ion velocities, electron and ion temperatures and electrical charges, a set of 10 nonlinear coupled partial differential equations (PDE) can be set up. Our investigations rely on the development of two solvers for this set of PDE. In the frame of her thesis, A. Bonnement (co-direction H. Guillard and R. Pasquetti, financial support of INRIA and PACA region, industrial partner ASSYSTEM) uses a Finite volume / element (FV/FE) approach. The other code, developed by R. Pasquetti, focuses on the use of high order approximations: a spectral element method (SEM) is used in the poloidal plane whereas Fourier expansions are used in the toroidal direction. Each of these codes has allowed in 2010 to solve strongly anisotropic diffusion equations in 2D and axisymmetric 3D geometries for the FV/FE code and in 2D and fully 3D for the Fourier-SEM code. The FV/FE code has also been used to carry out a study of radiative layers evolution in a 2D annular configuration but also in a realistic 3D ITER configuration.

Works carried out during the year 2011 are described hereafter: – On the basis of a Godunov scheme, the FV/FEd approach has been extended to solve the axisymmetric Euler, Navier-Stokes or Braginskii like systems in the jet-tokamak geometry. With respect to Navier-Stokes, the Braginskii system is characterized by anisotropies in the transport coefficients. The mesh is unstructured: A finite element approximation is used for the diffusion terms whereas for the convective terms a finite volume approximation is used on the dual mesh. We have especially focused on the treatment of the toroidal geometry. We have also focused on the implementation of the so-called Bohm conditions, which are enforced by imposing that at the limiter the fluid velocity is colinear to the magnetic field and that the « parallel Mach number » equals or is greater than one. In these studies the governing equations are completed by constant force terms in order to model the Lorentz forces as well as sources of mass and energy. Such forcing terms allow to preserve an equilibrium state, e.g. obtained on the basis of simulations that make use of the drift assumption (see next paragraph). We have then introduced perturbations of this equilibrium state to study the evolution of the different variables (density, velocity and temperature). Various perturbations have been used to this end, fully random in order to check the stability of the equilibrium or on the contrary localized, in order to roughly model the injection of pellets inside the SOL. It is planned that Audrey Bonnement will defend her thesis in Spring 2012.

The SEM-Fourier 3D code has been extended to solve the full set of governing equations. The unknowns are then the density, the velocity of the center of mass, the electric current and potential, the ion and electron internal energies. In time, this set of PDE is solved by using an IMEX (Implicit – Explicit) approach, based on the combination of an explicit Runge-Kutta scheme for the flux terms and on a DIRK (diagonal implicit Runge Kutta) for the Lorentz terms, which indeed lead to an unconditionally unstable scheme if treated explicitly. A projection method is used to enforce the divergence free constraint of the current, so that an additional solve of a Poisson equation is required to obtain the potential. The Bohm conditions are here implemented by enforcing that the ion pressure is such that the parallel Mach number shows the expected value, i.e. equal or greater than one. The main difficulty that we presently meet is that the initial condition that we use does not correspond to an equilibrium state, so that instabilities quickly develop till yielding an unsteady flow not consistent at the limiter with the Bohm condition. Detailed analyses of these results are presently carried out with Sebastian Minjeaud (new CNRS researcher of the LJAD) to provide relevant explanations of the observed phenomena.

5.2.4. Drift approximation modelling of the Scrape-Off-Layer Plasma

Participants: Marco Bilanceri, Hervé Guillard.

Based on a fluid model using the drift velocity approximation, a simulation method have been designed to compute the flow in the scrape-off-layer of a Tokamak. The variables used by the model are the particle number, the parallel (to the magnetic field) velocity and the electric potential. The spatial approximation uses a finite volume/finite element approach and is therefore easy to apply to complex geometries. Bohms boundary conditions are used on the divertor plates of the machine. The figure 1. shows the density in a poloidal cut where the influence of the separatrix can be clearly seen.
Figure 1. Density plot in a poloidal section of a jet-like Tokamak.
5.3. Fluid Turbulence

5.3.1. Hybrid RANS-LES models

Participants: Anca Belme [Tropics], Alain Dervieux, Bruno Koobus [University of Montpellier 2], Carine Moussaed [University of Montpellier 2], Hilde Ouvrard [IMF-Toulouse], Maria-Vittoria Salvetti [University of Pisa], Stephen Wornom [Lemma].

The purpose of our works in hybrid RANS/LES is to develop new approaches for industrial applications of LES-based analyses. In the foreseen applications (aeronautics, hydraulics), the Reynolds number can be as high as several tenth millions, a far too large number for pure LES models. However, certain regions in the flow can be much better predicted with LES than with usual statistical RANS (Reynolds averaged Navier-Stokes) models. These are mainly vortical separated regions as assumed in one of the most popular hybrid model, the hybrid Detached Eddy Simulation model. Here, “hybrid” means that a blending is applied between LES and RANS. The french-italian team has designed a novel type of hybrid model. This year, the new model has been adapted to very high Reynolds number. Our benchmark is the flow past a circular cylinder, an ECINADS test case. Reynolds number as high as 3 Millions could be passed with good prediction of main properties like mean drag, root mean square of lift fluctuation, base pressure.

5.3.2. Acoustics

Participants: Anca Belme [Tropics], ILya Abalakin [IMM-Moscou], Alain Dervieux [Tropics], Alexandre Carabias.

A method for the simulation of aeroacoustics on the basis of these models has been designed and developed by a cooperation between the Computational Aeroacoustics Laboratory (CAL) of Intitute for Mathematical Modeling at Moscow and INRIA. Further applications has been developed by the Russian team from the two common numerical scheme, the Mixed-Element-Volume at sixth-order, and the quadratic reconstruction scheme. This year the cooperation is concentrated on the study by Alexandre Carabias of a new quadratic reconstruction scheme, which extends the one developed by Hilde Ouvrard and Ilya Abalakin. A second research topic was the calculation of acoustic propagation with unsteady mesh adaptation.

5.4. Environmental flows

5.4.1. Mobile bed and sediment transport

Participants: Hervé Guillard, Boniface Nkonga, Marco Bilanceri, Maria-Vittoria Salvetti [University of Pisa, Italy], Imad Elmahi [University of Oudja, Morocco].

The numerical approximation of a model coupling the shallow-water equations with a sediment transport equation for the morphodynamics have been studied. In shallow-water problems, time advancing can be carried out by explicit schemes. However, if the interaction with the mobile bed is weak, the characteristic time scales of the flow and of the sediment transport can be very different introducing time stiffness in the global problem. For this case, it is of great interest to use implicit schemes. The time integration stategy that we have devised is based on a defect-correction approach and on a time linearization, in which the flux Jacobians are computed through automatic differentiation. The aim of the present work is to investigate the behaviour of this time scheme in different situations related to environmental flows. This work has been presented in [18] and [19]
5. New Results

5.1. Mathematical Modelling

5.1.1. Mathematical and numerical model for nonlinear viscoplasticity

Participants: Nicolas Favrie, Sergey Gavrilyuk.

A macroscopic model describing elastic plastic solids is derived in a special case of the internal specific energy taken in separable form: it is the sum of a hydrodynamic part depending only on the density and entropy, and a shear part depending on other invariants of the Finger tensor. In particular, the relaxation terms are constructed compatible with the von Mises yield criteria. In addition, Maxwell-type material behaviour is shown up: the deviatoric part of the stress tensor decays during plastic deformations. Numerical examples show the ability of this model to deal with real physical phenomena [15].

5.1.2. A discrete model for compressible flows in heterogeneous media

Participants: Olivier Le Métayer, Alexandre Massol, Nicolas Favrie, Sarah Hank.

This work deals with the building of a discrete model able to describe and to predict the evolution of complex gas flows in heterogeneous media. In many physical applications, large scales numerical simulation is no longer possible because of a lack of computing resources. Indeed the medium topology may be complex due to the presence of many obstacles (walls, pipes, equipments, geometric singularities etc.). Aircraft powerplant compartments are examples where topology is complex due to the presence of pipes, ducts, coolers and other equipment. Other important examples are gas explosions and large scale dispersion of hazardous materials in urban places, cities or underground involving obstacles such as buildings and various infrastructures. In all cases efficient safety responses are required. Then a new discrete model is built and solved in reasonable execution times for large cell volumes including such obstacles. Quantitative comparisons between experimental and numerical results are shown for different significant test cases, showing excellent agreement [18].

5.1.3. A hyperbolic Eulerian model for dilute two-phase suspensions

Participants: Sarah Hank, Richard Saurel, Olivier Le Métayer.

Conventional modeling of two-phase dilute suspensions is achieved with the Euler equations for the gas phase and gas dynamics pressureless equations for the dispersed phase, the two systems being coupled by various relaxation terms. The gas phase equations form a hyperbolic system but the particle phase corresponds to a hyperbolic degenerated one. Numerical difficulties are thus present when dealing with the dilute phase system. In the present work, we consider the addition of turbulent effects in both phases in a thermodynamically consistent way. It results in two strictly hyperbolic systems describing phase's dynamics. Another important feature is that the new model has improved physical capabilities. It is able, for example, to predict particle dispersion, while the conventional approach fails. These features are highlighted on several test problems involving particles jets dispersion and are compared against experimental data. With the help of a single parameter (a turbulent viscosity), excellent agreement is obtained for various experimental configurations studied by different authors [17].

5.1.4. Diffuse interface model for compressible fluid - Compressible elastic-plastic solid interaction

Participants: Nicolas Favrie, Sergey Gavrilyuk.
An Eulerian hyperbolic diffuse interface model for elastic plastic solid fluid interaction is constructed. The system of governing equations couples Euler equations of compressible fluids and a visco-plastic model of Maxwell type materials (the deviatoric part of the stress tensor decreases during plastic deformations) in the same manner as models of multicomponent fluids. In particular, the model is able to create interfaces which were not present initially.

The model is thermodynamically compatible: it verifies the entropy inequality. However, a numerical treatment of the model is particularly challenging. Indeed, the model is non-conservative, so a special numerical splitting is proposed to overcome this difficulty. The numerical algorithm contains two relaxation procedures. One of them is physical and is related to the plastic relaxation mechanism (relaxation toward the yield surface). The second one is numerical. It consists in replacing the algebraic equation expressing a mechanical equilibrium between components by a partial differential equation with a short relaxation time. The numerical method was tested in one dimensional case (Wilkins’ flying plate problem), two-dimensional plane case (impact of a projectile on a plate) and axisymmetrical case (Taylor test problem, impact with penetration effects, etc.). Numerical examples show the ability of the model to deal with real physical phenomena [13].

5.1.5. Criterion of hyperbolicity for non-conservative quasilinear systems admitting a partially convex conservation law

Participants: Alain Forestier, Sergey Gavrilyuk.

A system of conservation laws admitting an additional convex conservation law can be written as a symmetric -hyperbolic in the sense of Friedrichs system. However, in mathematical modeling of complex physical phenomena, it is customary to use non-conservative hyperbolic models. We generalize the Godunov Friedrichs Lax approach to this new class of models [16].

5.1.6. A new model of roll waves: comparison with Brock’s experiments

Participants: Gaël Richard, Sergey Gavrilyuk.

We derive a mathematical model of shear flows of shallow water down an inclined plane. Periodic stationary solutions to this model describing roll waves were obtained. The solutions are in good agreement with experimental profiles of roll waves measured in Brock’s experiments (1967). In particular, the height of the vertical front of the waves, the shock thickness and the wave amplitude are well captured by the model [21].

5.1.7. Modelling gas dynamics in one-dimensional ducts with abrupt area change

Participants: R Menina, Richard Saurel, M Zereg, Lazhar Houas.

Most gas dynamic computations in industrial ducts are done in one dimension with cross-section-averaged Euler equations. This poses a fundamental difficulty as soon as geometrical discontinuities are present. The momentum equation contains a non-conservative term involving a surface pressure integral, responsible for momentum loss. Definition of this integral is very difficult from a mathematical standpoint as the flow may contain other discontinuities (shocks, contact discontinuities). From a physical standpoint, geometrical discontinuities induce multidimensional vortices that modify the surface pressure integral. In the present paper, an improved one-dimensional flow model is proposed. An extra energy (or entropy) equation is added to the Euler equations expressing the energy and turbulent pressure stored in the vortices generated by the abrupt area variation. The turbulent energy created by the flow area change interaction is determined by a specific estimate of the surface pressure integral. Model’s predictions are compared with two-dimensional averaged results from numerical solution of the Euler equations. Comparison with shock tube experiments is also presented. The new one dimensional averaged model improves the conventional cross-section-averaged Euler equations and is able to reproduce the main flow features [19].

5.2. Applications for specific flow problems

5.2.1. Modelling cavitating flow around underwater missiles

The diffuse interface model of Saurel et al. [9] is used for the computation of compressible cavitating flows around underwater missiles. Such systems use gas injection and natural cavitation to reduce drag effects. Consequently material interfaces appear, separating liquid and gas. These interfaces may have a really complex dynamics such that only a few formulations are able to predict their evolution. Contrarily to front tracking or interface reconstruction method the interfaces are computed as diffused numerical zones, that are captured in a routinely manner, as is done usually with gas dynamics solvers for shocks and contact discontinuity. With the present approach, a single set of partial differential equations is solved everywhere, with a single numerical scheme. This leads to very efficient solvers. The algorithm derived in Saurel et al. [43] is used to compute cavitation pockets around solid bodies. It is first validated against experiments done in cavitation tunnel at CNU. Then it is used to compute flows around high speed underwater systems (Shkval-like missile). Performance data are then computed showing method ability to predict forces acting on the system [20].

5.2.2. Propagation of a planar shock wave through a two-phase gas-liquid medium

Participants: Alain Chauvin, Georges Jourdan, Éric Daniel, Lazhar Houas, R Tosello.

We conducted a series of shock tube experiments to study the influence of a cloud of water droplets on the propagation of a planar shock wave. In a vertically oriented shock tube, the cloud of droplets was released downwards into the air at atmospheric pressure while the shock wave propagated upwards. Two shock wave Mach numbers, 1.3 and 1.5, and three different heights of clouds, 150 mm, 400 mm, and 700 mm, were tested with an air-water volume fraction and a droplet diameter fixed at 1.2 % and 500 \( \mu \text{m} \), respectively. From high-speed visualization and pressure measurements, we analyzed the effect of water clouds on the propagation of the shock wave. It was shown that the pressure histories recorded in the two-phase gas-liquid mixture are different from those previously obtained in the gas-solid case. This different behavior is attributed to the process of atomization of the droplets, which is absent in the gas-solid medium. Finally, it was observed that the shock wave attenuation was dependent on the exchange surface crossed by the shock combined with the breakup criterion [12].
6. New Results

6.1. Probabilistic numerical methods, stochastic modelling and applications

Participants: Mireille Bossy, Nicolas Champagnat, Julia Charrier, Julien Claisse, Madalina Deaconu, Samuel Herrmann, James Inglis, Pierre-Emmanuel Jabin, Antoine Lejay, Sylvain Maire, Sebastien Niklitschek Soto, Nicolas Perrin, Denis Talay, Etienne Tanré, Laurent Violeau.

6.1.1. Published works and preprints

- M. Bossy in collaboration with J.-F. Jabir (Univ. Chile) proved the well posedness of the confined Lagrangian models, in association with no-permeability boundary conditions. When the confining domain is a hyperplane, they proved the strong existence of the trace of the density of particles following the kinetic stochastic equation of a simplified McKean Vlasov Lagrangian model in [12], http://hal.inria.fr/inria-00515481/en. When the confining domain $\mathcal{D}$ is bounded with smooth boundary, they constructed a confined primitive of Brownian motion in $\mathcal{D}$ and characterized the solution to the corresponding martingale problem by showing that the time marginal density is the unique solution to a mild equation with specular condition. This key step allowed them to finish the construction in the non linear case, using previous work on Vlasov-Fokker-Plank PDE with specular boundary condition. Two papers are being written.

- In collaboration with J.-F. Jabir and J. Fontbona (CMM and Universidad de Chile, Santiago de Chile), M. Bossy and P.-E. Jabin have studied the link between the Lagrangian version of divergence free constraint (and the uniform density constraint), with an additional potential term, in the Lagrangian equation, having some similarity with the role of the Eulerian pressure term. They obtained the local existence of analytical solutions for an incompressible Lagrangian stochastic model in periodic domain. An article is currently being written.

- N. Champagnat worked with A. Lambert (Univ. Paris 6) on splitting trees with Poissonian mutations. Assuming that each mutation is neutral and gives a new type in the population, they obtained in [15] explicit expressions for the expected number of types carried by a fixed number of individuals living in the population at time $t$. In [31], they also obtained large time convergence results on the sizes of the largest families and the ages of the oldest families in the population. http://hal.inria.fr/inria-00515481/en, http://hal.inria.fr/inria-00616765/en.

- N. Champagnat and P.-E. Jabin studied the limit of some population dynamics models under the assumption that the time scale for mutations is much larger than the time scale for reproduction. They are able to provide the first full characterization of the corresponding limit equation [14], http://hal.inria.fr/inria-00488979/en.

- M. Deaconu and S. Herrmann developed a new method for the simulation of the hitting time of nonlinear boundaries for Bessel processes. This method is based on a walk on moving spheres algorithm and can be applied for the hitting time of a given level for the Cox-Ingersoll-Ross process [32], http://hal.inria.fr/hal-00636056/en. This work is part of the ANR MANDy project.

- S. Herrmann and E. Tanré worked on a scheme to construct an efficient algorithm to simulate the first hitting time of curves by a one dimensional Brownian motion. They apply the result to estimate the spiking time of leaky integrate fire models in neurosciences. This work is part of the ANR MANDy project.
- P.-E. Jabin and F. Ben Belgacem (Univ. of Monastir, Tunisia) have studied a new class of models which have seen considerable development in applications for biosciences (flocking, chemotaxis, pedestrian flows...). These models include some non linear corrections to classical linear continuity equations. In [30], they introduce new, critical regularity estimates to obtain well posedness. http://www2.cscamm.umd.edu/~jabin/transportlcs2.pdf.

- P.-E. Jabin and M. Hauray (Aix-Marseille Université) have studied the mean field limit for systems of many interacting particles. It is the only result able to deal with singular forces and physically realistic initial configurations [33], http://hal.inria.fr/hal-00609453/en.

- P.-E. Jabin and A. Nouri (Aix-Marseille Université) studied a highly singular kinetic equation in dimension 1. This equation is obtained as a quasi-neutral limit in plasma physics. In [18], they were able to prove well posedness in short time of analytic solutions. http://dx.doi.org/10.1016/j.crma.2011.03.024.

- P.-E. Jabin and G. Raoul (Cambridge University) prove the convergence to a unique stable equilibrium for a wide class of competitive models in population dynamics [19], http://dx.doi.org/10.1007/s00285-010-0370-8.

- P.-E. Jabin and J. Calvo (Universidad de Granada) investigate the long time asymptotics of a new class of models for interacting particles inspired from various phenomena in the biosciences. In this model, when two particles collide they may coalesce and then completely stop moving [13], http://hal.inria.fr/hal-00601969/en.

- In collaboration with G. Pichot (INRIA Rennes Bretagne Atlantique), A. Lejay has developed a new Monte Carlo methods for discontinuous media that relies on the simulation of the Skew Brownian motion [22], [35], http://hal.inria.fr/hal-00642194/en, http://hal.inria.fr/hal-00649170/en.

- A. Lejay developed a new method for the simulation of a stochastic process in a layered media using the properties of the Brownian path [20], http://hal.inria.fr/inria-00583127/en.


- S. Maire and E. Tanré have generalised the spectral methods for elliptic PDEs developed in [39], [40] to the case of pure Neumann boundary conditions. Some additional difficulties occur because the stochastic representation of the solutions is defined only up to an additive constant and as a limit involving local time approximations [38]. By taking into account these additional properties, they still obtained a spectral matrix having a condition number converging to one.

- D. Talay and E. Tanré, in collaboration with F. Delarue and S. Rubenthaler (Univ. Nice – Sophia Antipolis), have given a precise approximation of the interspike intervals for the LIF model, describing the activity of a single neuron. This work is part of the ANR MANDy project (see Section 7.1.1).

- D. Talay, in collaboration with M. Martinez (Univ. Paris-Est), achieved to develop their stochastic approach for one-dimensional transmission parabolic problems. Owing to their stochastic representation of the solutions, they obtained accurate pointwise estimates for the derivatives of these solutions, from which they got accurate convergence rate estimates in the weak sense for a numerically effective discretization scheme of stochastic differential equations with weighted local times which are related to elliptic partial differential operators under divergence form with a discontinuous coefficient [36], http://hal.inria.fr/inria-00607967/en.

### 6.1.2. Other works in progress

- N. Champagnat studies in collaboration with S. Méléard (Ecole Polytechnique, Palaiseau) adaptive dynamics and evolutionary branching in individual-based models of populations competing for resources, similar to those involved in chemostat systems of ODEs.
N. Champagnat studies in collaboration with A. Lambert the process of the time to the most recent common ancestor in a family of subcritical branching processes whose genealogy is given by splitting trees.

J. Charrier joined the team in September as a post-doctoral researcher and began working with M. Bossy and D. Talay on the long time behaviour of stochastic particules systems in McKean-Vlasov interaction.

J. Claisse continued his PhD. under the supervision of N. Champagnat and D. Talay on stochastic control of population dynamics. He completed a finite-horizon and an infinite-horizon optimal control problem on a birth-death process. He is currently working on a birth-death process whose parameters depend on a controlled ordinary differential equation. In addition, he is working on applications of branching processes in biology and optimal control theory, and more specifically in cancer therapy.

M. Deaconu and S. Herrmann continue the study of the hitting time for Bessel processes in the situation of noninteger dimensions.

J. Inglis joined the team in October 2011 as a post-doctoral researcher (ANR MANDy), and began working with E. Tanré, D. Talay, F. Delarue (University of Nice) and S. Rubenthaler (University of Nice) on problems related to the rigorous justification of mean field models used in neuroscience.

J. Inglis, E. Tanré and M. Tejo (PUC, Chile) started a collaboration on the numerical simulation of spiking times of neurons described by some new stochastic models related to the Hodgkin-Huxley equation. This work is a part of Anestoch associated team.

A. Lejay and S. Maire study some new Monte Carlo methods for multi-dimensional discontinuous media.

In collaboration with J.-R. Li (INRIA Rocquencourt & Neurospin), A. Lejay studies some probabilistic representation for interface condition arising in diffusion Magnetic Resonance Imaging.

In collaboration with G. Pichot and J. Erhel (INRIA Rennes Bretagne Atlantique), A. Lejay studies Monte Carlo methods for discontinuous media as well as benchmarks and test on existing methods.

With L. Coutin (Univ. Toulouse), A. Lejay studies some perturbation results for solutions of Rough Differential Equations.

S. Maire develops with C. de Luigi (Univ. du Sud – Toulon – Var) and Jerôme Lelong (IMAG, Grenoble) resolution algorithms for the price of various european options in high dimension by coupling an adaptive deterministic integration algorithm and Principal Component Analysis tools.

S. Niklitschek continued his PhD. under the supervision of D. Talay on discretized stochastic differential equations related to one-dimensional partial differential equations of parabolic type involving a discontinuous drift coefficient. He obtained accurate pointwise estimates for the derivatives of these solutions, from which he gets convergence rate estimates in the weak sense of the stochastic discretization scheme. Now he is working on the extension of these results to the multi-dimensional setup.

N. Perrin continued his PhD. on stochastic methods in molecular dynamics under the supervision of M. Bossy, N. Champagnat and D. Talay. He is studying a method due to P. Malliavin (French Academy of Science) based on the Fourier analysis of covariance matrices with delay in order to identify the fast and slow components of a molecular dynamics and to construct simplified projected dynamics. He also studied probabilistic interpretation of the nonlinear Poisson-Boltzmann equation in Molecular Dynamics with BSDEs [ 37 ].

L. Violeau continued his PhD. on Stochastic Lagrangian Models and Applications to Downscaling in Fluid Dynamics under the supervision of M. Bossy and A. Rousseau (MÈRe team, INRIA Sophia Antipolis – Méditerranée, Montpellier). He studied the convergence in law of a sequence of penalized processes to the so called reflected langevin process in a convex domain. He is currently working on the rate of convergence of the particle approximation of conditional McKean stochastic models.
• P-E. Jabin and D. Talay continue to develop their innovating approach, which combines stochastic analysis and PDE analysis, for the time varying Hamilton-Jacobi-Bellman-McKean-Vlasov equations of the Lasry and Lions mean-field stochastic control theory.

6.2. Financial Mathematics

Participants: Mireille Bossy, Paul Charton, El Hadj Aly Dia, Dalia Ibrahim, Denis Talay, Etienne Tanré.

6.2.1. Published works and preprints

• In collaboration with N. Maïzi (CMA – Mines ParisTech) and O. Pourtallier (COPRIN team, INRIA Sophia Antipolis – Méditerranée), M. Bossy, and E.H.A. Dia studied the indifference pricing for carbon emission allowances, as a short term model value of carbon (see Section 7.1.2 ). The indifference pricing methodology describes the way an industrial agent on the emission allowances market chooses his production strategy. An utility function represents the preferences of the producer and its risk aversion. The outputs of its production have stochastic prices on the market, so that the optimal production strategy arises as the solution of a stochastic control problem. We extended the model hypotheses under which we get the well-posedness of the stochastic control problem and the associated HJB equation. We exhibited a simple case (marginal costs constant in time) where we proved the regularity of the value function via the explicit solution of the stochastic control problem [24], http://hal.inria.fr/hal-00645033/en . This particular case now can serve as a benchmark for the numerical solver currently developed in the framework of the ADEME Convention. It will also serve as a demonstrator case, with the objective of a public diffusion of the simulator CarbonQuant.

• M. Cissé (ENSAE-Sénégal), P. Patie (Univ. libre de Bruxelles) and E. Tanré have solved explicitly the optimal stopping problem with random discounting and an additive functional as cost of observations for a regular linear diffusion [17], http://hal.inria.fr/inria-00458901/en/.

6.2.2. Other works in progress

• P. Charton continued his PhD. under the supervision of M. Deaconu and A. Lejay. He studied some hedging strategies for day ahead markets of wind energy.

• Mathematical modelling for technical analysis techniques Since November 2009, D. Ibrahim has been working on her PhD. thesis on Mathematical modeling of technical analysis in finance, under supervision of D. Talay and E. Tanré. The aim of her work is to study the performances of a technical analysis tool designed to detect changes in the volatility term: the Bollinger Bands. First, she studied the performances of this indicator in a modified Black-Scholes model such that the rate of volatility changes at an unknown random time \( \tau \), independent of the Brownian motion governing the prices. She is interested to study whether this indicator can detect the changes in the volatility. So, she aims to study the tail probability of this indicator by using Karamata’s Tauberian Theorem for Laplace-Stieltjes transforms. Secondly, she exhibited a mathematical optimal strategy by modifying usual techniques in both the dual and the classical PDE approaches in stochastic control theory, in order to circumvent the discontinuity of the filtration generated by the price process. This work is part of the contract with FINRISK (see Section 8.3 ).

• P. Protter (Columbia University) and D. Talay started to develop a new bubble time evolution model.
TROPICS Project-Team

6. New Results

6.1. Automatic Differentiation and parallel codes

Participants: Valérie Pascual, Laurent Hascoët, Hubert Alcin, Jean Utke [Argonne National Lab. (Illinois, USA)], Uwe Naumann [RWTH Aachen University (Germany)].

This research is an ongoing joint work between three teams working on AD. We study differentiation in reverse mode of programs that contain MPI communication calls. Instead of the commonly used approach that encapsulates the MPI calls into black-box subroutines that will be differentiated by hand, we are looking for a native differentiation of the MPI calls by the AD tool. The ultimate goal of this work is to generate the adjoint of MPI-parallel codes.

One issue is to reduce the variability of the available MPI procedures and parameters to a smaller number of elementary concepts. We then address the basic question of sends and recvs, that may be blocking or nonblocking, individual or collective, and so on. Essentially the adjoint of a send is a recv, and vice-versa, but the possibility of nonblocking isend’s and irecv’s requires more subtlety.

We continue adaptation of the tool’s static analysis to programs with message-passing communication. In the framework of flow-sensitive and context-sensitive data-flow analysis, we introduce new “channel” variables and modify the general static data propagation mechanism.

Implementation in TAPENADE is well advanced: the main MPI procedures are now correctly understood and all data-flow analyses are adapted. In 2011, we obtained a first valid tangent differentiated code of the team’s CFD code AIRONUM. Work is continuing to obtain the adjoint.

Our team focuses on AD based on program transformation. On the other hand, we closely follow the developments of operator-overloading AD tools towards message-passing communication. This requires a more complex definition of the overloaded communication primitives [33].

6.2. AD adjoints and Dynamic Memory

Participants: Laurent Hascoët, Jean Utke [Argonne National Lab. (Illinois, USA)].

Again in the adjoint mode, dynamic memory allocation and the associated pointer manipulations pose difficult problems. As the adjoint mode is bound to recover past values of variables, the need arises to recover past values of pointers too. However, these values are addresses often relative to some dynamically allocated memory. The original program often manages memory by allocating and deallocating memory on the run. Correspondingly, the adjoint program will have to reallocate memory as it walks backwards along the original allocation. If recovery of past values is implemented through storage, e.g. on a stack, the stored addresses refer to memory that has been deallocated, and do not correspond to the reallocated memory zones.

In 2011 we investigated this problem in two principal directions:

- We may consider not storing the address, but rather recompute it by repeating in the reverse sweep its calculation from the forward sweep. This is not always possible as pointer assignment and manipulation may be complex and distributed in the code. However, when possible, it is certainly a very efficient approach. We propose an data-flow algorithm to detect applicable cases.

- For the remaining cases where storage must be used, we study an address mapping mechanism, such that addresses inside some allocated memory can be dynamically converted into addresses inside the corresponding reallocated memory.
6.3. Resolution of linearised systems

**Participants:** Hubert Alcin, Olivier Allain [Lemma], Anca Belme, Marianna Braza [IMF-Toulouse], Alexandre Carabias, Alain Dervieux, Bruno Koobus [Université Montpellier 2], Carine Moussaed [Université Montpellier 2], Hilde Ouvrard [IMF-Toulouse], Stephen Wornom [Lemma].

The interaction between the sophisticated solution algorithm inside a program and the Automatic Differentiation of the program is a non-trivial issue. An iterative algorithm generally does not store the successive updates of the iterated solution vector. Furthermore, a modern iterative solution algorithm involves several nonlinear processes, like in:

- the evaluation of an optimal step, which results at least from a homographic function of the unknown,
- the orthonormalisation of the updates (Gram-Schmidt method, Hessenberg method).

Applying reverse AD to the iterative solution algorithm produces a linearised iterative algorithm which is transposed and therefore follows a reverse order, with exactly the same number of iterations, and needing exactly each of the iterated state solution vectors. This effect is considerably amplified in the case of the numerical simulation of unsteady phenomena with implicit numerical schemes. For example, the simulation of high Reynolds turbulent flows by a Large Eddy Simulation (LES) requires hundreds of thousands time steps, each of them involving a modern iterative solution algorithm.

In the 4-year ECINADS ANR project, we design more efficient solution algorithms and we examine the questions risen by their reverse differentiation. The application domain is the computation of high Reynolds turbulent flows with LES and hybrid RANS-LES models. The efficiency will be evaluated through the practical scalability on a large number of processors. This efficiency criterion also extends to the scalability of the reverse/adjoint algorithm. ECINADS also addresses the scalable solution of new approximations. ECINADS associates the university of Montpellier 2, the Institut de Mécanique des fluides de Toulouse and Lemma company. The kick off meeting of ECINADS was held at end of 2009.

In 2011, Hubert Alcin has performed a study of deflation and balancing coarse grid methods for a set of scalar models. The methods has been extended to the incompressible Navier-Stokes model in Lemma’s software ANANAS by Olivier Allain and to compressible Navier-Stokes by Bruno Koobus and Hubert Alcin. A collection of benchmark tests on these models has been performed and show a good scalability for the tested algorithms. An article is prepared on these results. Hubert Alcin has also studied a novel method for three-level preconditioning. The new method will be extended to compressible flows in cooperation with the Montpellier team (Carine Moussaed and Bruno Koobus).

6.4. Perturbation Methods

**Participants:** Alain Dervieux, Laurent Hascoët.

In the context of the European project NODESIM-CFD, the contribution of Tropics involved mainly the differentiation of perturbation methods and reduced order models for the management of uncertainties. These methods rely on Taylor series with second-order terms. The production of second derivative code is obtained through repeated application of Automatic Differentiation. Three strategies can be applied to obtain (elements of) the Hessian matrix, named Tangent-on-Tangent (ToT), Tangent-on-Reverse (ToR), and Reverse-on-Tangent (RoT). These new methods are disseminated through short courses, as those given by Alain Dervieux at ERCOFTAC sessions (Munich and Hampton). The application and extension of these methods have motivated a joint application from the Italian Aircraft company Alenia and Tropics in a FP7 proposal (Proposal CARDINA, nov. 2011).

6.5. Control of approximation errors

**Participants:** Frédéric Alauzet [GAMMA team, INRIA-Rocquencourt], Estelle Mbinky [GAMMA team, INRIA-Rocquencourt], Olivier Allain [Lemma], Anca Belme, Alexandre Carabias, Hubert Alcin, Alain Dervieux.
This is a joint research between three INRIA teams GAMMA (Rocquencourt), TROPICS, PUMAS and Lemma company. Roughly speaking, GAMMA brings mesh and approximation expertise, TROPICS contributes to adjoint methods, and CFD applications are developed in the context of PUMAS and Lemma.

The resolution of the optimum problem using the innovative approach of an AD-generated adjoint can be used in a slightly different context than the optimal shape design, namely the mesh adaptation. This will be possible if we can map the mesh adaptation problem into a differentiable optimal control problem. To this end, we have introduced a new methodology that consists in stating the mesh adaptation problem in a purely functional form: the mesh is reduced to a continuous property of the computational domain: the continuous metric. We minimize a continuous model of the error resulting from that metric. Thus the problem of searching an adapted mesh is transformed into the search of an optimal metric.

In 2011, a work on goal-oriented mesh adaptation for unsteady Euler flows has been extended, with further analysis, and a paper has been written and submitted to a journal. Its extension to the compressible Navier-Stokes model has been developed, [ 11 ] and a paper is being written. A further extension to Large Eddy Simulation is started. The method is being extended to a third-order approximation, the Vertex-CENO. This approximation was defined during a collaboration between university of Montpellier, IMM-Moscow and Tropics. A more accurate version has been studied by Alexandre Carabias and presented in Honom-Trento. A new theory involving error estimates and criteria has been developed by Gamma and Tropics. The extension of the multiscale adaptation method is considered by Estelle Mbinky at Rocquencourt. The extension of the goal-oriented method is considered by Alexandre Carabias at Sophia. Anisotropic mesh adaptation allows for better convergence to continuous solutions, and in particular more accurate a posteriori error estimates and correctors. The synergy between correctors and mesh adaptation is currently analysed and is the subject of a joint contribution (Gamma and Tropics) for the FP7 CARDINA proposal (nov. 2011).
5. New Results

5.1. Modeling Interfaces and Contacts

5.1.1. On the Morphology of Protein Binding Patches

Participants: Frédéric Cazals, Noël Malod-Dognin.

In collaboration with A. Bansal, former summer intern from IIT Bombay.

Understanding the specificity of protein interactions is a central question in structural biology, whence the importance of models for protein binding patches—a patch refers to the collection of atoms of a given partner accounting for the interaction. To improve our understanding of the relationship between the structure of binding patches and the biological function of protein complexes, we present a binding patch model decoupling the topological and geometric properties [21]. While the geometry is classically encoded by the 3D positions of the atoms, the topology is recorded in a graph encoding the relative position of concentric shells partitioning the interface atoms. The topological - geometric duality provides the basis of a generic dynamic programming based algorithm to compare patches, which is instantiated to respectively favor topological or geometric comparisons.

On the biological side, using a dataset of 92 co-crystallized structures organized in biological sub-families, we exploit our encoding and the two comparison algorithms in two directions. First, we show that Nature enjoyed the topological and geometric degrees of freedom independently while retaining a finite set of qualitatively distinct topological signatures, and show that topological similarity is a less stringent notion that the ubiquitously used geometric similarity. Second, we analyze the topological and geometric coherence of binding patches within sub-families and across the whole database, and show that complexes related to the same biological function can encompass geometrically distinct shapes. Previous work on binding patches focused on the investigation of correlations between structural parameters and biochemical properties on the one hand, and on structural comparison algorithms on the other hand. We believe that the abstraction coded by the topological - geometric duality paves the way to new classifications, in particular in the context of flexible docking.

The corresponding software is presented in section 4.1.1.

5.2. Modeling Macro-molecular Assemblies

5.2.1. Assessing the Reconstruction of Macro-molecular Assemblies with Toleranced Models

Participants: Frédéric Cazals, Tom Dreyfus.

In collaboration with Valérie Doye, Institut Jacques Monod, Paris.

In [20], we introduce TOleranced Models (TOM), a generic and versatile framework meant to handle models of macro-molecular assemblies featuring uncertainties on the shapes and the positions of proteins. A TOM being a continuum of nested shapes, the inner (resp. outer) ones representing high (low) confidence regions, we present statistics to assess features of this continuum at multiple scales. While selected statistics target topological aspects (pairwise contacts, complexes involving proteins of prescribed types), others are of geometric nature (geometric accuracy of complexes). We validate the TOM framework on recent average models of the Nuclear Pore Complex (NPC) obtained from reconstruction by data integration, and confront our statistics against experimental findings related to sub-complexes of the NPC. In a broader perspective, the TOM framework should prove instrumental to handle uncertainties of various kind, in particular in electron-microscopy and crystallography.
5.2.2. Probing a Continuum of Macro-molecular Assembly Models with Graph Templates of Sub-complexes

**Participants:** Frédéric Cazals, Tom Dreyfus.

Reconstruction by data integration is an emerging trend to reconstruct large protein assemblies, but uncertainties on the input data yield average models whose quantitative interpretation is challenging. This paper presents methods to probe fuzzy models of large assemblies against atomic resolution models of sub-systems.

Consider a Toleranced Model (TOM) of a macro-molecular assembly, namely a continuum of nested shapes representing the assembly at multiple scales. Also consider a template namely an atomic resolution 3D model of a sub-system of this assembly—also called a complex. We present algorithms performing a multi-scale assessment of the complexes of the TOM, by comparing the pairwise contacts which appear in the TOM against those of the template. These operations reduce to the comparison of graphs, which we perform by computing Maximal Common Induced Sub-graphs (MCIS) and Maximal Common Edge Sub-graphs (MCES).

We apply this machinery to recent average models of the NPC. First, we show how our contact analysis allows assessing the quality of probability density maps. Regarding particular sub-systems of the NPC, we focus on the Y-complex and on the T-complex. In particular for the latter, our analysis suggests a new 3D template of pairwise contacts.

We believe that these tools should become standard to assess the reconstruction of fuzzy assemblies.

The software associated to these developments is presented in section 4.1.2.

5.3. Protein Shape Matching and Family Identification

5.3.1. Using Dominances for Solving the Protein Family Identification Problem

**Participant:** Noël Malod-Dognin.

*In collaboration with R. Andonov (IRISA), M. Le Boudic-Jamin (IRISA) and P. Kamath (former summer intern within the SYMBIOSE project at IRISA).*

The 3D structure of macro-molecules underpins all biological functions. Similarities between protein structures may come from evolutionary relationships, and similar protein structures relate to similar functions.

The exponential growth of the number of known protein structures in the Protein Data Bank over the past decade led to the problem of protein classification. We mean here how to automatically insert new protein structures into an already existing classified database \( Q = \{ q_1, q_2, \ldots, q_m \} \) such as CATH or SCOP. The problem of determining in which classes new structures \( P = \{ p_1, p_2, \ldots, p_n \} \) belong, according to a similarity function \( S: Q \times P \rightarrow \mathbb{R}^+ \), is referred here as the Protein Family Identification Problem (FIP).

There are computational pitfalls in the FIP. The number of similarity scores \( S(q_i, p_j) \) that need to be computed is \( |Q| \times |P| \), where \( |P| \) can be very large (there are currently 152920 classified protein structures in the expert classification CATH). Moreover, computing a single similarity score is often equivalent to solving a NP-hard problem (ex: DALI, DAST, CMO, VAST, etc...).

In [17] and [18], we propose a notion of dominance between the protein structure comparison instances that allows the computation of optimal FIP without optimally solving all the comparison instances, and thus reduces the effect of the NP-Hardness of the similarity score.

5.4. Algorithmic Foundations

5.4.1. Shape Matching by Localized Calculations of Quasi-isometric Subsets

**Participants:** Frédéric Cazals, Noël Malod-Dognin.
Consider a protein complex involving two partners, the receptor and the ligand. In [16], we address the problem of comparing their binding patches, i.e., the sets of atoms accounting for their interaction. This problem has been classically addressed by searching quasi-isometric subsets of atoms within the patches, a task equivalent to a maximum clique problem, a NP-hard problem, so that practical binding patches involving up to 300 atoms cannot be handled. We extend previous work in two directions. First, we present a generic encoding of shapes represented as cell complexes. We partition a shape into concentric shells, based on the shelling order of the cells of the complex. The shelling order yields a shelling tree encoding the geometry and the topology of the shape. Second, for the particular case of cell complexes representing protein binding patches, we present three novel shape comparison algorithms. These algorithms combine a Tree Edit Distance calculation (TED) on shelling trees, together with Edit operations respectively favoring a topological or a geometric comparison of the patches. We show in particular that the geometric TED calculation strikes a balance, in terms of accuracy and running time between a purely geometric and topological comparisons, and we briefly comment on the biological findings reported in a companion paper [21].
5. New Results

5.1. Medical Image Analysis

5.1.1. Spatial Decision Forests for MS Lesion Segmentation in Multi-Channel MR Images

Participants: Ezequiel Geremia, Nicholas Ayache, Olivier Clatz, Antonio Criminisi [MSR], Ender Konukoglu [MSR], Bjoern Menze [MIT].

- A new approach for MS lesions segmentation was proposed [33].
- Random forest for automatic segmentation of MS lesions in 3D MR images.
- Features: multi-channel MR intensities, priors, long-range spatial context, symmetry.
- Quantitative evaluation shows significant improvement over the MICCAI Grand Challenge 2008.
- The automatically learned decision sequence mimics the state-of-the-art pipeline.
- Independent validation carried out by the MICCAI Challenge website 2008.
- Exhaustive analysis of the discriminative power of channels and features.
- Analysis of the influence of random forest’s meta-parameters on the classification performance.

5.1.2. Left Ventricle Segmentation from Cardiac 4D Cine MRI Sequences

Participants: Jan Margeta [Correspondant], Ezequiel Geremia, Nicholas Ayache, Antonio Criminisi [MSR].

This work was performed in collaboration with Microsoft Research and was partly supported through its PhD Scholarship Programme.

- We extend the previous work for multiple sclerosis lesion segmentation of Geremia et al. [33] for spatio-temporal cardiac images.
- A fully automatic two layer left ventricle segmentation algorithm from 4D cardiac cine MRI sequences (See Fig. 2) was proposed for MICCAI STACOM LV segmentation challenge [63] using a random forest classification algorithm.
- Spatio-temporal features are used in the random forest framework to learn the segmentation task without explicitly defining the segmentation rules.
- Machine learning based MRI intensity standardization and pose normalization preprocessing pipeline was proposed to deal with diverse cardiac MRI datasets.

5.1.3. Design and use of anatomical atlases for automatic segmentation: application to radiotherapy of the head and neck region

Participants: Liliane Ramus [Correspondant], Grégoire Malandain, Vincent Grégoire [UCL], Juliette Thariat [CAL].

This work is done in collaboration with DOSIsoft S.A., Centre Antoine Lacassagne (CAL) and Université Catholique de Louvain.
Figure 1. Segmenting Case CHB05 from the public MSGC dataset. From left to right: preprocessed T1-weighted ($I_{T1}$), T2-weighted ($I_{T2}$) and FLAIR MR images ($I_{FLAIR}$) overlayed with the associated ground truth GT, the posterior map $Posterior = \{Posterior_v_k\}_k$ displayed using an inverted grey scale and the FLAIR sequence overlayed with the segmentation $Seg = \{Posterior \geq \tau_{posterior}\}$ with $\tau_{posterior} = 0.5$. Segmentation results show that most of lesions are detected. Although some lesions are not detected, e.g. peri-ventricular lesion in slice 38, they appear enhanced in the posterior map. Moreover the segmentations of slices 38 and 42 show peri-ventricular regions, visually very similar to MS lesions, but not delineated in the ground truth.
Figure 2. Two classification layers are used for left ventricle segmentation with random forests. (1) first layer posterior probability is used as a weight map for context aware MRI intensity standardization and cardiac pose normalization, (2) second layer is then used for a more confident left ventricle segmentation, (3) second level posterior probability isocontour overlay, (4) volumetric visualisation of the obtained segmentation.

In the context of radiotherapy of the head and neck, we propose different strategies to design anatomical atlases and we compare their performances for automatic segmentation [28]:

- We investigate average atlas construction, atlas stratification and patient-specific strategies based on the selection and fusion of the most appropriate atlases for each patient. We compared global, regional and local selection and fusion of the atlases.
- We show that the proposed patient-specific strategies enable to significantly improve the quality of the automatic segmentation in comparison with average atlas strategies. Visual results are presented on figure 3.
- We evaluated the proposed algorithms in two different contexts: segmentation of the lymph node levels and the organs at risk for radiotherapy planning, and segmentation of the teeth for post-irradiation dental care management. Automatic segmentations of the teeth are shown on figure 4.

5.2. Biological Image Analysis

5.2.1. Pre-clinical molecular imaging: reconstruction of tumors in rodents with SPECT imaging

Participants: Marine Breuilly [Correspondant], Grégoire Malandain, Nicholas Ayache, Jacques Darcourt [CAL], Philippe Franken [CAL], Thierry Pourcher [CEA].

This work is jointly conducted with the Transporter in Imagery and Oncologic Radiotherapy team (TIRO, CEA-CAL-UNSA) located in Nice.
Figure 3. Atlas-based segmentation of the lymph node level II using the average atlas (blue contours on left figure) and using the atlas that is locally adapted to the patient’s anatomy (yellow contours on right figure), compared with the manual contours (red contours on both figures).

Figure 4. Automatic segmentation of the teeth using a multi-atlas framework (in green) in comparison with the manual segmentation (in blue). Both segmentations are represented on the upper jaw.
The coupled CT and SPECT device allows to image both the anatomy (with the CT) and physiology information targeted by a dedicated radio-pharmaceutical tracer (here the tumors, with the SPECT). However, tumor quantification is impaired by the respiratory motion that induces an artificial enlargement of the moving structures. We propose then to select all the motion-less phases from a 4D SPECT images to reconstruct a motion-free 3D image. In addition, we also propose to correct for the heterogeneity of the respiratory cycles by re-tagging the SPECT raw data.

The resulting 3D motionless gated image shows improvement of volume accuracy compared to the non gated SPECT image; and noise reduction compared to the 4D SPECT image (see Figure 5).

![Figure 5. Coronal slices of fused SPECT and CT images from a NOD-SCID mouse (data acquired with GE eXplore speCZT CT 120): central hot spots reveal intraperitoneal metastasis from adenocarcinoma of the colon (PROb-mNIS). Left to right: non gated, phase at end of expiration, phase at end of inhalation, and motionless gated.](image)

**5.3. Computational Anatomy**

**5.3.1. The Kernel Bundle framework: Sparse Multiscale Diffeomorphic Deformations**

**Participants:** Stefan Sommer [DIKU], Mads Nielsen [DIKU], François Lauze [DIKU], Xavier Pennec [Correspondant].

*This work is performed in collaboration between DIKU (University of Copenhagen) and Asclepios (Inria). It was initiated during a 5 month visit of Stefan Sommer in 2011.*

In the analysis and modeling of anatomical deformations, we expect deformations to have both large and small scale components. However, we expect these large and small scale deformation to occur at different places. Thus, one would like to represent anatomical deformations with a small number of deformation atoms are different scales that are sparsely distributed across space and scale.

- In [71], we propose a multi-scale kernel bundle framework (LDDKBKM) that extends the LDDMM framework by incorporating multiple kernels at multiple scales in the registration. Experiments show that the method automatically adapts to the right scales, and it therefore removes the need for classical scale selection methods.
- In [72], we derive the Kernel Bundle EPDiff evolution equations, which provide optimal warps in this new framework.
5.3.2. Longitudinal modeling of the structural changes of the brain in Alzheimer’s disease.

**Participants:** Marco Lorenzi [Correspondant], Xavier Pennec, Giovanni Frisoni [IRCCS Fatebenefratelli Brescia, Italy], Nicholas Ayache.

This work is done in collaboration with LENITEM, IRCCS San Giovanni di Dio Fatebenefratelli, Brescia, Italy.

This work addresses the analysis and the quantification of the longitudinal structural changes of the brain affected by Alzheimer’s disease (AD). We propose a framework based on the non-rigid registration of brain MRIs using stationary velocity fields. In 2011, the main scientific developments were:

- Unbiased detection of the structural changes in the brain [60]. The method robustifies the Demons diffeomorphic registration by estimating and removing the multiplicative and additive intensity biases affecting the images.
- Definition of a population-based atlas for the longitudinal brain structural changes. In this work we proposed the Schild’s Ladder as a general method for parallel transporting the subject-specific longitudinal deformation trajectories in a reference space [61]. The work was awarded with the runner-up prize at the “Information Processing in Medical Imaging (IPMI)” conference in Irsee, Germany. Other transport methods from the Lie group theory have been successively proposed and investigated [62].
- Group-wise statistical analysis of the brain longitudinal changes in multiple time points [60]. The framework has been applied for the analysis of the longitudinal brain changes in healthy subjects at risk of AD, and the results showed an accelerated progression of atrophy for the subjects positive to the marker of Alzheimer Aβ42 (see Figure 6).

Finally, the above framework has been promoted to the neuroscience community as diagnostic tool and support for clinical trials [75].

5.3.3. Statistical Analysis of White Matter Fiber Bundles

**Participants:** Stanley Durrleman [Correspondant], Pierre Fillard [Parietal, INRIA Saclay], Xavier Pennec, Alain Trouvé [CMLA, ENS Cachan], Nicholas Ayache.

This work is an application of the generic morphometric method developed in 2009 to the statistical analysis of the shape and texture of white matter fiber bundles extracted from Diffusion Tensor Images (DTI).

- Registration, Atlas Estimation and Variability Analysis of White Matter Fiber Bundles Modeled as Currents [32].

5.3.4. Comparison of endocranial ontogenies in chimpanzees and bonobos

**Participants:** Stanley Durrleman [Correspondant], Xavier Pennec, Alain Trouvé [CMLA, ENS Cachan], Nicholas Ayache, José Braga [AMIS, Univ. Toulouse 3].

This work has been performed in the context of the INRIA collaborative project ARC 3D-Morphine (PI: Sylvain Prima, IRISA) and a follow-up collaboration with José Braga at Université Paul Sabatier, Toulouse.

This work quantifies ontogenetic differences between bonobo and chimpanzee endocrania, using dental development as a timeline. Synthetic endocranial trajectories are estimated from time series cross-sectional data. Then, differences in morphology and in rate of shape changes is quantified using the spatiotemporal registration introduced in 2009.

- Comparison of the endocranial ontogenies between chimpanzees and bonobos via temporal regression and spatiotemporal registration.
Figure 6. Top. Average longitudinal brain atrophy for the healthy subjects at risk for Alzheimer’s disease (Aβ42 positive). Bottom. Annual percentage differential evolution modelled for the Aβ42 positive group with respect to the healthy aging. The analysis shows an increased ventricular expansion and the matter loss in the cortex and in the temporal areas.
Figure 7. The proposed method estimates a variability model from the white matter fiber bundles extracted from six subjects (a-original data, b- same data approximated at the current resolution). The model allows us to synthesize artificial bundles that reproduce the variability in shape and in fiber density estimated from the original data (c).

Image taken from [32]

Figure 8. Typical endocranial trajectory of bonobos (top) and chimpanzees (bottom) estimated from time series cross-sectional surface data (59 chimpanzees and 60 bonobos)
Figure 9. Graph of time-wrap (magenta curve) putting into correspondence the developmental stages of the bonobos to that of the chimpanzees. The function measures the differences in rate of shape changes over time between the endocranial trajectories estimated in Fig. 8. This shows that the bonobos endocranial ontogeny is retarded by a factor 0.25 compared to that of the chimpanzees.

5.3.5. Statistical Modelling of Cardiac Growth and Deformation from Medical Images

Participants: Kristin McLeod [Correspondant], Tommaso Mansi, Adityo Prakosa, Maxime Sermesant, Xavier Pennec.

Parts of this work were performed within the framework of the EU project Care4me ITEA2, and the INRIA ARC Sirap, in collaboration with St Thomas Hospital, King’s College London, the REO team from INRIA Rocquencourt and Necker Paediatric Hospital in Paris.

This work builds on the statistical analysis framework for surfaces developed by Durrleman and Mansi in 2009.

- The iLogDemons motion tracking algorithm of Mansi et al. [37] was applied to a data-set of 15 subjects and 1 phantom each with a cine-MR, tagged-MR and echocardiography sequence as a part of the STACOM workshop challenge at MICCAI 2011 [64]. The paper received the Best Paper Award for the Motion Challenge.
- The work of Mansi et al. for a statistical model of cardiac growth in the right ventricle [38] was extended to the left ventricle to obtain a bi-ventricular cardiac growth model for patients with repaired tetralogy of Fallot (see Fig. 10).
- The preliminary analysis of a statistical model for reduced blood flow simulations in the pulmonary artery proposed in 2010 is currently being extended to a journal version to further analyse the method on a larger data-set.

5.3.6. Statistical Modeling of Shapes Using Trees of Locally Affine Transformations

Participants: Christof Seiler [Correspondant], Xavier Pennec, Mauricio Reyes [Institute for Surgical Technology and Biomechanics, University of Bern, Switzerland].
Figure 10. Mean growth model computed from a population of 13 repaired tetralogy of Fallot patients. Both ventricles grow as body surface area (used as an index of growth) increases.

This work is performed in the context of the joint PhD of Christof Seiler at the Institute for Surgical Technology and Biomechanics, University of Bern, Switzerland and Asclepios INRIA.

The goal of this work is to analyze anatomical shapes through deformations defined with few but important and intelligible parameters. Advances towards this goal were the following in 2011.

- Fusion of the Log-demons registration and the Log-Euclidean polyaframe framework. The results of the new registration method applied to femur CT’s was presented at SPIE [69].
- Decomposition of diffeomorphic deformations into a tree of locally affine transformations applied to mandible CT’s. This work won the young scientist award at MICCAI 2011 in Toronto [68].

5.3.7. Atlas of Cardiac Fiber Architecture from DT-MRI

Participants: Hervé Lombaert, Hervé Delingette [Correspondant], Nicholas Ayache, Jean-Marc Peyrat, Pierre Croisille.

This work is a collaboration between Creatis, Lyon and INRIA Sophia Antipolis, including members of Ecole Polytechnique of Montreal. Financial support is partly from the National Science and Engineering Research Council of Canada, and the Michael Smith Foreign Study Program.

The variability of the cardiac fiber architecture has been investigated in a human population. An automatic method has been developed to construct an atlas of DTMRI images. A statistical analysis has been carried on using the Log-Euclidean metric.

- An automatic method has been developed to construct an atlas of DTMRI images. The first human atlas of DTMRI has been build [56]. This work received the Best Paper Award at the FIMH 2011 conference in New York City.
- Results have also been published in [59] where the variability of the cardiac architecture is measured using the Log-Euclidean metric.
Figure 11. First three levels of space decomposition. Left: Our new method decomposes the image domain using a tree of oriented bounding boxes. Right: For comparison, a traditional approach using a dyadic multiresolution scheme.

- The variability of the cardiac laminar sheets in human DTMRI has been studied [58].
- Differences in a population of normal and abnormal hearts have been investigated [57].

5.4. Computational Physiology

5.4.1. Tumor Growth Modeling

Participants: Erin Stretton [Correspondant], Nicholas Ayache, Hervé Delingette, Bjoern Menze, Ender Konukoglu, Ezequiel Geremia, Emmanuel Mandonnet.

This work was funded by Care4me.

- Performing a sensitivity analysis of long time series of multi-modal data.
- Improving our current models and their inputs, including creating a white matter and brain mask template with the help of a neurosurgeon and integrating the Powell bound constrained optimization into the minimization routine.
- Performed comparison of using a patient DTI, an atlas DTI or no DTI at all showing the difference in accuracies in the simulation results since the patient DTI is not always available in a clinical setting or is of poor quality. We found that using an atlas DTI produced only slightly less accurate results than using a patient DTI, where as using no DTI at all did not produce accurate results.

5.4.2. Synthetic Echocardiographic Image Sequences for Cardiac Inverse Electro-Kinematic Learning

Participants: Adityo Prakosa [Correspondent], Maxime Sermesant, Hervé Delingette, Eric Saloux [CHU Caen], Pascal Allain [Philips Healthcare], Pascal Cathier [Philips Healthcare], Patrick Ettyngier [Philips Healthcare], Nicolas Villain [Philips Healthcare], Nicholas Ayache.
Figure 12. Tracked fibers from the human atlas of DTMRI of the myocardium.
A database of 120 synthetic 3D echocardiography (US) sequences is created based on a cardiac electromechanical model (see Figure 13).

Kinematic descriptors are extracted from the displacement field estimated from the synthetic 3D US sequence using the iLogDemons non-rigid registration method [37].

Cardiac inverse electro-kinematic learning is done by using the database of synthetic 3D US sequences in order to estimate the cardiac depolarization times for the given kinematic descriptors [67]. First evaluation on two clinical sequences from patients with Left Bundle Branch Block shows encouraging results.

5.4.3. Prediction of patient-specific Ventricular Tachycardia for radio-frequency ablation therapy planning

Participants: Jatin Relan [Correspondant], Maxime Sermesant, Hervé Delingette, Nicholas Ayache.

This work is funded by the FP7 European Project euHeart.

In this work, we build a patient-specific cardiac electrophysiology (EP) model derived from hybrid XMR imaging and non-contact electro-anatomical mapping procedure on a patient with heart failure. The model is then used to predict patient-specific arrhythmias, such as induced ischemic Ventricular Tachycardia (VT) (Fig. 14) and leads in generation and evaluation of patient-specific VT circuits, with critical exit points for Radio Frequency (RF) ablation. These predictions are now validated with some clinical cases, with electrophysiology mapping of induced VT in patients undergoing the clinical VT-Stim procedures (Fig. 14).

5.4.4. Real-time simulation of catheter ablation in the context of cardiac arrhythmia

Participants: Hugo Talbot [Correspondant], Federico Spadoni, Maxime Sermesant, Hervé Delingette, Stephane Cotin.
Figure 14. (a & b) Analysis of electroanatomical mapping data, to personalise the depolarisation and repolarisation wavefront dynamics. (c) Anatomical data personalisation along with ischemic regions (red). (d) A VT inducibility map showing regions with high probability of inducing VT with a VT-Stim procedure. (e) Critical exit point map showing the most eligible regions for RF ablation success. (f & g) An example of one of the various induced VT circuits, with VT-Stim model prediction. (h & i) Induced VT circuit in a clinical case (red = exit point).

This work is performed in the context of the euHeart project and the PhD of Hugo Talbot in collaboration with the Shacra (INRIA, Lille Nord Europe) team.

- This work aims at simulating in real-time the endovascular procedure of radiofrequency ablation of the left ventricle for patient suffering from Ventricular Tachycardia.
- Fast simulation of electrophysiology has been reached with a Eikonal model\[40\].
- Use the SOFA platform for simulating endovascular navigation and cardiac electrophysiology.

5.4.5. Personalized model of the heart for cardiac therapy planning

**Participants:** Stéphanie Marchesseau [Correspondant], Ken C.L. Wong, Hervé Delingette, Maxime Serme- sant, Nicholas Ayache.

This work has been performed in the context of the euHeart european project.

- We improved the existing electromechanical model of the heart (see Fig 16 ) to include mechanical non linearity, viscosity and strain rate dependent contractility. It was implemented in SOFA with a new four valves model to deal with the cardiac phases and enforce isovolumetric constraint.
- We have obtained first personalization of cardiac mechanics from 3 cine-MRI cases using a variational approach (adjoint method)[54].
Figure 15. (Left) Simulation of catheterization with a force-feedback device; Simulation of cardiac electrophysiology.

Figure 16. (Left) Electromechanical model of the heart that is coupled with medical images; (Right) Pressures and volumes curves of the left ventricle resulting from the simulation of one cardiac cycle.
6. New Results

6.1. Computational Diffusion MRI

This sub-theme is dedicated to describe our various contributions performed within the framework of Computational Diffusion MRI. In 6.1.1, we start by presenting our contributions to improving dMRI signal and optimize dMRI acquisition schemes. Then, we present our modeling contributions related to the problem of reconstructing and characterizing important Diffusion MRI features such as the Orientation Distribution Function (ODF) and the Ensemble Average Propagator (EAP) in 6.1.2, including contributions of the compressed sensing theory to dMRI and contributions to online motion detection. Finally, we end up, in 6.1.3, with some general applications such as tractography, clustering and microstructures recovery with pore size distribution estimation.

6.1.1. Improving dMRI Signal and Acquisitions

6.1.1.1. Optimal Design of Multiple Q-shells experiments for Diffusion MRI

Participants: Rachid Deriche, Emmanuel Caruyer, Iman Aganj [Department of Electrical and Computer Engineering, University of Minnesota], Christophe Lenglet [Department of Electrical and Computer Engineering, University of Minnesota], Guillermo Sapiro [Department of Electrical and Computer Engineering, University of Minnesota], Jian Cheng [ATHENA and LIAMA, China], Jiang Tianzi [LIAMA, China].

This work was partly supported by the CD-MRI Associate Team.

Recent advances in diffusion MRI make use of the diffusion signal sampled on the whole Q-space, rather than on a single sphere. While much effort has been done to design uniform sampling schemes for single shell experiment, it is yet not clear how to build a strategy to sample the diffusion signal in the whole Fourier domain. In this work, we proposed a method to generate acquisition schemes for multiple Q-shells experiment in diffusion MRI. The acquisition protocols we designed are incremental, which means they remain approximately optimal when truncated before the acquisition is complete. Our method is fast, incremental, and we can generate diffusion gradients schemes for any number of acquisitions, any number of shells, and any number of points per shell. The samples arranged on different shells do not share the same directions. The method has been tested for Spherical Polar Fourier reconstruction of the diffusion signal, and is based on Monte-Carlo simulations. Several preferred acquisition parameters are identified.

This work has been published in [20].

6.1.1.2. Incremental gradient table for multiple Q-shells diffusion MRI

Participants: Rachid Deriche, Emmanuel Caruyer, Iman Aganj [Department of Electrical and Computer Engineering, University of Minnesota], Christophe Lenglet [Department of Electrical and Computer Engineering, University of Minnesota], Guillermo Sapiro [Department of Electrical and Computer Engineering, University of Minnesota].

This work was partly supported by the CD-MRI Associate Team.

Most studies on sampling optimality for diffusion MRI deal with single Q-shell acquisition. For single Q-shell acquisition, incremental gradient table has proved useful in clinical setup, where the subject is likely to move, or for online reconstruction. In this work, we proposed a generalization of the electrostatic repulsion to generate gradient tables for multiple Q-shells acquisitions, designed for incremental reconstruction or processing of data prematurely aborted.

This work has been published in [21].

6.1.1.3. Impact of radial and angular sampling on multiple shells acquisition in diffusion MRI

Participants: Rachid Deriche, Sylvain Merlet, Emmanuel Caruyer.
In this work, we evaluated the impact of radial and angular sampling on multiple shells (MS) acquisition in diffusion MRI. The validation of our results is based on a new and efficient method to accurately reconstruct the Ensemble Average Propagator (EAP) in term of the Spherical Polar Fourier (SPF) basis from very few diffusion weighted magnetic resonance images (DW-MRI). This approach nicely exploited the duality between SPF and a closely related basis in which one can respectively represent the EAP and the diffusion signal using the same coefficients. We efficiently combined this relation to the recent acquisition and reconstruction technique called Compressed Sensing (CS). Based on results of multi-tensors models reconstruction, we showed how to construct a robust acquisition scheme for both neural fibre orientation detection and attenuation signal/EAP reconstruction.

This work has been published in [32].

6.1.1.4. Simultaneous Smoothing and Estimation of DTI via Robust Variational Non-local Means

Participants: Rachid Deriche, Meizhu Liu [Department of CISE, University of Florida, Gainesville, USA], Baba Vemuri [Department of CISE, University of Florida, Gainesville, USA].

Regularized diffusion tensor estimation is an essential step in DTI analysis. There are many methods proposed in literature for this task but most of them are neither statistically robust nor feature preserving denoising techniques that can simultaneously estimate symmetric positive definite (SPD) diffusion tensors from diffusion MRI. One of the most popular techniques in recent times for feature preserving scalar-valued image denoising is the non-local means filtering method that has recently been generalized to the case of diffusion MRI denoising. However, these techniques denoise the multi-gradient volumes first and then estimate the tensors rather than achieving it simultaneously in a unified approach. Moreover, some of them do not guarantee the positive definiteness of the estimated diffusion tensors. In this work, we proposed a novel and robust variational framework for the simultaneous smoothing and estimation of diffusion tensors from diffusion MRI. Our variational principle makes use of a recently introduced total Kullback-Leibler (tKL) divergence, which is a statistically robust similarity measure between diffusion tensors, weighted by a non-local factor adapted from the traditional non-local means filters. For the data fidelity, we use the nonlinear least-squares term derived from the Stejskal-Tanner model. We have performed experimental results depicting the positive performance of our method in comparison to competing methods on synthetic and real data examples.

This work has been published in [31].

6.1.1.5. Anisotropic LMMSE denoising of MRI based on statistical tissue models

Participants: Rachid Deriche, Gonzalo Vegas-Sánchez-Ferrero [Universidad de Valladolid, Spain], Santiago Aja Fernández [Universidad de Valladolid, Spain].

Linear Minimum Mean Squared Error Estimation (LMMSE) is a simple, yet powerful denoising technique within MRI. It is based on the computation of the mean and variance of the data being filtered according to a noise model assumed, which is usually accomplished by calculating local moments over squared neighborhoods. When these neighborhoods are centered in pixels corresponding to image contours, the estimation is not accurate due to the presence of two or more tissues with different statistical properties. In this work, we overcome this limitation by introducing an anisotropic LMMSE scheme: the grey levels of each tissue in the MRI volume are modeled as a Gamma-mixture, such that we can discriminate between the different matters to construct anisotropic neighborhoods containing only one kind of tissue. The potential of the Gamma distribution relies on its ability to fit both the Rician distribution traditionally used to model the noise in MRI and the non-central Chi noise found in modern parallel MRI systems.

This work is currently under submission.

6.1.2. Modeling in Diffusion MRI

6.1.2.1. Multiple q-Shell Diffusion Propagator Imaging

Participants: Rachid Deriche, Maxime Descoteaux [Sherbrooke University, Quebec], Denis Le Bihan [NeuroSpin, IFR 49 CEA Saclay], Jean-François Mangin [NeuroSpin, IFR 49 CEA Saclay], Cyril Poupon [NeuroSpin, IFR 49 CEA Saclay].
Many recent high angular resolution diffusion imaging (HARDI) reconstruction techniques have been introduced to infer an orientation distribution function (ODF) of the underlying tissue structure. These methods are more often based on a single-shell (one b-value) acquisition and can only recover angular structure information contained in the ensemble average propagator (EAP) describing the three-dimensional (3D) average diffusion process of water molecules. The EAP can thus provide richer information about complex tissue microstructure properties than the ODF by also considering the radial part of the diffusion signal. In this work, we presented a novel technique for analytical EAP reconstruction from multiple q-shell acquisitions. The solution is based on a Laplace equation by part estimation between the diffusion signal for each shell acquisition. This simplifies greatly the Fourier integral relating diffusion signal and EAP, which leads to an analytical, linear and compact EAP reconstruction. An important part of this work is dedicated to validate the diffusion signal estimation and EAP reconstruction on real datasets from ex vivo phantoms. We also illustrated multiple q-shell diffusion propagator imaging (mq-DPI) on a real in vivo human brain and performed a qualitative comparison against state-of-the-art diffusion spectrum imaging (DSI) on the same subject. mq-DPI is shown to reconstruct robust EAP from only several different b-value shells and less diffusion measurements than DSI. This opens interesting perspectives for new q-space sampling schemes and tissue microstructure investigation.

This work has been published in [13].

6.1.2.2. A Riemannian Framework for Ensemble Average Propagator Computing

Participants: Rachid Deriche, Jian Cheng [ATHENA and LIAMA, China], Aurobrata Ghosh, Jiang Tianzi [LIAMA, China].

This work was partly supported by the Association France Parkinson and the ANR NucleiPark project.

In Diffusion Tensor Imaging (DTI), Riemannian framework (RF) has been proposed for processing tensors, which is based on Information Geometry theory. Many papers have shown that RF is useful in tensor estimation, interpolation, smoothing, regularization, segmentation and so on. Recently RF also has been proposed for Orientation Distribution Function (ODF) computing and it is applicable to any Probability Density Function (PDF) based on any orthonormal basis representation. Spherical Polar Fourier Imaging (SPFI) was proposed recently to fast and robustly estimate the ODF and Ensemble Average Propagator (EAP) from arbitrary sampled DWI signals. In this work, we proposed the RF for EAPs and implemented it via SPFI. We proved that the RF for EAPs is diffeomorphism invariant, which is the natural extension of affine invariant RF for tensors. It could avoid the so-called swelling effect for interpolating EAPs, just like the RF for tensors. We also proposed the Log-Euclidean framework (LEF), Affine-Euclidean framework (AEF), for fast processing EAPs, and Geometric Anisotropy (GA) for measuring the anisotropy of EAPs, which are all the extensions of previous concepts in RM for tensors respectively.

This work has been published in [22].

6.1.2.3. Diffeomorphism Invariant Riemannian Framework for Ensemble Average Propagator Computing

Participants: Rachid Deriche, Jian Cheng [ATHENA and LIAMA, China], Aurobrata Ghosh, Jiang Tianzi [LIAMA, China].

This work was partly supported by the Association France Parkinson and the ANR NucleiPark project.

In Diffusion Tensor Imaging (DTI), Riemannian framework based on Information Geometry theory has been proposed for processing tensors on estimation, interpolation, smoothing, regularization, segmentation, statistical test and so on. Recently Riemannian framework has been generalized to Orientation Distribution Function (ODF) and it is applicable to any Probability Density Function (PDF) under orthonormal basis representation. Spherical Polar Fourier Imaging (SPFI) was proposed for ODF and Ensemble Average Propagator (EAP) estimation from arbitrary sampled signals without any assumption. Tensors only can represent Gaussian EAP and ODF is the radial integration of EAP, while EAP has full information for diffusion process. To our knowledge, so far there is no work on how to process EAP data. In this work, we presented a Riemannian framework as a mathematical tool for such task. We proposed a state-of-the-art Riemannian framework for EAPs by representing the square root of EAP, called wavefunction based
on quantum mechanics, with the Fourier dual Spherical Polar Fourier (dSPF) basis. In this framework, the exponential map, logarithmic map and geodesic have closed forms, and weighted Riemannian mean and median uniquely exist. We analyzed theoretically the similarities and differences between Riemannian frameworks for EAPs and for ODFs and tensors. The Riemannian metric for EAPs is diffeomorphism invariant, which is the natural extension of the affine-invariant metric for tensors. We proposed Log-Euclidean framework to fast process EAPs, and Geodesic Anisotropy (GA) to measure the anisotropy of EAPs. With this framework, many important data processing operations, such as interpolation, smoothing, atlas estimation, Principal Geodesic Analysis (PGA), can be performed on EAP data. The proposed Riemannian framework was validated in synthetic data for interpolation, smoothing, PGA and in real data for GA and atlas estimation. Riemannian median is much robust for atlas estimation.

This work has been published in [23].

6.1.2.4. Theoretical Analysis and Practical Insights on EAP Estimation via a Unified HARDI Framework

Participants: Rachid Deriche, Jian Cheng [ATHENA and LIAMA, China], Jiang Tianzi [LIAMA, China].

This work was partly supported by the Association France Parkinson and the ANR NucleiPark project.

Since Diffusion Tensor Imaging (DTI) cannot describe complex non-Gaussian diffusion process, many techniques, called as single shell High Angular Resolution Diffusion Imaging (sHARDI) methods, reconstruct the Ensemble Average Propagator (EAP) or its feature Orientation Distribution Function (ODF) from diffusion weighted signals only in single shell. Q-Ball Imaging (QBI) and Diffusion Orientation Transform (DOT) are two famous sHARDI methods. However, these sHARDI methods have some intrinsic modeling errors or need some unreal assumptions. Moreover they are hard to deal with signals from different q-shells. Most recently several novel multiple shell HARDI (mHARDI) methods, including Diffusion Propagator Imaging (DPI), Spherical Polar Fourier Imaging (SPFI) and Simple Harmonic Oscillator based Reconstruction and Estimation (SHORE), were proposed to analytically estimate EAP or ODF from multiple shell (or arbitrarily sampled) signals. These three methods all represent diffusion signal with some basis functions in spherical coordinate and use plane wave formula to analytically solve the Fourier transform. To our knowledge, there is no theoretical analysis and practical comparison among these sHARDI and mHARDI methods. In this work, we proposed a unified computational framework, named Analytical Fourier Transform in Spherical Coordinate (AFT-SC), to perform such theoretical analysis and practical comparison among all these five state-of-the-art diffusion MRI methods. We compared these five methods in both theoretical and experimental aspects. With respect to the theoretical aspect, some criteria are proposed for evaluation and some differences together with some similarities among the methods are highlighted. Regarding the experimental aspect, all the methods were compared in synthetic, phantom and real data. The shortcomings and advantages of each method were highlighted from which SPFI appears to be among the best because it uses an orthonormal basis that completely separates the spherical and radial information.

This work has been published in [24].

6.1.2.5. Compressive Sensing Ensemble Average Propagator Estimation via L1 Spherical Polar Fourier Imaging

Participants: Rachid Deriche, Sylvain Merlet, Emmanuel Caruyer, Jian Cheng [ATHENA and LIAMA, China], Jiang Tianzi [LIAMA, China].

In diffusion MRI (dMRI) domain, many High Angular Resolution Diffusion Imaging (HARDI) methods were proposed to estimate Ensemble Average Propagator (EAP) and Orientation Distribution Function (ODF). They normally need many samples, which limits their applications. Some Compressive Sensing (CS) based methods were proposed to estimate ODF in Q-Ball Imaging (QBI) from limited samples. However EAP estimation is much more difficult than ODF in QBI. Recently Spherical Polar Fourier Imaging (SPFI) was proposed to represent diffusion signal using Spherical Polar Fourier (SPF) basis without specific assumption on diffusion signals and analytically obtain EAP and ODF via the Fourier dual SPF (dSPF) basis from arbitrarily sampled signal. Normally the coefficients of SPF basis are estimated via Least Square with weighted L2 norm regularization (L2-SPFI). However, L2-SPFI needs a truncated basis to avoid overfitting, which brings some estimation errors. By considering the Fourier relationship between EAP and signal and the Fourier basis pair provided in SPFI, we proposed a novel EAP estimation method, named L1-SPFI, to estimate
EAP from limited samples using CS technique, and favorably compared it to the classical L2-SPFI method. L1-SPFI estimates the coefficients in SPFI using least square with weighted L1 norm regularization. The weights are designed to enhance the sparsity. L1-SPFI significantly accelerates the ordinary CS based Fourier reconstruction method. This is performed by using SPF basis pair in CS estimation process which avoids the numerical Fourier transform in each iteration step. By considering high order basis in L1 optimization, L1-SPFI improves EAP reconstruction especially for the angular resolution. The proposed L1-SPFI was validated by synthetic, phantom and real data. The CS EAP and ODF estimations are discussed in detail and we showed that recovering the angular information from CS EAP requires much less samples than exact CS EAP reconstruction. Various experiments on synthetic, phantom and real data validate the fact that SPF basis can sparsely represent DW-MRI signals and L1-SPFI largely improves the CS EAP reconstruction especially the angular resolution.

This work has been published in [25], [26].

6.1.2.6. Spherical Polar Fourier EAP and ODF Reconstruction via Compressed Sensing in Diffusion MRI

Participants: Rachid Deriche, Sylvain Merlet, Aurobrata Ghosh, Jian Cheng [ATHENA and LIAMA, China].

In diffusion magnetic resonance imaging (dMRI), the Ensemble Average Propagator (EAP), also known as the propagator, describes completely the water molecule diffusion in the brain white matter without any prior knowledge about the tissue shape. In this work, we described a new and efficient method to accurately reconstruct the EAP in terms of the Spherical Polar Fourier (SPF) basis from very few diffusion weighted magnetic resonance images (DW-MRI). This approach exploits the duality between SPF and a closely related basis in which one can respectively represent the EAP and the diffusion signal using the same coefficients, and efficiently combines it to the recent acquisition and reconstruction technique called Compressed Sensing. Our work provides an efficient analytical solution to estimate, from few measurements, the diffusion propagator at any radius. We also provide a new analytical solution to extract an important feature characterising the tissue microstructure: the Orientation Distribution Function (ODF). We illustrate and prove the effectiveness of our method in reconstructing the propagator and the ODF on both noisy multiple q-shell synthetic and phantom data.

This work has been published in [33].

6.1.2.7. On Line Reconstruction and Motion Detection in HARDI

Participants: Rachid Deriche, Emmanuel Caruyer, Iman Aganj [Department of Electrical and Computer Engineering, University of Minnesota], Christophe Lenglet [Department of Electrical and Computer Engineering, University of Minnesota], Guillermo Sapiro [Department of Electrical and Computer Engineering, University of Minnesota].

This work was partly supported by the CD-MRI Associate Team.

With acquisition protocols such as high angular resolution diffusion imaging, head motion can become an issue. Although the misalignment between diffusion-weighted images (DWIs) can be corrected in a post-processing step, this might increase partial volume effects, because of the relatively low spatial resolution of DWIs and interpolation in the registration procedure. If able to detect motion online, the scanner technician could be issued a warning and make a decision accordingly. Orientation distribution functions (ODF) can be reconstructed online using a Kalman filter (KF). In this work, we presented three contributions related to the problem of online ODF reconstruction and motion detection in HARDI. First, we developed a proper error propagation accounting for the non-linear transform on the diffusion signal. Next, we developed two motion detection algorithms, based on the monitoring of residuals, and compared them using synthetic data.

This work has been published in [18].

6.1.2.8. Online Motion Detection in High Angular Resolution Diffusion Imaging

Participants: Rachid Deriche, Emmanuel Caruyer, Iman Aganj [Department of Electrical and Computer Engineering, University of Minnesota], Christophe Lenglet [Department of Electrical and Computer Engineering, University of Minnesota], Guillermo Sapiro [Department of Electrical and Computer Engineering, University of Minnesota].
This work was partly supported by the CD-MRI Associate Team. The orientation distribution function (ODF) can be reconstructed online incrementally from diffusion-weighted MRI with a Kalman filtering framework. This online reconstruction can provide real-time feedback to the practitioner, especially appreciated for long acquisition protocols typical in Q-ball imaging. On top of the Kalman filter, we proposed a method to evaluate online the reconstruction accuracy of the estimated ODF in constant solid angle. In addition, monitoring the residuals of the Kalman filter, we designed, based on statistical tests, two algorithms for online detection of subject motion. The proposed techniques, tested on real and synthetic data under various experimental conditions, can detect rotation by angle less than 3°.

This work has been published in [19].

6.1.3. From DW-MRI to Fiber Pathways and Microstructures Recovery

6.1.3.1. A Polynomial Approach for Maxima Extraction and Its Application to Tractography in HARDI

Participants: Rachid Deriche, Aurobrata Ghosh, Demian Wassermann [Harvard Medical School].

A number of non-parametrically represented High Angular Resolution Diffusion Imaging (HARDI) spherical diffusion functions have been proposed to infer more and more accurately the heterogeneous and complex tissue microarchitecture of the cerebral white-matter. These spherical functions overcome the limitation of Diffusion Tensor Imaging (DTI) at discerning crossing, merging and fanning axonal fiber bundle configurations inside a voxel. Tractography graphically reconstructs the axonal connectivity of the cerebral white-matter in vivo and non-invasively, by integrating along the direction indicated by the local geometry of the spherical diffusion functions. Tractography is acutely sensitive to the local geometry and its correct estimation. In this work, we first proposed a polynomial approach for analytically bracketing and numerically refining with high precision all the maxima, or fiber directions, of any spherical diffusion function represented non-parametrically. This permits an accurate inference of the fiber layout from the spherical diffusion function. Then we proposed an extension of the deterministic Streamline tractography to HARDI diffusion functions that clearly discern fiber crossings. We also extended the Tensorline algorithm to these HARDI functions, to improve on the extended Streamline tractography. We illustrated our proposed methods using the Solid Angle diffusion Orientation Distribution Function (ODF-SA). We presented results on multi-tensor synthetic data, and real in vivo data of the cerebral white-matter that show markedly improved tractography results.

This work has been published in [30].

6.1.3.2. Tract-based statistical analyzes in dMRI in autism spectrum disorder

Participants: Rachid Deriche, Anne-Charlotte Philippe, Demian Wassermann [Harvard Medical School, Boston, MA], Pablo Barttfeld [Integrative Neuroscience Laboratory, Physics Dept. University of Buenos Aires, Buenos Aires, Argentina], Jorge Calvar [Fundación para Lucha contra las Enfermedades Neurológicas del Niño, Buenos Aires, Argentina], Ramon Leiguarda [Fundación para la Lucha contra las Enfermedades Neurológicas del Niño, Buenos Aires, Argentina], Bruno Wicker [INCM CNRS, Marseille, France], Mariano Sigman [Integrative Neuroscience Laboratory, Physics Dept. University of Buenos Aires, Buenos Aires, Argentina].

Abnormal face processing is one of the hallmark features of social impairments in autism spectrum disorder (ASD). Previous neuroimaging studies showed that the fusiform gyrus is involved in face perception and is not or abnormally activated in autistic subjects. The aim of this study was to quantify potential anatomical differences in the white matter tracts that traverse the fusiform gyrus, the prefrontal cortex and the superior temporal gyrus, and correlate them with ADOS scores in ASD subjects. We used Diffusion Tensor MRI (DT) images to assess the integrity of automatically segmented white matter bundles connecting these brain areas. Then, we performed statistical analysis on these fiber bundles using diffusivity measures calculated from DT to characterize tissue microstructure changes and correlate these changes with ADOS scores. 7 adults with high functioning autism or Asperger syndrome and 11 typical adults participated in the study. We found several clusters with dissimilarities between ASD and control subjects in FA measures on tracts traversing the fusiform gyrus in both hemispheres of the brain. We observed a significant reduction of FA values in a cluster on a bundle joining the superior temporal gyrus to the prefrontal.
This work has been published in [37]. A related work on large-scale network analysis reflecting big-world characteristics in ASD has been published in [35].

6.1.3.3. Unsupervised automatic white matter fiber clustering using a Gaussian mixture model

Participants: Rachid Deriche, Meizhu Liu [Department of CISE, University of Florida, Gainesville, USA], Baba Vemuri [Department of CISE, University of Florida, Gainesville, USA].

Fiber tracking from diffusion tensor images is an essential step in numerous clinical applications. There is a growing demand for an accurate and efficient framework to perform quantitative analysis of white matter fiber bundles. In this work, we proposed a robust framework for fiber clustering. This framework is composed of two parts: accessible fiber representation, and a statistically robust divergence measure for comparing fibers. Each fiber is represented using a Gaussian mixture model (GMM), which is the linear combination of Gaussian distributions. The dissimilarity between two fibers is measured using the total square loss function between their corresponding GMMs (which is statistically robust). Finally, we performed the hierarchical total Bregman soft clustering algorithm on the GMMs, yielding clustered fiber bundles. Further, our method is able to determine the number of clusters automatically. We performed experimental results depicting favorable performance of our method on both synthetic and real data examples.

This work is currently under submission.

6.1.3.4. Riemannian geometry based brain white matter fiber clustering

Participants: Rachid Deriche, Ali Demir [Sabanci University, Istanbul, Turkey], Gozde Unal [Sabanci University, Istanbul, Turkey].

This work was partly supported by the PHC Bosphore grant.

Clustering of reconstructed brain white matter fibers into meaningful anatomical bundles becomes an important tool for detailed analysis of brain white matter diseases via diffusion MRI. In this work we developed a Riemannian geometry based geodesic distance measure between fiber pairs, which is then utilized in fiber clustering. A second contribution is a fiber selection algorithm, which compresses the dataset and speed up the computation time. We demonstrated methods on synthetic kissing fibers dataset, and validated on a brain white matter atlas.

This work is currently under submission.

6.1.3.5. White Matter Clustering Revisited

Participants: Rachid Deriche, Alvaro-Alejandro Sanchez-Moscosa.

This work was partly supported by the Association France Parkinson and the ANR NucleiPark project.

In [70], we have introduced an interesting hierarchical agglomerative based algorithm that represents and clusters white matter bundles under a Gaussian Process framework. In this work, a new implementation which drastically improves the performance of the clustering algorithm before mentioned is proposed and validated. This implementation notably improves the performance of the clustering algorithm [70] and has been validated on real data. This approach has the advantage of running in a lower abstraction level, which leads to lowered memory requirements, shorter running times and ultimately provides the possibility to process more densely seeded tractographies. The new implementation is then used to process densely seeded streamline tractographies which provide more localized fiber bundles. Additionally, as Parkinson’s Disease is believed to induce changes in the axonal bundles, tract-based statistical analysis is performed on interesting fiber tracts to find significant differences in diffusion scalar measures.

6.1.3.6. Using Radial NMR Profiles to Characterize Pore Size Distributions

Participants: Rachid Deriche, John Treilhard [Queen’s University, CA].
Extracting information about axon diameter distributions in the brain is a challenging task which provides useful information for medical purposes; for example, the ability to characterize and monitor axon diameters would be useful in diagnosing and investigating diseases like amyotrophic lateral sclerosis (ALS) or autism. In [75], three families of operators are defined, whose action upon an NMR attenuation signal extracts the moments of the pore size distribution of the ensemble under consideration; also a numerical method is proposed to continuously reconstruct a discretely sampled attenuation profile using the eigenfunctions of the simple harmonic oscillator Hamiltonian – the SHORE basis. The work we have performed here extends this method to other bases that can offer a better description of attenuation signal behaviour – in particular, we proposed the use of the radial Spherical Polar Fourier (SPF) basis. Testing is performed to contrast the efficacy of the radial SPF basis and SHORE basis in practical attenuation signal reconstruction. The robustness of the method to additive noise is tested and analyzed. We demonstrated that a low-order attenuation signal reconstruction outperforms a higher-order reconstruction in subsequent moment estimation under noisy conditions. We proposed the simulated annealing algorithm for basis function scale parameter estimation. Finally, analytic expressions are derived and presented for the action of the operators on the radial SPF basis (obviating the need for numerical integration, thus avoiding a spectrum of possible sources of error).

This work is currently under submission.

6.2. Brain functional imaging using MEG/EEG

6.2.1. EEG forward problem

6.2.1.1. OpenMEEG software library

Participants: Maureen Clerc, Emmanuel Olivi, Alexandre Gramfort, Théodore Papadopoulo.

This work was partly supported by the Regional Council of Provence Alpes Côte d'Azur and the ANR ViMAGE.

To recover the sources giving rise to electro- and magnetoencephalography in individual measurements, realistic physiological modeling is required, and accurate numerical solutions must be computed. The OpenMEEG software library solves the electromagnetic forward problem in the quasistatic regime, for head models with piecewise constant conductivity. The core of OpenMEEG consists of the symmetric Boundary Element Method, which is based on an extended Green Representation theorem. OpenMEEG is able to provide lead fields for four different electromagnetic forward problems: Electroencephalography (EEG), Magnetoencephalography (MEG), Electrical Impedance Tomography (EIT), and intracranial electric potentials (IPs). OpenMEEG is open source and multipurpose. It can be used from Python and Matlab in conjunction with toolboxes that solve the inverse problem; its integration within FieldTrip is operational since release 2.0.

Some new developments have concerned the organization of the computations to compute the lead fields. A lead-field matrix is obtained by concatenating the forward fields computed for thousands of sources characterized by their positions, orientations and strengths. A line of this lead-field matrix represents the physical quantity (potential for EEG, or some components of the magnetic field for MEG) at a sensor for each source. The number of sources largely exceeds the number of sensors (up to 256 electrodes for EEG, and less than 600 squids for MEG). When solving the forward problem with a BEM (Boundary Element Method), the lead-field matrix is generally computed column-by-column, i.e. source by source, which represents nsources resolutions of the forward problem. Using the adjoint operator of the forward problem, one can reduce the computations to sensors resolutions. Some previous works [72], [73] have used similar techniques for efficient computation of the lead-fields using finite element methods. The adjoint method [69] generalizes the Helmholtz reciprocity theorem and here is proposed its implementation using the BEM provided by the open-source software OpenMEEG.

This work has been published in [15] and [36].

6.2.1.2. Conductivity calibration for the EEG forward problem

Participants: Maureen Clerc, Emmanuel Olivi, Alexandre Gramfort, Théodore Papadopoulo, Jean-Michel Badier [INSERM U751, La Timone, Marseille], Martine Gavaret [INSERM U751, La Timone, Marseille], Laurent Koessler [CRAN Nancy].
Bioelectric phenomena at low temporal frequency can be described by the electrostatic Poisson equation

\[ \text{div } \sigma \nabla V = \text{div } J^p, \]

where \( J^p \) are primary current sources and \( \sigma \nabla V \) the Ohmic current. Appropriate boundary conditions (b.c.) must be set on the domain boundary, typically imposing the potential (Dirichlet b.c.) or the normal current (Neumann b.c.).

Solving this electrostatics equation for the electric potential is a problem common to different fields such as electroencephalography (EEG), electrocardiography (ECG), functional electrical stimulation. The main difficulty of the model concerns the conductivity field, which is not homogeneous, and whose values depend on the tissue type. Although the tissue structure can be revealed by imaging methods such as CT, Magnetic Resonance Imaging, diffusion Magnetic Resonance Imaging, conductivity values must nevertheless be assigned to the different tissues. In order to calibrate conductivity, injected current Electrical Impedance Tomography (EIT) can be used: it consists of injecting current on the outer boundary, and measuring the associated electric potential.

The thrust of OpenMEEG is to propose accurate forward problems, in several instances, including electro- and magneto-encephalography (EEG-MEG) and Electrical Impedance Tomography (EIT). OpenMEEG allows to compute the electric potential and magnetic fields due to boundary current injection or to sources within the compartments.

We apply this methodology to conductivity calibration in the context of high-density EEG pre-surgical epileptic exploration. High-density EEG measurements may be used to solve the inverse problem of source localization, in order to localize the foci of the epileptic activity within the brain. The solution of the inverse problem relies on a forward problem, linking sources to measurements. In turn, this forward problem is dependent on the conductivity of the head tissues.

Prior work has shown that the scalp-to-skull conductivity ratio is a sensitive parameter for source localization, because it has an influence on the depth of the estimated dipoles \[ 68 \]. There have been several studies demonstrating the feasibility of injected-current Electrical Impedance Tomography to calibrate the scalp-to-skull conductivity ratio \[ 57 \], \[ 49 \]. We are currently collaborating with our partners from La Timone in Marseille in a clinical assessment of the use of injected current Electrical Impedance Tomography to calibrate the scalp-to-skull conductivity ratio.

This work has been presented at the international conference on Electrical Impedance Tomography, see \[ 27 \] and \[ 28 \].

### 6.2.1.3. Coupling numerical methods for the forward problem

**Participants:** Maureen Clerc, Emmanuel Olivi, Théodore Papadopoulo, Mariette Yvinec [Geometrica Project-Team, INRIA Sophia Antipolis Méditerranée].

This work was partly supported by the ANR grant ViMAGINE.

Electroencephalography (EEG) and magnetoencephalography (MEG) are two modalities that aim at measuring brain activity. Source localization from external data such EEG or MEG, requires a good understanding of the electromagnetic behavior of the patient head. Several models can be used, representing more or less complex geometrical shapes, and conductivity profiles. Different numerical methods allow to cope with different types of models: the Finite Element Method (FEM) can handle very general conductivity models, whereas the Boundary Element Method (BEM) is limited to piecewise constant conductivity. On the other hand, BEM is more capable than FEM to accurately represent sources in isotropic media.
Using a domain decomposition approach, we propose to independently use BEM or FEM in different sub-domains. In the EEG forward problem considered, the BEM is limited to the domain where the sources lie (the brain) while the other tissues are handled with the FEM. This leads to an accurate description of the sources while allowing for inhomogeneous and anisotropic conductivity. The proposed method is first validated against analytical solutions in multi-sphere models. Results of the forward problem are presented for a four-layer realistic head-model incorporating a burr-hole in the skull. Convergence of the iterative algorithm is analyzed numerically. The domain decomposition framework provides a way of taking the best advantage of both methods, thus significantly improving the accuracy in the resolution of the forward EEG problem, as well as time and memory consumption.

This work is part of Emmanuel Olivi’s PhD thesis [11], and a journal article is under submission.

6.2.1.4. White matter anisotropy

**Participants:** Maureen Clerc, Emmanuel Olivi, Alexandre Gramfort, Théodore Papadopoulo.

*This work was partly supported by the ANR grant ViMAGINE.*

Conductivity of tissues in the vicinity of the sources is especially influential on the MEG and EEG forward fields. Those tissues include white matter, whose conductivity is anisotropic because of its fiber structure. While white matter anisotropy can be measured thanks to Diffusion-Weighted MRI, it is rarely incorporated in MEG and EEG head models. Boundary Element Methods can only deal with piecewise constant conductivities, therefore ruling out white matter anisotropy that has a complex structure, and Finite Element Methods have been developed to deal with anisotropic conductivity, but require very fine meshes, thus huge linear systems. We have extended the BEM framework to incorporate white matter anisotropy by treating anisotropic conductivity as a perturbation of an isotropic one. With this extension we have been able to compute the influence of a fiber within the brain white matter on the electric potential, and to validate the results by comparison with a Finite Element Method.

This work has been published in [34].

6.2.2. Inverse problems in MEG and EEG

6.2.2.1. Rational Approximations

**Participants:** Maureen Clerc, Théodore Papadopoulo, Juliette Leblond [APICS Project Team, INRIA Sophia Antipolis Méditerranée], Jean-Paul Marmorat [Centre de Mathématiques Appliquées, Ecole des Mines].

In functional neuroimaging, a crucial problem is to localize active sources within the brain non-invasively, from the knowledge of the electromagnetic measurements outside the head. Identification of point sources from boundary measurements is an ill-posed inverse problem. In the case of electroencephalography (EEG), measurements are only available at electrode positions, the number of sources is not known in advance, and the medium within the head is inhomogeneous. We pursue our ongoing work on EEG source localization, based on rational approximation techniques in the complex plane. The method is used in the context of a nested sphere head model, in combination with a cortical mapping procedure. Results on simulated data prove the applicability of the method in the context of realistic measurement configurations.

This work has been submitted to a journal and published as a Research Report in [38].

6.2.3. Brain Computer Interfaces

6.2.3.1. New features for Motor Imagery

**Participants:** Maureen Clerc, Joan Fruittet, Théodore Papadopoulo, Eoin Thomas.

*This work was partly supported by ANR grant CoAdapt.*

Our goal is to build a training free BCI based on beta rebound detection and discrimination during the first stage of use, while the learning of the conventional sensorimotor rhythms is done. We show in this preliminary study that it is possible to use the beta rebound to discriminate, real and imagined, right versus left hand movement with either no or very little training.
6.2.3.2. A bandit algorithm for exploring mental imagery

Participants: Maureen Clerc, Joan Fruite, Alexandra Carpentier [Sequel Project Team, INRIA Lille Nord Europe], Rémi Munos [Sequel Project Team, INRIA Lille Nord Europe].

This work was partly supported by ANR grant Co-Adapt.

This study presents a new procedure to automatically select a discriminant motor task for an asynchronous brain-controlled button. This type of control pertains to Brain Computer Interfaces (BCI). When using sensorimotor rhythms in a BCI, several motor tasks, such as moving the right or left hand, the feet or the tongue, can be considered as candidates for the control. This report presents a method to select as fast as possible the most promising task. We develop for this purpose an adaptive algorithm UCB-classif based on the stochastic bandit theory and build an EEG experiment to test our method. By not wasting time on inefficient tasks, our algorithm can focus on the most promising ones, resulting in a faster task selection and a more efficient use of the BCI training session. This leads to better classification rates for a fixed time budget, compared to a standard task selection.

This work has been published in [39] and is currently under submission to a journal.

6.3. Multi-Imaging Modalities

6.3.1. Coupling neuronal and haemodynamic models

6.3.1.1. Modeling of the neurovascular coupling in epileptic discharges

Participants: Maureen Clerc, Théodore Papadopoulos, Nicole Yoges [former Athena postdoc], Christian Bénar [INSERM U751 Marseille], Solenna Blanchard [INSERM U642 Rennes], Fabrice Wendling [INSERM U642 Rennes], Habib Benali [INSERM U678 Paris], Olivier David [INSERM U594 Grenoble].

This work was partly supported by ANR MultiModel.

Despite the interest in simultaneous EEG-fMRI studies of epileptic spikes, the link between epileptic discharges and their corresponding hemodynamic responses is poorly understood. In this context, biophysical models are promising tools for investigating the mechanisms underlying observed signals. We have applied a metabolic-hemodynamic model to simulated epileptic discharges, in part generated by a neural mass model. We analyzed the effect of features specific to epileptic neuronal activity on the blood oxygen level dependent (BOLD) response, focusing on the issues of linearity in neurovascular coupling and on the origin of negative BOLD signals. We found both sub- and supra-linearity in simulated BOLD signals, depending on whether one observes the early or the late part of the BOLD response. The size of these non-linear effects is determined by the spike frequency, as well as by the amplitude of the excitatory activity. Our results additionally indicate a minor deviation from linearity at the neuronal level. According to a phase space analysis, the possibility to obtain a negative BOLD response to an epileptic spike depends on the existence of a long and strong excitatory undershoot. Moreover, we strongly suggest that a combined EEG-fMRI modeling approach should include spatial assumptions. The present study is a step towards an increased understanding of the link between epileptic spikes and their BOLD responses, aiming to improve the interpretation of simultaneous EEG-fMRI recordings in epilepsy.

This work has been published in [17].

6.3.1.2. A nested cortex parcellation combining analysis of MEG forward problem and diffusion MRI tractography

Participants: Anne-Charlotte Philippe, Maureen Clerc, Théodore Papadopoulos, Rachid Deriche.
Understanding the relationship between structure and function is a major challenge in neuroscience. Diffusion MRI (dMRI) is the only non-invasive modality allowing to have access to the neural structure. Magnetoencephalography (MEG) is another non-invasive modality that allows a direct access to the temporal succession of cognitive processes. Functional cortex parcellation being one of the most important ways to understanding structure-function relationship, we propose an innovative method merging MEG and dMRI to parcellate the cortex. The combination of MEG forward problem and connectivity information reveals cortical areas generating a similar magnetic field at sensors while having a similar connectivity. Results show suitable clusters that forecast interesting studies for inter- and intra-subject comparisons of the cortex parcellations. The automatic nested cortex parcellation we propose could be a first step to analyse sources that are seeds of long or short range connectivity and to differentiate these connectivities in the white matter.

This work is currently under submission.

6.3.1.3. Improved computer-aided detection of small polyps in CT colonography using interpolation for curvature estimation

Participants: Rachid Deriche, Jiamin Liu [Imaging Biomarkers and Computer-Aided Diagnosis Laboratory, Radiology and Imaging Sciences, Clinical Center, National Institutes of Health, Bethesda], Suraj Kabadi [Imaging Biomarkers and Computer-Aided Diagnosis Laboratory, Radiology and Imaging Sciences, Clinical Center, National Institutes of Health, Bethesda], Robert Van Uitert [Imaging Biomarkers and Computer-Aided Diagnosis Laboratory, Radiology and Imaging Sciences, Clinical Center, National Institutes of Health, Bethesda], Nicholas Petrick [Center for Devices and Radiological Health, U.S. Food and Drug Administration, Maryland], Ronald M. Summers [Imaging Biomarkers and Computer-Aided Diagnosis Laboratory, Radiology and Imaging Sciences, Clinical Center, National Institutes of Health, Bethesda].

Surface curvatures are important geometric features for the computer-aided analysis and detection of polyps in CT colonography (CTC). However, the general kernel approach for curvature computation can yield erroneous results for small polyps and for polyps that lie on haustral folds. Those erroneous curvatures will reduce the performance of polyp detection. This work presents an analysis of interpolation’s effect on curvature estimation for thin structures and its application on computer-aided detection of small polyps in CTC. In this work, we demonstrated that a simple technique, image interpolation, can improve the accuracy of curvature estimation for thin structures and thus significantly improve the sensitivity of small polyp detection in CTC. Our experiments showed that the merits of interpolating included more accurate curvature values for simulated data, and isolation of polyps near folds for clinical data. After testing on a large clinical data set, it was observed that sensitivities with linear, quadratic B-spline and cubic B-spline interpolations significantly improved the sensitivity for small polyp detection. In conclusion, the image interpolation can improve the accuracy of curvature estimation for thin structures and thus improve the computer-aided detection of small polyps in CTC.

This work has been published in [16].
6. New Results

6.1. Mathematical methods and methodological approach to biology

6.1.1. Mathematical analysis of biological models

Participants: Jean-Luc Gouzé, Olivier Bernard, Frédéric Grognard, Ludovic Mailleret, Pierre Bernhard, Andrei Akhmetzhanov.

Mathematical study of semi-discrete models

Semi-discrete models have shown their relevance in the modelling of biological phenomena whose nature presents abrupt changes over the course of their evolution [77]. We used such models and analysed their properties in several situations that are developed in 6.2.3, most of them requiring such a modelling in order to take seasonality into account. Such is the case when the year is divided into a cropping season and a 'winter' season, where the crop is absent, as in our analysis of the sustainable management of crop resistance to pathogens [58] or in the co-existence analysis of epidemiological strains [17], [21]. Seasonality also plays a big role in the semi-discrete modelling required for the analysis of consumers' adaptive behaviour in seasonal consumer-resource dynamics, where only dormant offspring survives the ‘winter’ [12], [52].

Mathematical study of models of competing species

When several species are in competition for a single substrate in a chemostat, and when the growth rates of the different species only depend on the substrate, it is known that the generic equilibrium state for a given dilution rate consists in the survival of only one of the species. In [28], we propose a model of competition of \( n \) species in a chemostat, where we add constant inputs of some species. We achieve a thorough study of all the situations that can arise when having an arbitrary number of species in the chemostat inputs; this always results in a Globally Asymptotically Stable equilibrium where all input species are present with at most one of the other species.

6.1.2. Model design, identification and validation

Participants: Olivier Bernard, Jean-Luc Gouzé.

Model design and identification

One of the main families of biological systems that we have studied involves mass transfer between compartments, whether these compartments are microorganisms or substrates in a bioreactor, or species populations in an ecosystem. We have developed methods to estimate the models of such systems [62]. These systems can be represented by models having the general structure popularized by [61], [65], [66], and based on an underlying reaction network:

\[
\frac{d\xi}{dt} = Kr(\xi, \psi) + D(\xi_{in} - \xi) - Q(\xi)
\]

We address two problems: the determination of the pseudo-stoichiometric matrix \( K \) and the modeling of the reaction rates \( r(\xi, \psi) \).

In order to identify \( K \), a two-step procedure has been proposed. The first step is the identification of the minimum number of reactions to be taken into account to explain a set of data. If additional information on the process structure is available, we showed how to apply the second step: the estimation of the pseudo-stoichiometric coefficients.

This approach has been applied to various bioproduction processes, most recently on activated sludge processes [60], anaerobic digestion [71], [87] and anaerobic digestion of microalgae [22].
6.1.3. Nonlinear observers

**Participants:** Jean-Luc Gouzé, Olivier Bernard.

*Interval observers*

Interval observers give an interval estimation of the state variables, provided that intervals for the unknown quantities (initial conditions, parameters, inputs) are known [8]. We have extended the interval observer design to new classes of systems. First, we designed interval observers, even when it was not possible in the original basis, by introducing a linear, time-varying change of coordinates, [80]. This approach was then extended to \( n \)-dimensional linear systems, leading to the design of interval observers in high dimensions [25]. Extension to time-delay systems have also been proposed [26]. The combination of the observers has also been improved in the case where various types of interval observers are run in parallel in a so-called “bundle of observers” [64]. The approach has been applied to estimation of the microalgal growth and lipid production [39].

In order to demonstrate the efficiency of the interval observer design, even with chaotic systems, a special application of the interval observer has been developed for Chua’s chaotic systems. The interval estimation of the state variables are performed considering parameters uncertainties of the system and biased output [83], [80].

6.1.4. Metabolic and genomic models

**Participants:** Jean-Luc Gouzé, Madalena Chaves, Alfonso Carta, Ismail Belgacem, Xiao Dong Li, Olivier Bernard, Christian Breindl, Frédéric Grognard.

*Qualitative control of piecewise affine models*

In the control of genetic networks, the construction of feedback control laws is subject to many specific constraints, including positivity, appropriate bounds and forms of the input. In addition, control laws should be liable to implementation in the laboratory using gene and protein components. In this context, under the hypothesis that both the observations and control functions are qualitative (or piecewise constant), and using sliding mode solutions, we analysed the controllability and stabilizability with respect to either of the steady states, for a piecewise affine model of the bistable switch with single input. It is also possible to find a qualitative control law that leads the system to a periodic orbit passing through the unstable steady state [14]. Moreover, we designed some preliminary control laws for the negative loop with two genes, which has an oscillatory behaviour [54].

*Interconnections of Boolean modules: asymptotic behaviour*

A biological network can be schematically described as an input/output Boolean module: that is, both the states, the outputs and inputs are Boolean. The dynamics of a Boolean network can be represented by an asynchronous transition graph, whose attractors describe the system’s asymptotic behaviour. We have shown that the attractors of the feedback interconnection of two Boolean modules can be fully identified in terms of cross-products of the attractors of each module. Based on this result, a model reduction technique is proposed in [33], to predict the asymptotic dynamics of high-dimensional biological networks through the computation of the dynamics of two isolated smaller subnetworks. Applications include a model of cell-fate decision (represented as an interconnection of two 3-input/3-output modules).

*Structure estimation for unate Boolean models of gene regulation networks*

Estimation or identification of the network of interactions among a group of genes is a recurrent problem in the biological sciences. Together with collaborators from the University of Stuttgart, we have worked on the reconstruction of the interaction structure of a gene regulation network from qualitative data in a Boolean framework. The idea is to restrict the search space to the class of unate functions. Using sign-representations, the problem of exploring this reduced search space is transformed into a convex feasibility problem. The sign-representation furthermore allows to incorporate robustness considerations and gives rise to a new measure which can be used to further reduce the uncertainties. The proposed methodology is demonstrated with a Boolean apoptosis signaling model [70].
Multistability and oscillations in genetic control of metabolism

Genetic feedback is one of the mechanisms that enables metabolic adaptations to environmental changes. The stable equilibria of these feedback circuits determine the observable metabolic phenotypes. Together with D. Oyarzun from Imperial College, we considered an unbranched metabolic network with one metabolite acting as a global regulator of enzyme expression. Under switch-like regulation and exploiting the time scale separation between metabolic and genetic dynamics, we developed geometric criteria to characterize the equilibria of a given network. These results can be used to detect mono- and bistability in terms of the gene regulation parameters for any combination of activation and repression loops [40]. We also find that metabolic oscillations can emerge in the case of operon-controlled networks; further analysis reveals how nutrient-induced bistability and oscillations can emerge as a consequence of the transcriptional feedback [27].

Uniqueness and global stability for metabolic models

We are interested in the uniqueness and stability of the equilibrium of reversible metabolic models. For biologists, it seems clear that realistic metabolic systems have a single stable equilibrium. However, it is known that some type of metabolic systems can have no or multiple equilibria. We have made some contribution to this problem, in the case of a totally reversible enzymatic system. We prove that the equilibrium is globally asymptotically stable if it exists; we give conditions for existence and behaviour in a more general genetic-metabolic loop [84].

Birhythmicity in the p53-Mdm2 network

The p53-Mdm2 network is one of the key protein module involved in the control of proliferation of abnormal cells in mammals. Recently, a differential model of the p53-Mdm2 biochemical network which shows birhythmicity has been proposed to reproduce the two experimentally observed frequencies of oscillations of p53. Our study aimed at investigating the mechanisms at the origin of this birhythmic behaviour. To do so, we approximated this continuous non-linear model into a lower dimensional piecewise affine model and performed a first return map analysis. Based on this analysis, an experimental strategy has been proposed to test the existence of birhythmicity in the p53-Mdm2 network [31], [11].

E. coli modelling and control

In the framework of ANR project Gemco, with the aim of better understanding how to build a controller for Escherichia coli growth, three reduced models of E. coli gene expression machinery have been developed: the wild-type model, the open-loop model and the closed-loop model. Each one of these models is made up of two piece-wise non-linear differential equations.

Notably, the wild-type model describes the qualitative dynamics of the unmodified bacterium in terms of two relevant protein concentrations (RNAP and CRP) and a carbon source as input, which can be either glucose or maltose. Bacteria prefer glucose, which leads to a higher growth rate in wild type.

The open-loop model describes the qualitative dynamics of RNAP and CRP concentrations when the gene encoding for RNAP is controlled externally by an IPTG-inducible promoter. This control, biologically implemented by our collaborators in Grenoble—yielding different E. coli growth rates depending on IPTG concentration—shows that a controller for the bacterial growth can be built acting on the gene expression machinery level. Moreover, lumped parameters related to RNAP dynamics and IPTG regulation function have been estimated by means of E. coli growth curves.

Finally, the closed-loop model implements a possible feedback control-loop able to theoretically generate inverse diauxie, i.e. higher growth on maltose than on glucose.

Observation problems of a class of genetic regulatory networks

A state reconstruction problem with Boolean measurements is considered for a piecewise affine genetic network model. The problem has two distinct aspects with respect to classical ones: 1) the model is a hybrid system, and 2) the measurements (of genes expression) are only qualitative due to the experimental techniques. A Luenberger-like observer is proposed which can present some sliding modes and has finite-time convergence. A transition graph is given for the coupled observer-nominal system. To minimize the
convergence time, different convergence scenarios are discussed for optimizing the choice of initial condition of the observer \[35\], \[57\]. Robustness of the observer is checked for two types of parametric perturbed systems: 1) the observer is used to identify an unknown but fixed variation on the synthesis coefficient, via an adaptive dichotomy algorithm; and 2) the observer is robust in practical sense for the model with an uncertainty on the threshold value \[76\].

6.2. Fields of application

6.2.1. Bioenergy

6.2.1.1. Modelling and optimization of microalgae production

**Participants:** Olivier Bernard, Antoine Sciandra, Frédéric Grognard, Philipp Hartmann, Rafael Munoz Tamayo, Andrei Akhmetzhanov, Nina Moelants, Hubert Bonnefond.

Experiments have been carried out to study the effects of nitrogen limitation on the lipid production in microalgae \[74\] and support model development. We have proposed a new model for lipid production by microalgae which describes the fate of the CO$_2$ incorporated during photosynthesis \[23\]. This model describes the accumulation of neutral lipids (which can be turned into biofuel), carbohydrates and structural carbon. It has been calibrated and validated with experimental data. Experiments have also attempted to simultaneously represent the effect of an osmotic stress \[55\]. This model highlights and explains the phenomenon of hysteresis in lipid production which has been experimentally verified. It has been extended to account for light/dark cycles \[36\].

On the other hand, a new dynamical model has been developed to describe microalgal growth in a photobioreactor under light and nitrogen limitations \[13\]. The strength of this model is that it takes into account the strong interactions between the biological phenomena (effects of light and nitrogen on growth, photoacclimation \[34\], \[48\] ...) and the radiative transfer in the photobioreactor (light attenuation due to the microalgae). Using these two approaches, we have developed a model which describes lipid production in a photobioreactor under light limitation. This model is used to predict lipid production in the perspective of large scale biofuel production. Simpler models have also been developed and have been used to provide optimization strategies: first, biomass production has been optimized in a constant light environment \[79\], yielding results emphasizing the importance of the optical depth of the reactor. In a second work, we focused on the optimal operating conditions for the biomass productivity under day/night cycles using Pontryagin’s maximum principle (assuming a periodic working mode) \[72\], \[73\].

Another model has been developed to represent growth of microalgae colimited by nitrogen and phosphorus \[69\]. It has been shown, from qualitative analysis of the model that uptake of nitrogen and phosphorus are non symmetric.

6.2.1.2. Modelling the effect of light and temperature on microalgae

**Participants:** Olivier Bernard, Antoine Sciandra, Frédéric Grognard, Philipp Hartmann, Rafael Munoz Tamayo, Kerstin Ebert, Nina Moelants, Hubert Bonnefond.

The light distribution within a photobioreactor was estimated thanks to a multi photon Monte-Carlo simulation. From measurements of absorption and scattering properties, it was thus possible to extrapolate and validate the light distribution within a photobioreactor or a raceway.

The impact of the hydrodynamics on the light percept by a single cell was studied thanks to fluid dynamics simulations of raceway pond \[48\]. The light signals that a cell experiences at the Lagrangian scale, depending on the fluid velocity, were then estimated. A Droop-Han model was used to assess the impact of light fluctuation on photosynthesis \[48\].

Finally, the effect of temperature on microalgae has been represented by adapting the CTMI model developed for bacteria \[88\]. The proposed model \[59\], associated with a parameter identification procedure, was able to correctly represent the growth response to temperature for 12 different species \[48\].
6.2.2. CO₂ fixation by microalgae

Participants: Olivier Bernard, Antoine Sciandra, Philipp Hartmann, Nina Moelants.

We have run experiments to observe the response of a population of microalgal cells to various periodic light/dark or nitrate signals. The measurements show the synchronicity of the cells for some conditions. These experiments support the hypothesis that uptake of nitrogen stops during cell division [82]. On this basis, we have developed a structured model representing the development of microalgal cells through three main phases of their cell cycle: G1, G2 and M. The model is made of three interdependent Droop models [13]. The model was validated through extensive comparison with experimental results in both condition of periodic light forcing and nitrogen limitation. The model turns out to accurately reproduce the experimental observations [81]. The effect of cell synchronization on lipid content were experimentally studied [18] and included into microalgae growth models [36].

The effect of CO₂ partial pressure increase on photosynthesis and calcification of the calcareous microalgae *Emiliania huxleyi* have been experimentally observed. It results that an increase of the coccolith size together with a decrease in the calcification rate has been observed.

Three models accounting for the possible coupling between photosynthesis and calcification were included in an ocean model, including settling and predation by grazers, and a bloom of coccolithophorids was simulated [67], [68]. It was shown using Monte Carlo simulations that the uncertainty on the mechanisms driving calcification together with parametric uncertainties lead to uncertainties which are in the same range as the effect of an increase or the CO₂ partial pressure.

6.2.3. Design of ecologically friendly plant production systems

6.2.3.1. Controlling plant pests

Participants: Frédéric Grognard, Ludovic Mailleret, Mickaël Teixeira-Alves.

*The influence of an alternative prey in biological control programs*

We have developed a model based on the classical Leslie-Gower predator prey model, that allows for the choice that a predator might have for its diet. In a biological control framework, this choice might be between a pest that we want to eradicate and another prey that could be fed to the predator in order to help the biological control efficiency or between the pest and an alternative prey that is present in the field and might keep the predator from acting as a natural enemy of the pest. We put the problem in a time-partitioning framework: the predator has to split its time between the two prey. We then compared two time-partitioning strategies: one where the predator always spends a fixed proportion of its time on each prey and one where the predator always chooses the prey that is instantaneously most profitable (adaptive foraging). We then studied the effect of the presence of one prey on the other (indirect effect since it is mediated by the presence of the predator). We showed that, in the Leslie-Gower framework, one of the two prey always benefits from the presence of the other and that this effect is even stronger in the adaptive foraging framework, where the presence of the other prey is never detrimental to the one considered. That way, with very little assumptions, we showed the existence of apparent predation, commensalism and apparent mutualism, while most existing theoretical results tend to evidence apparent competition [51]. Such mechanisms may explain why generalist biological control agents are, in general, not as efficient as specialists in controlling crop pests.

*The influence of plant dynamics on pest eradication*

Pests-biocontrol agents models have been developed in order to build biological control strategies that can achieve pest eradication through augmentative biological control [85]. In the present work, we aim at introducing a plant compartment since its dynamics clearly have an influence on that of the pests and since the final objective of biological control is to maintain a sufficient plant yield. In a first step, we focused on plant-insect interactions and showed how the level and timing of the pest invasion could influence the final plant yield. Introducing pests control interventions and studying its timing, we showed how it eventually could have important effects on the growth pattern and the final biomass. As a reference, we consider the novel invasive pest Tuta absoluta on tomato plants [56].
6.2.3.2. Controlling plant pathogens

Participants: Frédéric Grognard, Ludovic Mailleret.

Sustainable management of plant resistance

The introduction of plants strains that are resistant to one pathogen often leads to the appearance of virulent pathogenic strains that are capable of infecting the resistant plants. The resistance strain then becomes useless. It is therefore necessary to develop ways of introducing such resistance into crop production without jeopardizing its future efficiency. We did so by choosing the proportion of resistant plants that are mixed with the non-resistant ones. In this work, we studied a vector borne pathogen in a seasonal environment, with healthy crop being planted at the beginning of each season and cropped at its end, the pathogen surviving in the environment during the ‘winter’. Two strategies have been proposed, one that aims at minimizing the cumulated damage over a 15 years horizon and one that aims at preventing the virulent strain outbreak. In the first case, both plant strains need to be mixed, but it results in the loss of the resistance at the end of the 15 year period; in the second case, the damage is higher and the maximal proportion of resistant plant is smaller, but the resistance is preserved [58], [16].

This work is done in collaboration with Frédéric Fabre (INRA Avignon).

Plant pathogen dynamics and cropping management practices

The coexistence of closely related plant parasites is widespread. Yet, understanding the ecological determinants of evolutionary divergence in plant parasites remains an issue. Niche differentiation through resource specialization has been widely researched, but it hardly explains the coexistence of parasites exploiting the same host plant. Agricultural systems are characterized by the cyclical presence and absence of the crop, due to cropping practices such as harvest and planting. We studied the influence that time-partitioning, i.e. the specialization of a parasite for the beginning or the end of crop presence, can have on co-existence. In modelling the epidemiology through a semi-discrete model we showed through an evolutionary invasion analysis that evolutionary divergence, and thus co-existence of different strains, of the parasite phenotype can occur [17], [44]. Also, in a similar context, we underlined why modelling seasonal plant epidemiology did not necessarily lead to competitive exclusion; indeed, generating a compact model by rigorously isolating the slow dynamics from a large detailed model of plant epidemiology, we found out the possibility of coexistence [49], [21]. Such a result contrasts with classical competitive exclusion principles found in compact models which rely on the arguable density independent nature of the pathogen infections occurring during the very beginning of the cropping seasons.

This work is done in collaboration with Frédéric Hamelin (Agrocampus Ouest).

6.2.4. Biological depollution

6.2.4.1. Coupling microalgae to anaerobic digestion

Participants: Olivier Bernard, Antoine Scandra, Jean-Philippe Steyer, Frédéric Grognard, Philipp Hartmann.

The coupling between a microalgal pond and an anaerobic digester is a promising alternative for sustainable energy production and wastewater treatment by transforming carbon dioxide into methane using light energy. The ANR Symbiose project is aiming at evaluating the potential of this process [90], [89].

In a first stage, we developed models for anaerobic digestion of microalgae. Two approaches were used: First, a dynamic model has been developed trying to keep a low level of complexity so that it can be mathematically tractable for optimisation [37], [32], [22]. Considering three main reactions, this model fits adequately the experimental data of an anaerobic digester fed with Chlorella vulgaris (data from INRA LBE). On the other hand, we have tested the ability of ADM1 [91] (a reference model which considers 19 biochemical reactions) to represent the same dataset. This model, after modification of the hydrolysis step [24], [38], [41] has then been used to evaluate process performances (methane yield, productivity...) and stability through numerical simulations.
In a second stage, a model describing the coupling between anaerobic digestion process and microalgae culture (including the feeding of the algae with anaerobic digestion effluents) has been developed. The model is based on the three steps model for anaerobic digestion, and on the photoacclimation model for microalgae [13]. The model also includes the modelling of heterotrophs in the microalgae pond.

6.2.4.2. Life Cycle Assessment of microalgae production

**Participants:** Olivier Bernard, Jean-Philippe Steyer.

This work is the result of a collaboration with Laurent Lardon and Arnaud Helias of INRA-LBE through the co-supervision of Pierre Collet’s PhD thesis.

An analysis of the potential environmental impacts of biodiesel production from microalgae has been carried out using the life cycle assessment (LCA) methodology [75]. This study has allowed to identify the obstacles and limitations which should receive specific research efforts to make this process environmentally sustainable. This study has been updated and the effects of technological improvements (leading to higher productivities) have been compared to the source of electricity. It turns out that the overall environmental balance can much more easily be improved when renewable electricity is produced on the plant [47], [46]. As a consequence, a new paradigm to transform solar energy (in the large) into transportation biofuel is proposed, including a simultaneous energy production stage.

A LCA has been carried out to assess the environmental impact of methane production by coupling microalgae and anaerobic digestion. The study highlights the limitation derived by the low biodegradability of the considered microalgae [15] which induces a large digester design and thus more energy to mix and heat it.

6.2.5. Models of ecosystems

6.2.5.1. Adaptive behaviour in seasonal consumer-resource dynamics

**Participants:** Frédéric Grognard, Ludovic Mailleret, Pierre Bernhard, Andrei Akhmetzhanov.

In this work we studied the evolution of a consumer-resource (or predator-prey) system with seasonal character of the dynamics. We specified two main parts of the process. First, we considered the system during one season with a fixed length: the prey lay eggs continuously and the predators lay eggs or hunt the preys (choose their behaviour) according to the solution of an optimal control problem [12]. Secondly, we studied the long-scale discrete dynamics over seasons. We investigated the qualitative behaviour of the dynamics with respect to the parameters of the problem and showed that, depending on the parameters of the model, extinction or co-existence of the predators and preys can be evidenced [12].

We then examined how (resident) predators adopting this behaviour would fare when faced with a small population of selfish mutants that would be identical to the resident but would have the freedom to choose a different behaviour. We studied the resulting optimal control problem where the mutants maximize their own number of offspring using the knowledge of the resident’s behaviour, and showed that, in most situations, mutants can take advantage of their low frequency and fare better than the residents. Over the course of a large number of seasons, the mutants replace the residents, only to find themselves applying the original resident behaviour [52]. We have then proposed a strategy for the predator in which it would prevent the invasion by the mutant instead of maximizing its number of offspring, which corresponds to the computation of evolutionarily stable life history strategies.

We have then considered that the resource itself could adapt its behaviour over time to limit the damage caused by the consumer, and maximize its own offspring. This problem requires the solution of a non zero-sum differential game, the consumer and the resource being the two players. We showed that the patterns of the strategies of the consumers and the resources are identical to the ones that can be obtained if the opposing player adopts a constant behaviour; the timing of the switchings varies however [42].

6.2.5.2. Including phytoplankton photoadaptation into biogeochemical models

**Participant:** Olivier Bernard.
The complexity of the marine ecosystem models and the representation of biological processes, such as photoadaptation, remain open questions. We compared several marine ecosystem models with increasing complexity in the phytoplankton physiology representation in order to assess the consequences of photoadaptation model complexity in biogeochemical model predictions. Three models of increasing complexity were considered, and the models were calibrated to reproduce ocean data acquired at the Bermuda Atlantic Time-series Study (BATS) from in situ JGOFS data. It turns out that the more complex model are trickier to calibrate and that intermediate complexity models, with an adapted calibration procedure, have a better prediction capability [43].

6.2.5.3. Growth models of zooplankton

Participants: Jean-Luc Gouzé, Jonathan Rault, Eric Benoît.

The model built to describe a zooplankton community is some variant of the McKendrick-Von Foerster Equation. The model includes cannibalism within zooplankton and predation on phytoplankton. Dynamic mass budget theory is used in order to describe individual behaviour and allows mass conservation. Also we have added phytoplankton dynamics, and we use environmental data as an input for the model. The aim is to compare simulations with data provided by the Laboratoire d'Océanographie de Villefranche. Since the model incorporates lots of parameters, which are not always known in the literature, we have to use optimization techniques to find them. Further, equilibria of such models and their local stability is studied in using strongly continuous semigroup approach [50]

We have also built a discrete size-structured model. Discrete models are less numerically demanding and so can be more easily incorporated into bigger models. Moreover the study of discrete models are often easier than that of continuous ones. We focus our study on the impact of cannibalism within zooplankton community and show that under some hypotheses, cannibalism can stabilize the equilibrium of the model [86].

6.3. Software design

Participants: Olivier Bernard, Florian Guenn, Mélaine Gautier.

Over the years, BIOCORE has been developing a software framework for bioprocess control and supervision called ODIN [63]. This C++ application (working under Windows and Linux) enables researchers and industrials to easily develop and deploy advanced control algorithms through the use of a Scilab interpreter. It also contains a Scilab-based process simulator which can be harnessed for experimentation and training purposes. ODIN is primarily developed in the C++ programming language and uses CORBA to define component interfaces and provide component isolation. ODIN is a distributed platform, enabling remote monitoring of the controlled processes as well as remote data acquisition. Recently, a software development effort has been directed to the graphical user interface, a synoptic view component, new drivers for the experimental hardware and integration of the PlantML data exchange format. ODIN has been tested on four different processes and currently supervises the 66m$^2$ high rate pond at the LBE, INRA Narbonne.
6. New Results

6.1. Modelling and Identification

6.1.1. Multi-Dimensional Wrist Musculoskeletal Modeling for Tremor Simulation

Participants: Peng Yao, Mitsuhiro Hayashibe, Dingguo Zhang (Shanghai Jiao Tong Univ.).

In this work, we established multi-dimensional wrist musculoskeletal model to be used for suppressing the wrist joint’s tremor by functional electrical stimulation. Often the wrist model for FES control is based on 1DOF biomechanical model for the simplicity and convenience to develop the controller. However, wrist motion is generated by complex interactions of multiple muscles spanning the wrist joint. Here, we have tried to have 3DOF wrist model considering main muscles involved in flex-extension, radial-ulnar deviation, and pron-supination as in the left of Fig. 1 (opensim model). Inertia of 3DOF joint model was obtained based on the work of de Leva which modified Zatsiorsky parameters. Joint dynamics was formed by inertia, gravitational torque, and passive visco-elastisity. As for the mapping between torque and muscles, the moment arm matrix \((3 \times 4)\) was obtained from opensim software, thus anatomical information could be considered along with muscle parameters such as muscle-length, isometric maximal forces. For this first trial, 4 muscles were considered based on Hill-type model. In order to confirm the generated motion by muscle activation, we activated the ECRB (extensor carpi radialis brevis) and FCU (flexor carpi ulnaris) in antiphase, the corresponding 3DOF wrist angles were obtained as in Fig. 1. Qualitatively, we know that ECRB can generate negative flexion angle, negative deviation angle and positive pronation when it is activated; while FCU can produce positive flexion angle, positive deviation angle and positive pronation angle. When both ECRB and FCU are activated in order, the reasonable wrist angles could be generated. We will work on the development of tremor controller in the future work.

![Figure 1. 3DOF Wrist Musculoskeletal Modeling. Anatomical layout of principle wrist muscles, moment arm matrix was obtained from opensim (left). 3DOF wrist angles were generated by tremor like activation in ECRB and FCU (right).](image)

6.1.2. Multi-functional EMG classification for dynamic EMG-motion modeling

Participants: Lizhi Pan, Mitsuhiro Hayashibe, Dingguo Zhang, Xiangyang Zhu (Shanghai Jiao Tong Univ.).
EMG signal is widely used to control a limb protheses such as exoskeleton. It would be also useful to control prosthetic hand for upper-extremity amputees. Especially for upper limb control, hybrid control both for position and torque is required for dynamic motions. In this work, we aimed at establishing the EMG-joint angle model for dynamic motions, to effectively decode EMG signals to reproduce the corresponding motion in different velocity. In the experiment, the subject performed the wrist flexion-extension with different speeds along with EMG measurements on flexor and extensor muscles. ARX model, of which the parameters are adaptively identified by extended kalman filter, was applied to represent 1 DoF (degree of freedom) EMG-joint angle model. The result shows that the EMG-angle model could produce good angle tracking as shown in Fig. 2. We have observed that three principal parameters $a_1$, $a_2$, $a_3$ for autoregressive term are almost constant at the same angular speed and different angular speeds correspond to different model parameters. With this regular pattern, SR (Switch Regime) model would be possible to be used to switch the model to reproduce the motion in different conditions. In the further work, we work on creating a SR EMG-angle model using learning and classification technique to dynamicaly decode the EMG signals into the corresponding motions only based on EMG without angle information.

![Figure 2. Dynamic EMG-motion modeling and identification. Subject performed the wrist flexion-extension with different speeds along with EMG measurements on flexor and extensor muscles. Adaptive identification was implemented with Kalman filtering. Model parameters and speed will be used for classification.](image)

6.1.3. Real-time Volumetric Skeletal Muscle Deformation

Participants: Yacine Berranen, Mitsuhiro Hayashibe, Benjamin Gilles.

In this work, we explore skeletal muscle volumetric deformation. The current available simulation for musculoskeletal model is basically using wire-type muscle model which considers only principal longitudinal path of the muscle-tendon units. If we aim at the simulation of the interaction between muscles and objects like orthosis, exoskeleton keeping classical biomechanical property, wire-type modeling is not sufficient. In addition, muscle modeling in volumetric way gives another advantage to reflect microscopic muscle fiber direction and function. We have tried to implemente real-time volumetric skeletal muscle deformation as in Fig. 3 using the INRIA SOFA environment. The idea is making more realistic musculoskeletal simulation from the current approximation of muscle model as wire element to physically and functionally detailed simulation as volumetric element.

6.1.4. Muscle Strength and Mass Distribution Identification in Musculoskeletal Modeling

Participants: Mitsuhiro Hayashibe, Gentiane Venture (Tokyo Univ. of Agriculture and Technology), Ko Ayusawa (Univ. of Tokyo), Yoshihiko Nakamura (Univ. of Tokyo).
In current biomechanics approach, the assumptions are commonly used in body-segment parameters and muscle strength parameters due to the difficulty in accessing those subject-specific values. Especially in the rehabilitation and sports science where each subject can easily have quite different anthropometry and muscle condition due to disease, age or training history, it would be important to identify those parameters to take benefits correctly from the recent advances in computational musculoskeletal modeling. In this paper, Mass Distribution Identification to improve the joint torque estimation and Muscle Strength Identification to improve the muscle force estimation were performed combined with previously proposed methods in muscle tension optimization. This first result highlights that the reliable muscle force estimation could be extracted after these identifications. Fig. 4 shows the estimated muscle forces of Rectus Femoris, Vastus Lateralis and Vastus Medialis with different speeds (first two series are normal speed, second two are slow and last two are fast). The corresponding visualizations of estimated muscle tensions at the indicated time instant are depicted in the bottom. The proposed framework toward subject-specific musculoskeletal modeling would contribute to a patient-oriented computational rehabilitation [22].

6.1.5. Joint Angle Estimation with Inertial Sensors Calibrated by Kinect

Participants: Mitsuhiro Hayashibe, Antonio Padilha Lanari Bo (Univ. of Brasilia), Philippe Poignet.

In this work, we explore the combined use of inertial sensors and the Kinect for applications on rehabilitation robotics and assistive devices. In view of the deficiencies of each individual system, a new method based on Kalman Filtering was developed in order to perform online calibration of sensor errors automatically whenever measurements from Kinect are available. The method was evaluated on experiments involving healthy subjects performing multiple DOF tasks. Accelerometers and gyrometers are used to estimate joint angle, while the Kinect is used for initializing the inertial system and for enabling 3D visualization of the performed task as in Fig. 5 [32].

6.1.6. Investigation of fibre size stimulation selectivity using earthworm model

Participants: Pawel Maciejasz, Christine Azevedo Coste, David Andreu, David Guiraud.

Fibre type and diameter selective stimulation may allow to restore various motor and sensory functions of human body that have been lost due to disease or injury. For example in people unable to voluntarily empty the bladder, selective stimulation of small fibres within the ventral branch of the sacral nerve roots (S2-S4/5) would induce detrusor contraction and those cause bladder emptying closest to normal physiology. Currently, it is not possible to perform it in such a way, because stimulation of sacral nerve roots activates also bigger fibres innervating the urethral sphincter, which closest the outlet of the bladder.

Already many stimulation techniques have been proposed for fibre type and diameter selective stimulation. They were verified performing computer simulations and in some cases also by in vivo experiments on mammalian models. However, results of computer simulations still need to be confirmed by in vivo experiments, whereas experiments on mammalian models, due to high number of fibres within stimulated nerve, can be very
Figure 4. Estimated muscle forces of Rectus Femoris, Vastus Lateralis and Vastus Medialis using the identified model in squat motion with 4kg load with different speeds (first two series are normal speed, second two are slow and last two are fast) (up). The corresponding visualization of estimated muscle tensions at the indicated time instant (bottom).

Figure 5. Angles estimated using inertial sensors in a sit-to-stand task.
complex to perform and obtained results difficult to interpret. As a result, it is still unclear which stimulation parameters may allow for selective stimulation of only particular group of fibers. Therefore we propose the earthworm (Lumbricus terrestris) as a model for selective stimulation. The earthworm has three giant nerve fibres, with two distinctly different conduction velocities and diameters. Therefore it is very easy to distinguish between fibres that are firing at the moment. As a consequence the selectivity of stimulation may be immediately verified without application of sophisticated signal processing and averaging techniques.

We have investigated influence of various pulse amplitudes and durations on the selectivity of stimulation. Using a simple experimental set-up [29] shown in the fig. 6 A, we were able to achieve selective activation of small (fig. 6 C) and big (fig. 6 D) fibers, as well as concurrent activation of both fibers (fig. 6 B) [28]. For that purpose we have used so called "anodal block" technique.

Based on the results of the above experiments, the recommendations of the optimal parameters for selective stimulation of nerve fibres will be prepared. Afterwards we are going to verified these recommendations in mammalian models (rats).

Figure 6. A. The schematic representation of the stimulation ("-" - cathode, "+" - anode) and recording (R1, R2 and R3) sites during experiments on earthworms; B, C and D - examples of nerve responses recorded for various pulse amplitudes. It may be observed that although all fibres are activated under cathode (R1 recording site, first plot in each line), on the anode side (R2 and R3 recording sites, second and third plot in each line respectively) activation of all (B), only small (C) and only big (D) fibres could be achieved.

6.1.7. Neural network based identification for time-variant dynamics
Participants: Zhan Li, Mitsuhiro Hayashibe, David Guiraud.

Due to high nonlinearity and time variance in muscle dynamics under FES, the identification of muscle model is complex task. The time-variant of muscle response may come from muscle fatigue, but also the electrode attachment condition. Along with such long-term time-variant, short term time-variant may be created by reflex effect. In addition, the characteristics of such variance even may change in time in an unpredicted way.
Reinforcement learning framework may be applied to bring the robustness in adaptive identification. Current work is focused on the usage of discrete-time recurrent neural network for model identification.

6.2. Function control and synthesis

6.2.1. Correction of drop-foot in post-stroke hemiplegic patients

Participants: Christine Azevedo Coste, Roger Pissard-Gibollet (SED INRIA), Fabien Jammes (INRIA RA), Jérôme Froger (Rehab. Centre, Grau du Roi, CHU Nîmes).

Hemiplegia is a condition where one side of the body is paretic or paralyzed; it is usually the consequence of a cerebro-vascular accident. One of the main consequences of hemiplegia is the drop-foot syndrome. Due to lack of controllability of muscles involved in flexing the ankle and toes, the foot drops downward and impedes the normal walking motion. Today, there are commercially available assistive systems that use surface electrodes to stimulate Tibialis Anterior (TA) muscle and prevent drop-foot. The efficiency of drop-foot stimulators depends on the timing of stimulation and functionality of dorsiflexion motion. Classically, available stimulators use footswitches to detect foot on/off events. These discrete events allow only for triggering the stimulation and/or playing with the duration of the stimulation pattern, but does not allow for precise online modification of the pattern itself. We have developed algorithms to monitor the ongoing walking cycle by observing the valid limb movements. In order to ensure legs coordination during walking, we propose a robust phase estimation method based on the observer of a nonlinear oscillator. We have modified a commercial stimulator, ODSTOCK, in order to be able to trigger it using our own wireless sensors and algorithms. Agreement from Nîmes CPP (ethical committee) was obtained in June 2010 to run tests on patients. The protocol comprises 1) the control algorithm triggering the stimulator based on signals issued from one wireless inertial sensor placed on healthy shank, 2) a sensor setup including inertial sensors placed on deficient shank and foot, one goniometer measuring deficient ankle angle, one EMG sensor placed on stimulated TA and one instrumented carpet (GAITRITE) (fig. 7). Several patients have been included in the study and data is being processed [15][14].

![Figure 7. Stroke patient study protocol description.](image)

6.2.2. eEMG Feedback Torque Control in FES

Participants: Mitsuhiro Hayashibe, Qin Zhang, Christine Azevedo Coste.
Electrical stimulation (ES) is one of the solutions for drop foot correction. Conventional ES systems deliver predefined stimulation pattern to the affected muscles. However, time-variant muscle response may influence the gait performance as they are difficult to be taken into account in advance. Therefore, closed-loop ES control is important to obtain desired gait in presence of muscle response variation. In this work, a dual predictive control, which consists of two nonlinear generalized predictive controllers, is proposed to track desired torque. The stimulated muscle dynamics are modeled by Hammerstein cascades, with one representing stimulation to activation, the other representing activation to torque. Ankle dorsiflexion torque and ES-evoked EMG of tibialis anterior were recorded experimentally for model identification. The control scheme is validated by following desired torque trajectories with the identified model. The results show that the stimulation pattern obtained from the dual predictive control can produce good torque tracking according to the current muscle condition as shown in Fig. 8. The updates of model parameters were switched off after certain instant for both the excitation and contraction model. Consequently, the model prediction in the control was only driven by the model input and the last model parameter estimates. The dual predictive controller can still generate suitable control signal to obtain desired torque trajectory [23].

![Figure 8. Torque reference (solid blue) and reproduced torque output by the proposed controller (dashed red). The updates of model parameters were switched off after 19.2s. Consequently, the model prediction in the control was only driven by the model input and the model parameter estimates at the time of 19.2s.](image)

### 6.2.3. FES assisted Sit-to-Stand

**Participants:** Jovana Jovic, Christine Azevedo Coste, Philippe Fraisse, Charles Fattal.

Standing up is a common daily activity and a prerequisite to standing or walking. This frequently executed task is one of the most biomechanically demanding activities. The ability to rise from a sitting to a standing position is very important for individuals with paraplegia in order to achieve minimal mobility and has functional and therapeutic benefits related to bone loading, joint extension, cardio-circulatory stimulation, and pressure sore prevention. One method which has been widely investigated is functional electrical stimulation (FES) of the lower extremities. The sit-to-stand method, which is widely used in clinical practice, involves open-loop stimulation of knee extensors activated by hand switches. This technique works adequately in many cases, however, in applying this strategy, stimulation starts without reference respect to the upper body movement. Hence, the whole-body motion is not optimal and requires a high velocity of the joints and large upper limb forces during the rising motion, which may cause both damage of joint tissues and shoulder complications.
We propose a "patient-driven" FES method that would coordinate motion of the trunk, which is under voluntary control of the patient, and motion of the lower limbs, which are under FES control. The proposed approach is based on the observation of trunk movement during rising motion and a detection algorithm, which triggers a pre-programmed stimulation pattern. Trunk acceleration was acquired by a single one-axis wireless accelerometer positioned on the subject’s back. The detection algorithm consists of an online comparison of the movement acceleration of the ongoing motion with the reference pattern (a typical pattern characterizing the sit-to-stand transfer for each subject) using Pearson’s correlation coefficients [25]. Experiments on paraplegic subjects are ongoing in rehabilitation center PROPARA. The experimental setup and a paraplegic subject of the experiment are presented in Fig. 9.

![Figure 9. Experimental setup.](image)

We have shown that in the cases where the acceleration and reference signal are similar, our algorithm is able to recognize sit-to-stand motion and to properly trigger leg stimulation at the desired instant. Also, we have shown that there is an influence of stimulation timing on applied hand forces during the motion. The best results were achieved for trials in which motion was similar to the one of the able bodied subjects in terms of trunk motion and the beginning of the leg motion with respect to the trunk acceleration signal.

We also investigated dynamic optimization as a tool to improve FES assisted sit to stand transfers of paraplegic subjects. The objective would be to find optimal strategy for voluntary trunk movement, which would minimize hip, knee and ankle torques and demand minimal upper limb participation during the motion. Our results suggest paraplegic patients should bend their body forward in order to use linear momentum of the trunk in sit off phase. Figure 10 shows optimal coordination of ankle, knee, and hip angles during sit-to-stand motion [26], [24].

### 6.2.4. Signal-based segmentation of human locomotion using embedded sensor network

**Participants:** Maud Pasquier, Christine Azevedo Coste, Bernard Espiau, Christian Geny (CHU Montpellier), Fabien Jammes.

Last year, we introduced a simple approach to segment in homogeneous phases a long-duration record of locomotion data consisting of body segment acceleration and foot pressure information. We used a system based on a network of wireless nodes embedding various types of sensors [3]. Two cases were considered: walk and run around an indoor running track [35] and outdoor marathon [34].
Figure 10. Optimization results for lower limb trajectories. Blue line is ankle angle, red line is knee angle and green line is hip angle. The dashed bar marks the beginning of sit off phase.

We now use this system as part of a study of mobility impairment caused by Parkinson’s disease (fig. 11). Freezing of gait (FOG) has been identified as one of the main contributors to gait disturbances in this disease. We introduce an ambulatory gait analysis method using body attached gyroscopes and accelerometers to detect the freezing of gait. One hand, we aim at proposing a FOG detection algorithm more robust because the existing algorithms were not able to detect the FOG without tremor. On the other hand, we would like to anticipate the freezing before it is installed in order to reduce the risk of falling.

Before and during a FOG, a patient tends to walk slowly with short strides and fast rhythm. The detection of an increase of frequency is not enough, because there exists similar variations during the initiation of gait or a voluntary acceleration. The association of an increase of the gait rhythm together with a decrease of stride length allows us to detect a FOG. The computation of correlation coefficient in a moving window allows us to estimate the rhythm of strides. We are also working on different methods to estimate the stride length using one or two IMU (3-axis gyroscope and 3-axis accelerometer).

A time-frequency representation permits to show an increase of fundamental frequency and a duplication before a FOG, Fig. 12. The variations of the fundamental frequency are already detected with the correlation. In the future works, we aim at characterizing these duplications and to propose an algorithm of automatic detection.

6.2.5. Awake surgery: How to optimize functional brain mapping by improving per-operative testing?

Participants: Cheikh Niang, Pom Charras, Stephane Argon, Christine Azevedo, Hugues Duffau, David Guiraud, François Bonnetblanc.

It is now possible to perform resections of slowgrowing tumors in awake patients. Using direct electrical stimulation (DES), real-time functional mapping of the brain can be used to prevent the resection of essential areas near the tumor. For now, simple clinical tests are performed on conscious patients and combined with DES in order to discriminate functional and non-functional areas invaded by the tumors. In this work we try to develop a simple device based on a simple technology to better quantify the performances of the patients during the surgery itself and give a real-time feedback to the neurosurgeon that will help to further guide the surgery...
Figure 11. nodes disposition on the patient.

Figure 12. FOG: Smoothed pseudo Wigner-Ville time-frequency distribution with causal kernel.
by improving the sensibility of the functional mapping. This procedure should also allow building a strong
database that should serve retrospectively to improve the surgical procedure and reinforce the neurosurgeons’
experience as well as to monitor the patients’ performances all along their life.

6.2.6. Closed-loop CoM based posture control in FES

Participants: Alejandro González, Mitsuhiro Hayashibe, Philippe Fraisse.

Center of Mass control has been used in humanoid robotics to create stable standing postures and movements.
By controlling the CoM position and acceleration, joint trajectories which respect to the ZMP stability criterion
can be generated. FES may be used to drive joint torques in order to maintain a standing posture within a closed
loop controller. Current work is focused on locating a human’s CoM by creating a statically equivalent serial
chain (SESC) model using widely accessible equipment, such as the Kinect camera and the Wii balance board.

6.3. Neuroprostheses

6.3.1. Stimulator calibration

Participants: Jérémie Salles, Fabien Soulier, Serge Bernard, Guy Cathébras.

![Figure 13. Stimulator architecture overview.](image)

In the context of the TIME project, one CAFE12-based stimulator will be used for chronic experiment in human.
During the validation of the stimulator, it appeared that we needed to improve both the linearity and the current
matching of the 12 channels. We thus define a calibration process consisting in:

- PCB modification: To take advantage of the 10 bits of the DAC (only 8 were used before for compatibiliry reasons). The modification give the FPGA access to the two latter bits. Moreover, test points were added between the ASIC and the output capacitors that now can be removed for the calibration phase.
- Digital interface modification: To allow a 16th current amplifying ratio (only 0-15 were enable). Improvements in the activation sequences of output current mirrors have also been carried on.
- Reference voltage and current tuning: The stimulator use several level of power supply and voltage references. Nominal values are:

\[
V_{hv} = 16.0 \text{ V}, \quad V_{DD} = 3.3 \text{ V}, \quad V_{hv2} = 13.3 \text{ V},
\]

\[
V_{ref P} = 13.5 \text{ V}, \quad V_{ref N} = 1.5 \text{ V}.
\]
The DAC current reference is set to get a 5.46 mA maximum current at the stimulator output for ratio of 15. We have enabled modification of the biasing resistor in order to fine-tune this current reference.

- Raw data acquisition: Measurement of the 12 output currents are carried on independently with the following configuration:
  - all outputs configured as cathodes,
  - all ratios set to 16.

The measurement setup makes use of a characteristic analyzer (HP4156A) to maintain the voltage load to 7.5 V for all the DAC current values.

- Correction: The linearity and matching correction is specific to the association of a particular ASIC with a particular DAC. For human experiment, the configuration will be limited to common anode/controlled cathodes. A first-order linear regression is applied to the 12 raw current measurements. This gives gain and offset adjustments for each channel that are applied by a linear digital correction block (fig. 14). The DAC initial value (8-bit) is multiplied by the correction gain and summed to the correction offset (both channel-dependant), resulting in a 10-bit corrected command. At last, respective correction values are chosen and quantified to lower impact of these modifications to the precision and dynamic range of the stimulator output (no “lost” bit). Concerning the implementation, since the 3 most significant bits of the correction gain appear to be constant, it is possible to use channel-independent bit-shifts and a substractor to perform the 8 to 10-bit multiplication.

![Figure 14. Linear digital correction principle.](image)

The improvement in the matching of the 12 channels can be seen in figure 15. The results in terms of linearity are an integral non-linearity of ±2.5 LSB and a differential non-linearity of ±0.3 LSB.

### 6.3.2. Nerve Modelling for ENG recording

**Participants:** Olivier Rossel, Jonathan Coulombe, Fabien Soulier, Serge Bernard, Guy Cathébras.

In the context of FES, neural recording is one of the main issues, as the control requires information carried on afferent peripheral nerves. Because specific information are carried in different fascicles, we propose to realize a non-invasive and spatial-selective electrode. Last year, based on investigation on the topic of extracellular Action Potentials (AP), we proposed a new tripole design, where the tripolar output signal is the image of the activity in the close vicinity of this tripole, providing high spatial selectivity.

We showed however, that this high spatial selectivity is achieved at the expense of signal amplitude. This first result jeopardizes the feasibility of this kind of electrode since the signal amplitude appears to be on the same range of the expected noise. First, we propose to estimate the performance of the proposed electrode with a quantitative study of the electrode selectivity. Then, to conclude on the feasibility of this electrode, the SNR has to be determined. So with a more accurate model, we studied the sensitivity of the proposed tripole, allowing to determine precisely the amplitude level of the expected signal. Thus, the SNR can be estimated knowing the expected noise.
Figure 15. Mean gain (top) and worst case (bottom) before and after correction.
In short, the work of this year aims at characterizing the performances and evaluating the feasibility this new multi-contact cuff electrode.

6.3.2.1. Selectivity

We proposed an electrode configuration inspired from the FINE electrode (figure 16) designed for the same purpose. The electrode is composed of many tripoles, placed around the nerve. This disposition is used for two electrode, state-of-the-art electrode A and the proposed electrode B. The unique difference between both electrode resides on the longitudinal inter-pole distance ($d_e$), which is respectively 5 mm for the electrode A and 0.375 mm for the electrode B.

Figure 16. Electrodes A and B have the same shape but differ on the longitudinal inter-pole distance ($d_e$). Two fascicles are represented.

The electrodes performances are evaluated based on simulations using a model of a nerve comprising multiple fascicles [38]. The Selectivity Index (SI) quantifies the ability to record and distinguish between different active fascicles in such a manner that $SI = 0$ corresponds to a case where an active fascicle yields identical signals at every recording site, while $SI = 1$ occurs when one recorded signal is different from every other. This SI has to be presented according to the inter-fiber spacing.

The result of electrode selectivity are presented in the fig. 17. This figure shows that activity of two fascicles separated by as little as 1 mm can be distinguished for the proposed electrode (for this distance $SI$ for electrode B ($\approx 0.9$) is more than double that of electrode A ($< 0.4$)). The proposed electrode thus appears to be much more selective than the reference electrode.

6.3.2.2. Sensitivity

Using a more realistic model (inhomogeneous and anisotropic), we investigate the spatial properties of extracellular AP and that of the filtering done by the proposed tripole [37]. This allows us to represent the tripolar sensitivity. It was realized for the proposed tripole B and compared to a state-of-the-art tripole A 18. This sensitivity represents the amplitude of the tripolar output signal for a single unit action potential.

This figure shows that the classical tripole radial sensitivity is huge compared to that of the proposed electrode. This confirms the high spatial sensibility of the proposed tripole. We can also determine the expected amplitude, where the signal can reach 6 $\mu$V. Considering this amplitude and knowing that in natural case there will be superposition of action potentials, we can conclude that the signal amplitude could be higher than the expected noise (around 1 $\mu$V). So we can conclude positively to the feasibility of this kind of electrode.

6.3.3. Low-noise, low-power ENG amplifier design

Participants: Jonathan Coulombe, Olivier Rossel, Fabien Soulier, Serge Bernard, Guy Cathébras.

This year we proposed a method for enhancing the noise-power tradeoff of front-end amplifiers in parallel recording applications of analog signals with respect to a common reference. One example of application is shown in the Fig. 19 for spatial-selectivity ENG recordings.

The circuit architecture is based on a Shared-Input Amplifier (SIA), composed by shared-input transconduc-tance amplifiers and a differential stage. Averaging null signals and subtracting the result from every signal reduce the noise because the correlated noise between parallel outputs is attenuated. It results significant supply current savings without noise penalty. One example is shown in the Fig. 20 for the specific case of two average of two null signals.
Figure 17. Selectivity index computed for random combinations of simulated fascicles, plotted as a function of the distance between each couple of fascicles.

Figure 18. Comparison between classical and proposed tripole sensitivity. The peak-to-peak amplitude of a single unit action potential is represented, measured by a classical tripole on the left and by the proposed tripole on the right. The tripoles are placed on the surface of a nerve of 300 µm radius.
Figure 19. Conceptual representation of a selective ENG recording system.

Figure 20. SIA modified for the use of transconductance amplifiers in open-loop configuration for ENG recordings. And concept of parallel SIA with noise reduction using null output averaging and differential readout.
Also, a method for reducing the remaining noise with little power penalty is possible. And it is possible to combine both methods, either noise level, total supply current, or both can be significantly reduced. The benefits of combining both methods and the related trade-offs was validated by simulations using models of a 0.35 $\mu$m BiCMOS process. So we have shown that the total supply current can be reduced by more than 50% that of a comparable system using conventional differential amplifiers with equivalent output noise. Alternatively, the noise can be reduced by approximately 35% with comparable power consumption.

This should enable low-noise recording of signals with significantly better efficiency than even the theoretical limit of any conventional differential amplifier. Future work will include circuit optimization and investigation of the impact of the architecture over other performances of the system, such as crosstalk, linearity, distortion, and channel mismatch. Implementation of the circuit for full characterization is expected to be completed in the near future.
6. New Results

6.1. Theoretical results

6.1.1. Theory of competition for one limited resource

Participants: Claude Lobry, Tewfik Sari, Radhouene Fekih-Salem.

In the paper [24], we give a global asymptotic stability result for a mathematical model of competition between several species in a chemostat, by using a new Lyapunov function. The model includes both monotone and non-monotone response functions, distinct removal rates for the species and variable yields, depending on the concentration of substrate.

In the paper [14], we consider the mathematical model of two species microbial competition on a single food resource in a chemostat, when one takes into account species interactions between the two populations of microorganisms and intraspecific interactions between individuals themselves, using strictly monotonic growth functions and distinct dilution rates.

6.1.2. Study of input/output maps of interconnected chemostats

Participants: Alain Rapaport, Ihab Haidar.

Patch or island models are popular in ecology, and are a convenient way to study the influence of a spatial structure of a geography on the distribution of the abundance of resources. Coupling such a structure with abiotic/biotic models and studying its input-output properties has been very rarely tackled in the literature. In biotechnology engineering, dead-zones models, that distinguish two sub-domains (a “living” and a “dead” one) are often used for approximating non perfectly mixed tanks. No more sophisticated representation, apart continuous space models (systems of partial differential equations), have been investigated.

We consider an hydric capacity and an nutrient flow that are fixed, and analyze the influence of different structures, having the same total hydric volume, on the output concentrations at steady-state. Three configurations are compared, under the assumption of a monotonic growth rate: perfectly-mixed, serial and parallel with diffusion rate. In each case, we show the uniqueness of a steady-state different to the washout equilibrium and its global asymptotic stability in the positive orthant. We prove the existence of a threshold on the input concentration of nutrient for which the benefits of the serial and parallel configurations over the perfectly-mixed one are reversed. In addition, we show that the dependency of the output concentrations on the diffusion rate can be non-monotonic, and give precise conditions for the diffusion effect to be advantageous [19], [33]. The study encompasses the dead-zone models.

The possibly non-monotonic influence of the diffusion parameter on the output steady state is not intuitive, and leaves further investigations open for understanding or taking benefit of this property for natural ecosystems (such as saturated soils or wetlands) as well as for bioprocesses (such as wastewater treatments). This result can be also of interest for reverse engineering when deciding which among serial or parallel configurations is a better fit for the modelling of chemostat-like ecosystems, providing that one has an estimation of the hydric capacity of the system.

This work is part of the material thesis of I. Haidar [11].

6.1.3. Aggregation models in the chemostat

Participants: Claude Lobry, Alain Rapaport, Jérôme Harmand, Tewfik Sari, Radhouene Fekih-Salem.
Bacteria aggregation often occurs in bioprocesses, creating flocks or biofilms (the latter being attached to the tank walls). At a macroscopic level with large populations of aggregated and non-aggregated individuals, a simple way of modelling this phenomenon in the chemostat is to distinguish explicitly two populations: planktonic or free bacteria and attached ones. The main differences between flocks and biofilms rely in the attachment/detachment terms and the effective dilution rate (assumed to be zero or very small for biofilms). Typically, the specific growth rate of free bacteria is expected to be larger than the attached one (that have in average a restricted access to nutrient and use part of their energy to glue together).

Based on former works of the team and the main assumption that attachment and detachment dynamics are much faster than the biological one, we have shown that a significant difference between the specific dilution rates of the free and attached populations can surprisingly lead to bi-stability, even for a single species and monotonic growth rates [30], [42], [43].

A work in progress addresses the case of two species, one of them having a non-monotonic growth rate (due to substrate inhibition) and the ability to form flocks. Without flock, the Competitive Exclusion Principle extended by G. Wolkowicz and her co-authors [47] shows the possibility to have one of the two species winning the competition depending on their initial repartition. Here, the presence of flocks may lead on the contrary to a single winner.

6.1.4. Neutral community models for microbial ecology

Participant: Bart Haegeman.

Hubbell’s neutral model [50] describes the dynamics of an ecological community in terms of random birth, death, immigration and speciation events, attributing equivalent characteristics to all species. Despite the absurd simplicity of these assumptions, remarkable agreement between neutral model predictions (e.g., the distribution of the abundance of the species present in the community) and empirical observations has been reported for some, mostly rather diverse, ecological communities.

There is some evidence that also certain aspects of microbial communities can be well described by the neutral model. Highly diverse microbial communities have been difficult to deal with using more traditional modelling approaches from community ecology. The neutrality assumption could lead to an effective global description, without requiring quantitative species data (growth characteristics, interaction strengths, etc). We are actively participating in the development of neutral community models, with a focus on microbial systems.

(1) Effect of speciation process

It has been argued that the neutral model predictions are rather insensitive to its assumptions. However, we have found that the details of the way new species appear in the community (i.e., the speciation process) do matter, and can drastically change the model predictions. In particular, we have studied the neutral community model with random fission speciation. This speciation model is quite different from the point mutation model usually considered in neutral community model, and is generally believed to be more realistic.

Using a technique from theoretical physics, we have obtained the stationary distribution of species abundances for the random fission model. We have compared our solution with the well-known stationary distribution of species abundances for the point mutation model on empirical data (tree communities in tropical forests) [13]. Surprisingly, we found that the point mutation model fits the data better than the random fission model, although the latter is believed to be more realistic.

(2) Comparison with niche models

Neutral community models challenge more traditional, niche-based models in community ecology. Niche theory states that species can coexist only if they differ sufficiently in their characteristics (for example, their use of available substrates). Neutral theory assumes that all species have approximately equal characteristics. Hence, the two theories describe species coexistence in fundamentally different ways.

We have tried to narrow the gap between the two theories. We have proposed a mathematical model that combines essential features of niche-based and neutral community models [17]. It integrates species niches, described as Lotka-Volterra interactions, in the standard neutral community model. The analysis of this model
indicates that the addition of species interactions has a limited effect on the species abundance distribution. We have further clarified this result using a slightly different model that also combines niche and neutral features [16]. For the latter model we have proved that the niche structure does not affect at all the species abundance distribution.

6.1.5. Quantifying genetic diversity of bacteria

**Participant:** Bart Haegeman.

With the wide availability of DNA sequencing, microbiologists are now able to rapidly sequence entire bacterial genomes. Comparison of these genomes has revealed a large genetic diversity within bacterial species. For example, one genome of the bacteria *E. coli* has about 4000 different genes, but a set of 10 genomes of *E. coli* has typically over 10000 different genes. Some of these genes are shared by all or almost all of the genomes, but many other genes are only present in one or a few of the genomes. This observation has important implications for the definition of bacterial species and for the description of the functional characteristics of bacteria.

We have been dealing with the problem of how to quantify this observed genetic diversity. Microbiologists have introduced notions like the pan genome of a bacterial species (that is, the set of genes that are present in at least one of the organisms from the species) and the core genome of a bacterial species (that is, the set of genes that are present in every organisms from the species). However, we have argued that both the pan and core genome are difficult to estimate, and should not be used for quantitative purposes [21]. Instead, we have proposed a measure of genetic diversity that has much better estimation properties. It is based on the average number of genes shared by a pair of genomes when sequencing two randomly sampled organisms from the species under consideration. We have applied our estimator on six bacterial species (about 100 sequenced genomes in total). Software for our robust estimation procedure of genetic diversity is freely available, see http://ecotheory.biology.gatech.edu/downloads/genomic-fluidity-scripts.

6.1.6. Individual-based modelling

**Participants:** Fabien Campillo, Chloé Deygout, Coralie Fritsch, Marc Joannides, Claude Lobry.

In terms of computational modelling of ecosystems, individual-based models (IBMs) are an interesting path to explore. We can outline two types of IBMs. On the one hand “detailed IBM” attempt to integrate in an ad-hoc way all the knowledge available about an ecosystem. On the other hand, “simplified IBM” are limited to one or several mechanisms to simplify the analysis. The former may be more realistic but are often difficult to analyze. Although the latter are too simplistic in realistic situations they lend themselves to the analysis and numerical analysis. We focus on the latter.

The IBMs offer an interdisciplinary language between biologists, biotechnologists, mathematicians, and computer scientists, to develop models in the form of relatively simple rules. In the case of simplified IBMs it is possible to translate these rules in the form of a branching Markov process with values in a space of measures. Using scaling methods, the IBMs can be approximated by integro-differential equations; using model simplification methods IBMs can be reduced to stochastic or ordinary differential equations. The mathematical interpretation of the IBMs and their analysis is relatively recent and still very few studies exist [48]. The numerical analysis of these models is yet to be built. Under certain conditions, IBMs themselves can be simulated through adapted Monte Carlo procedures.

The MODEMIC project-team develops three studies in the field of IBMs. The first is part of the ANR MODEC COL on the modelling of clonal plant growth (see Section 7.5); the second is part of the ANR DISCO on modelling of biofilms (see Section 7.4); the last one is a starting thesis.

In all cases, we aim at developing the Monte Carlo simulation of the IBM as well as analyzing their links with integro-differential models. We also seek to make connections with non-IBM models proposed in Section 6.1.8.
In October 2011, Coralie Fritsch started a thesis at the École Doctorale I2E of the University of Montpellier 2, under the supervision of Fabien Campillo, Jérôme Harmand and Marc Joannides. This thesis is supported by a grant of the MESR and a grant of INRA from the MEM Meta-program (Méta-omiques des écosystèmes microbiens). The thesis aims at developing and analyzing individual-based microbial ecosystems models that capture both the spatial, biodiversity and function of these ecosystems. The thesis received the Agreenium label in December.

6.1.7. Hybrid modelling of biofilms in plug-flow reactors

Participants: Fabien Campillo, Chloé Deygout, Annick Lesne, Alain Rapaport.

Within the DISCO project of the SYSCOMM program founded by the ANR, we have proposed a multi-scaled modelling that combines three scales: a microscopic one for the individual bacteria, a mesoscopic or “coarse-grained” one that homogenises at an intermediate scale the quantities relevant to the attachment/detachment process, and a macroscopic one in terms of substrate concentration (see the Section 7.4).

Such an “hybrid” approach allows for modelling and understanding in plug-flow reactors [41] the interplay between

- the formation of the biofilm at a microscopic scale, that starts from a small number of bacteria (thus a stochastic individual based description),
- the limitation of the biofilm, due the carrying capacity of the wall attachment, at a mesoscopic scale,
- the consumption of nutrient along the flow at a macroscopic level, as a solution of a coupled transport-reaction partial differential equation.

The numerical computation of such a model requires a software architecture that allows the simultaneous simulation of stochastic events at the bacteria scale and the continuous evolution (in space an time) of the substrate density.

Experiments on real tubular plug-flow reactors are currently driven at Cemagref HBAN with the perspective of comparison with numerical simulations. After spending one year at Montpellier for deriving and simulating the theoretical model, our post-doctoral fellow C. Deygout is presently participating to the real experiments at Cemagref Antony.

The multi-species case with different bacteria specialized in different environments (poor or rich in nutrient) is a work in progress.

6.1.8. Stochastic modelling for biotechnology

Participants: Fabien Campillo, Marc Joannides, Claude Lobry.

This year we continue to study stochastic models for the chemostat [12], [39], [27]. Starting from the well-known ordinary differential equation systems, we propose first a pure jump process model at the microscopic scale that leads to a stochastic differential equation at the intermediate scale and to an ordinary differential equation at the macroscopic level (fluid limit model). After developing the model, we establish the Fokker-Planck partial differential equation for the diffusion model. This PDE integrates a specific washing-out term. We proposed an ad hoc numerical integration scheme for the simulation of this PDE [39].

In [40], we consider a stochastic version of the basic predator-prey differential equation model. The model, which contains a parameter ω which represents the number of individuals for one unit of prey – if x denotes the quantity of prey in the differential equation model x = 1 means that there are ω individuals in the discontinuous one – is derived from the classical birth and death process. It is shown by the mean of simulations and explained by a mathematical analysis based on results in singular perturbation theory (the so called theory of Canards) that qualitative properties of the model like persistence or extinction are dramatically sensitive to ω. For instance, in our example, if ω = 10^7 we have extinction and if ω = 10^8 we have persistence. This means that we must be very cautious when we use continuous variables in place of jump processes in dynamic population modelling even when we use stochastic differential equations in place of deterministic ones.
6.1.9. Minimal time control of batch bioprocesses

Participants: Denis Dochain, Alain Rapaport.

Minimal time control problems often occur in biotechnology when one has to fill tanks. Typically, the objective to be reached is to have the tank full with a prescribed value of substrate or product concentrations, the tank being filled with a high concentration of nutrient.

When a single reaction occurs, the optimal solution is already known and has been rigorously proved by J. Moreno in 1999 [51] using the Green’s theorem: it consists in a “bang-bang” strategy (fill as fast as possible or do not fill) and possibly a singular arc when the growth function presents an inhibition (i.e. a maximum growth for a precise concentration of nutrient). When impulse controls in addition to regular control are allowed, an extension of this result has been recently proposed with a different technique that do not use the Green’s theorem [4]. This technique has also allowed to solve partially the problems when several species compete for the degradation of the substrate, but when all of them have a monotonic growth.

In the presence of complex non monotonic kinetics, typically characterized by the combination of two non-monotonic growth functions, aimed at emphasizing the presence of two parallel metabolic pathways to transform the limiting substrate into the biomass, the candidate singular arcs are multiple and determining which singular arc is eventually optimal is clearly a crucial issue. The local optimality conditions based on the Pontryagin Maximum Principle allow to characterize the geometric structure of the extremal trajectories, in which there may be singular arcs, but these necessary conditions are not always sufficient for determining which extremals are (globally) optimal. Then one has to compute the cost of each extremal or use global optimization methods such as dynamic programming or Hamilton-Jacobi- Bellman equation. The extremals are traditionally determined numerically by considering shooting methods, but for bang-bang control, it is well known that one may face numerical troubles because the shooting function is in general not smooth.

For this problem, we have used an approximation technique first proposed and studied by C. Lobry and his students [53] and later by C. Silva and E. Trélat [52], that consist in adding an artificial control. In [22], we have proposed a new proof of convergence based on differential inclusions arguments that allows to relax the assumption of the uniqueness of the optimal solution for the convergence of the optimal paths of [52]. Then we have shown how to apply numerically this approximation procedure for analyzing the field of extremals on the whole state space. This technique appears to be quite effective for the practical determination of optimal synthesis in the planar case even in presence of multiple singular arcs.

6.1.10. Optimal control of continuous bioprocesses

Participants: Jérôme Harmand, Alain Rapaport, José Fernandez, Walid Bouhafs, Amel Ghouali.

In continuous bioprocesses, a usual objective is to stabilize the output of the bioreactors about a desired steady state (in wastewater industry, this value is typically chosen under the norm of authorized discharge). It happens more and more frequently that transient trajectories are expected also to maximize a product of interest.

We have begun to study the maximization of the gaseous production of methane in anaerobic processes over a given period of time on specific problems. For the moment we have proved that the optimal trajectory consists in approaching a unique singular arc as fast as possible when only one limiting substrate has to be converted, but the problem is still open when involving several substrates. These works are part of the PhD work of A. Ghouali and W. Bouhafs.

Reference points in batch processes can be mimicked by a series of continuously stirred bioreactors in series at steady state (see applications 6.2.2 and 7.2). We study the minimal time problem to drive the nutrients concentrations of a cascade of chemostats. The control variable is the dilution rates of each tank, under the constraint that each dilution rate is bounded by the one of the previous tank, that makes the system not locally controllable. For the particular case of two tanks with total mass at steady state, the planar feedback synthesis has been found but the problem is still under investigation for the general case.

6.1.11. Minimal time bioremediation of natural resources

Participants: Jérôme Harmand, Alain Rapaport, Antoine Rousseau.
In biological wastewater treatment (batch or continuous bioprocesses), one has always to separate biomass from the purified liquid phase at the output of the tanks, that is not possible when tanks are rather natural reservoirs such as lakes or water tables.

We have proposed a new operation strategy that consists in treating with the help of a bioreactor aside. No bacteria are introduced in the reservoir but water is pumped and treated by microorganisms in a smaller tank, and treated water returns to the reservoir after being separated from the biomass. Consequently, there is no need of a separation operation for the reservoir.

The minimal time control problem consists in controlling the flow rate for having the substrate concentration of the whole reservoir below a given reference value as fast as possible.

Last year, we have determined analytical expressions of optimal feedback strategies for a general class of growth functions under the assumptions that the volume of the bioreactor is much smaller than the reservoir one, and that the spatial repartition of the concentration of the pollutant in the reservoir can be modelled by simple spatial representations: either perfectly mixed or discrete one directional gradient [44], [15]. This year, we have studied more realistic spatial motifs:

- dead-zones: we have shown that the optimal synthesis is identical to the perfectly mixed case, even though the time to reach the target is larger [31].
- two parallel zones, allowing to control the repartition of the flow rate between the two zones. Without diffusion between the zones, the optimal solution is almost straightforward and under investigation in presence of lateral diffusion.

This work is mainly achieved in cooperation with Chilean researchers and PhD students within the associated team DYMECOS.

6.2. Applications

6.2.1. Modelling and control of Anaerobic Digestion processes

Participants: Amine Charfi, Radhouene Fekih-Salem, Jérôme Harmand, Boumediene Benyahia, Tewfik Sari.

We consider the AM2 or AMOCO model developed in [46] and extend both the model in itself and its analysis to the following cases:

- Depending on the AM2 model parameters, the steady states were analytically characterized and their stability were analyzed. Following this study, it was shown that the overloading tolerance, a parameter proposed in [49] to on-line monitoring anaerobic processes, may be not adapted under certain operating conditions and even lead to bad operating decisions.

- Within the framework of the PhD theses of Amine Charfi and Boumediene Benyahia, we have included the fouling dynamics of membranes into the AM2 and we have analyzed the resulting model (called the AM2b).

- We actually work towards two directions: (i) we are extending these results in including into the AM2 an additional process, *i. e.* the hydrolysis step in order to study bioprocesses treating solid waste (the resulting model being called the AM3); (ii) we try to find links between complex models such as the ADM1 model and simple models such as the AM2b or the AM3.

6.2.2. Modelling and control of cascade biosystems to mimic batch wine making processes

Participants: Jérôme Harmand, Alain Rapaport, José Fernandez.
An experimental setup of four tanks connected in series has been designed by the research unit SPO (Montpellier) for studying four physiological stages of yeast as steady state. The manipulated variables are the flow rates $Q_i$ of each tank with the constraint $Q_i \geq Q_{i-1} \geq 0$, and the objective is to reach simultaneously four set-points in the four tanks. We are studying two kinds of control strategies:

- a linearizing feedback law that drives exponentially the dynamics to the target. This is not the fastest strategy but is has good robustness properties. Nevertheless, the inputs constraint imposes to use saturation functions that provide satisfactory convergence in simulations but that is hard to prove mathematically.

- a minimal time feedback. Due to lack of local controllability imposed by the constraint on the inputs, the optimal synthesis is not smooth with the presence of “barriers” (see Section 6.1.10).

Those feedback laws will be implemented and tested on real pilot plant at SPO lab in the scope of the European project CAFE described in Section 7.2.

### 6.2.3. Modelling and simulating terrestrial plant ecological dynamics

**Participants:** Fabien Campillo, Ihab Haidar.

This study is part of the ANR Syscomm MODECOL that is done in collaboration particularly with the University of Rennes I, the University of La Rochelle and INRIA. This is the second year of the three years program. We propose a stochastic individual-based model for clonal plant dynamics in continuous time and space, focusing on the effects of the network structure of the plants on the reproductive strategy of ramets. This model is coupled with an explicit advection-diffusion dynamics for resources. We develop a partially exact simulation scheme of the model; the capacity of the model to reproduce specific features of clonal plants, such as their efficiency to forage resources over the field, is numerically studied. Next, we propose a large population approximation of the model for phalanx-type populations, taking the form of an advection-diffusion PDE for population densities, where the influence of the local graph structure of the plant takes the form of a nonlinear dependence in the gradient of resources. This year we improved the simulation code that was proposed last year and made three communications in international conferences [26] [36] [35].

### 6.2.4. Modelling and inferring agricultural dynamics

**Participants:** Fabien Campillo, Angelo Raherinirina.

The International Laboratory LIRMA supports this work that is done in collaboration with the University of Fianarantsoa in Madagascar and with Dominique Hervé (IRD, Fianarantsoa, Madagascar). The aim is to study the dynamics of agricultural plots on the edge of primary forest. In [38] we propose a Markov chain model where the transition matrix is estimated both by maximum likelihood and Bayesian approaches. We also test if the Markov chain model is adapted to this problem. In an ongoing work we develop semi-Markov models for an extended data set.

### 6.2.5. Modelling and simulating microbial ecosystems in soils

**Participants:** Ihab Haidar, Jérôme Harmand, Alain Rapaport.

The team studies simple representations of the spatial inhomogeneity for bioprocesses, in terms of networks of interconnected compartments. Each compartment is modelled as a perfectly mixed bioreactor. Simulation of such networks is performed with the software developed by the VITELBIO project (see 5.1). Comparisons of simple structures (cascade of bioreactors) with numerical simulators based on REV (Representative Elementary Volumes) have been performed. The objective is to understand the role of the topology of the network on the biological functions of the overall system. For the moment simple configurations with two to four nodes and one single species have been investigated.

In addition, we study in simulation how software implementations of transport reactions models such as MIN3P, can mimic such simple configurations. Transport reactions models are often used in the community of soil functioning modelling by simulation. Here we compare their numerical computation with the exact solutions that can be approximated with a good accuracy using classical o.d.e. solvers. We found that one has to be careful when the trajectories are close to a non-hyperbolic equilibrium [45], [18].
6.2.6. Numerical optimisation in non perfectly mixed tanks

Participants: Jérôme Harmand, Alain Rapaport, Antoine Rousseau.

The optimisation of series of bioreactors in terms of minimizing the total residence time have been already investigated in the literature. In such models, the space has a one-dimensional representation. There is comparatively much less work on 2D or 3D space models. We consider different shapes of reactors of the same volume, and simulate with the help of multi-physics numerical software the coupling of the hydrodynamics laws in 3D or 2D (under cylindrical symmetry), solved numerically by the Navier-Stokes equations, with the system of differential equations of biotic/abiotic concentrations [34].

We are studying the influence of the shape of the domain and possible obstacles on the output concentration at steady state.

With A. Rousseau (EPI MOISE, INRIA Rhône-Alpes) we study with simulation of 2D Navier-Stokes equations the benefits of having several pumping points and how to control the pumps speed to minimize the treatment duration for the bioremediation of natural reservoirs presented in Section 6.1.11. We evaluate the feedback strategies that are derived to be optimal for simple models of o.d.e. in the more realistic framework of hydrodynamics simulation.

6.2.7. Individual-based models for the bacterial degradation of the cellulose

Participants: Fabien Campillo, Chloé Deygout, Marc Joannides.

We propose an individual-based model for the degradation of one cellulose bead (dozens of micrometers in diameter) by cellulolytic bacteria. Our aim is to determine the macroscopic degradation behavior. The initial stages of the degradation process may involve a very limited number of bacteria that cannot be properly modelled by classical models based on deterministic equations. In the present work we only consider a two-dimensional model for the degradation of a cellulose disc.

6.2.8. Non-linear filtering for the chemostat

Participants: Boumediene Benyahia, Fabien Campillo, Jérôme Harmand.

We propose numerical non-linear filtering approaches for the identification of non-observed components of dynamical systems in the context of the chemostat. This recently started study relies on the work of the project-team in the stochastic modelling of the chemostat. In a preliminary work we consider the bootstrap particle filter.
5. New Results

5.1. Statistical analysis of spike trains

Modern advances in neurophysiology techniques, such as two-photon imaging of calcium signals or microelectrode arrays (MEA) electro-physiology, have made it possible to observe simultaneously the activity of large assemblies of neurons. Such experimental recordings provide a great opportunity to unravel the underlying interactions of neural assemblies and to understand how neural populations dynamically encode information. The goal of the present project is to propose to the neuroscientists community statistical methods and numerical tools to analysing the statistics of action potentials (spike trains) obtained from MEA recordings. Our work is grounded on one hand on theoretical results on Gibbs distributions in neural networks and the other hand on a C/C++ library of algorithms developed jointly with the CORTEX INRIA team, freely available at http://enas.gforge.inria.fr/. We have collaborations with several labs specialized in MEA recording from the retina: Centro Interdisciplinario de Neurociencia de Valparaiso, Universidad de Valparaiso, Chile http://www.cinv.cl/; Department of Molecular Biology and Princeton Neuroscience Institute, Princeton University, USA http://www.princeton.edu/neuroscience/; Institut de la Vision, Paris http://www.institut-vision.org/.

5.1.1. A discrete time neural network model with spiking neurons. Dynamics with noise.

Participant: Bruno Cessac [correspondent].

We provide rigorous and exact results characterizing the statistics of spike trains in a network of leaky Integrate-and-Fire neurons, where time is discrete and where neurons are submitted to noise, without restriction on the synaptic weights. We show the existence and uniqueness of an invariant measure of Gibbs type and discuss its properties. We also discuss Markovian approximations and relate them to the approaches currently used in computational neuroscience to analyse experimental spike trains statistics. This work has appeared in Journal of Mathematical Biology[ 17 ].

5.1.2. Statistics of spike trains in conductance-based neural networks: Rigorous results

Participant: Bruno Cessac [correspondent].

We consider a conductance-based neural network inspired by the generalized Integrate and Fire model introduced by Rudolph and Destexhe in 1996. We show the existence and uniqueness of a unique Gibbs distribution characterizing spike train statistics. The corresponding Gibbs potential is explicitly computed. These results hold in the presence of a time-dependent stimulus and apply therefore to non-stationary dynamics. This establishes a rigorous ground for the current investigations attempting to characterize real spike trains data with Gibbs distributions, such as the Ising-like distribution, using the maximal entropy principle. This work has appeared in Journal of Mathematical Neuroscience [ 18 ].


Participants: Bruno Cessac [correspondent], Adrian Palacios [Centro de Neurociencia, Valparaiso, Chile].

This work focuses on methods from statistical physics and probability theory allowing the analysis of spike trains in neural networks. Taking as an example the retina we present recent works attempting to understand how retina ganglion cells encode the information transmitted to the visual cortex via the optical nerve, by analyzing their spike train statistics. We compare the maximal entropy models used in the literature of retina spike train analysis to rigorous results establishing the exact form of spike train statistics in conductance-based Integrate-and-Fire neural networks. This work is submitted in “Mathematical Problems in Computational Biology and Biomedicine” Springer, [ 54 ].
5.1.4. Gibbs distribution analysis of temporal correlations structure in retina ganglion cells

Participants: Michael Berry II [Department of Molecular Biolog, Princeton University, USA], Bruno Cessac [correspondent], Olivier Marre, Adrian Palacios [Centro de Neurociencia, Valparaiso, Chile], Juan-Carlos Vasquez.

We present a method to estimate Gibbs distributions with spatio-temporal constraints on spike trains statistics. We apply this method to spike trains recorded from ganglion cells of the salamander retina, in response to natural movies. Our analysis, restricted to a few neurons, performs more accurately than pairwise synchronization models (Ising) or the 1-time step Markov models (Marre et al. (2009)) to describe the statistics of spatio-temporal spike patterns and emphasizes the role of higher order spatio-temporal interactions. This work has been accepted in Journal of Physiology, Paris [28] (in press).

5.1.5. A Markovian event-based framework for stochastic spiking neural networks

Participants: Olivier Faugeras, Jonathan Touboul.

In spiking neural networks, the information is conveyed by the spike times, that depend on the intrinsic dynamics of each neuron, the input they receive and on the connections between neurons. In this article we study the Markovian nature of the sequence of spike times in stochastic neural networks, and in particular the ability to deduce from a spike train the next spike time, and therefore produce a description of the network activity only based on the spike times regardless of the membrane potential process. To study this question in a rigorous manner, we introduce and study an event-based description of networks of noisy integrate-and-fire neurons, i.e. that is based on the computation of the spike times. We show that the firing times of the neurons in the networks constitute a Markov chain, whose transition probability is related to the probability distribution of the interspike interval of the neurons in the network. In the cases where the Markovian model can be developed, the transition probability is explicitly derived in such classical cases of neural networks as the linear integrate-and-fire neuron models with excitatory and inhibitory interactions, for different types of synapses, possibly featuring noisy synaptic integration, transmission delays and absolute and relative refractory period. This covers most of the cases that have been investigated in the event-based description of spiking deterministic neural networks.

This work has appeared in the Journal of Computational Neuroscience [26].

5.2. Coding by spikes

Our goal here is a better understanding of the extent to which computing and modeling with spiking neuron networks might be biologically plausible and computationally efficient. Based on a thorough characterization of the main constraints on spiking neural networks dynamics this has led us to propose new algorithms to infer the structure of the network from its spike trains and to propose an FPGA implementation of spiking neural networks.

5.2.1. Reverse-engineering of spiking neural networks parameters

Participants: Bruno Cessac [correspondent], Horacio Rostro-Gonzalez, Thierry Viéville [Cortex].

We consider the deterministic evolution of a time-discretized spiking network of neurons with connection weights having delays, modeled as a discretized neural network of the generalized integrate and fire (gIF) type. The purpose is to study a class of algorithmic methods allowing to calculate the proper parameters (synaptic weights) to reproduce exactly a given spike train generated by an hidden (unknown) neural network. This problem is linear (L) if the membrane potentials are observed and LP (Linear-Programming) if only spike times are observed, in the context of gIF models. The L or LP adjustment mechanism is local to each unit and has the same structure as an "Hebbian" rule. This paradigm is easily generalizable to the design of input-output spike train transformations. This means that we have a practical method to "program" a spiking network, i.e. find a set of parameters allowing us to exactly reproduce the network output, given an input.

This work has been submitted in the Journal of Neural Engineering, 2011[25].
5.2.2. Development of FPGA-based efficient reconfigurable architectures for spiking neural networks

**Participants:** Bruno Cessac, Bernard Girau [INRIA Cortex], Horacio Rostro-Gonzalez, Cesar Torres-Huitzil [Information Technology Department, Polytechnic University of Victoria (UPV), Tamaulipas, Mexico], Thierry Viéville [Cortex, correspondent].

Spiking neural networks are able to perform very powerful computations with precise timed spikes. We are developing an FPGA (Field Programmable Gate Array) reconfigurable platform that enables the simulation of in silico models of spiking neural networks. Since the model is directly mapped into a FPGA device, the neural processing is accelerated and the time consumption reduced. We use VHDL and Handel-C to design the reconfigurable architecture of a discrete time Integrate-and-Fire model coded in CUDA, running on GPU. This work has been accepted in Journal of Physiology, Paris [24].

5.2.3. Towards biologically inspired image coders

**Participants:** Marc Antonini [Laboratoire I3S, Sophia Antipolis, France], Pierre Kornprobst, Khaled Masmoudi [Laboratoire I3S, Sophia Antipolis, France].

In [51] we presented a novel bio-inspired and dynamic coding scheme for static images. Our coder aims at reproducing the main steps of the visual stimulus processing in the mammalian retina taking into account its time behavior. The main novelty of this work is to show how to exploit the time behavior of the retina cells to ensure, in a simple way, scalability and bit allocation. To do so, our main source of inspiration has been the biologically plausible retina model Virtual Retina described in Section 4.1. Following a similar structure, our model has two stages. The first stage is an image transform which is performed by the outer layers in the retina. Here it is modelled by filtering the image with a bank of difference of Gaussians with time-delays. The second stage is a time-dependent analog-to-digital conversion which is performed by the inner layers in the retina. Thanks to its conception, our coder enables scalability and bit allocation across time. Also, compared to the JPEG standards, our decoded images do not show annoying artefacts such as ringing and block effects. As a whole, this article shows how to capture the main properties of a biological system, here the retina, in order to design a new efficient coder.

5.3. Mean field methods

5.3.1. Noise-induced behaviors in neural mean field dynamics

**Participants:** Jonathan Touboul, Geoffroy Hermann, Olivier Faugeras.

The collective behavior of cortical neurons is strongly affected by the presence of noise at the level of individual cells. In order to study these phenomena in large-scale assemblies of neurons, we consider networks of firing-rate neurons with linear intrinsic dynamics and nonlinear coupling, belonging to a few types of cell populations and receiving noisy currents. Asymptotic equations as the number of neurons tends to infinity (mean field equations) are rigorously derived based on a probabilistic approach. These equations are implicit on the probability distribution of the solutions which generally makes their direct analysis difficult. However, in our case, the solutions are Gaussian, and their moments satisfy a closed system of nonlinear ordinary differential equations (ODEs), which are much easier to study than the original stochastic network equations, and the statistics of the empirical process uniformly converge towards the solutions of these ODEs. Based on this description, we analytically and numerically study the influence of noise on the collective behaviors, and compare these asymptotic regimes to simulations of the network. We observe that the mean field equations provide an accurate description of the solutions of the network equations for network sizes as small as a few hundreds of neurons. In particular, we observe that the level of noise in the system qualitatively modifies its collective behavior, producing for instance synchronized oscillations of the whole network, desynchronization of oscillating regimes, and stabilization or destabilization of stationary solutions. These results shed a new light on the role of noise in shaping collective dynamics of neurons, and gives us clues for understanding similar phenomena observed in biological networks.
This work has been accepted for publication in the SIAM Journal on Applied dynamical Systems [72].

5.3.2. Mean Field description of and propagation of chaos in recurrent multipopulation networks of Hodgkin-Huxley and FitzHugh-Nagumo neurons

Participants: Javier Baladron, Diego Fasoli, Olivier Faugeras, Jonathan Touboul.

We derive the mean-field equations arising as the limit of a network of interacting spiking neurons, as the number of neurons goes to infinity. The neurons belong to a fixed number of populations and are represented either by the Hodgkin-Huxley model or by one of its simplified version, the Fitzhugh-Nagumo model. The synapses between neurons are either electrical or chemical. The network is assumed to be fully connected. The maximum conductances vary randomly. Under the condition that all neurons initial conditions are drawn independently from the same law that depends only on the population they belong to, we prove that a propagation of chaos phenomenon takes places, namely that in the mean-field limit, any finite number of neurons become independent and, within each population, have the same probability distribution. This probability distribution is solution of a set of implicit equations, either nonlinear stochastic differential equations resembling the McKean-Vlasov equations, or non-local partial differential equations resembling the McKean-Vlasov-Fokker-Planck equations. We prove the well-posedness of these equations, i.e. the existence and uniqueness of a solution. We also show the results of some preliminary numerical experiments that indicate that the mean-field equations are a good representation of the mean activity of a infinite size network, even for modest sizes. These experiment also indicate that the McKean-Vlasov-Fokker-Planck equations may be a good way to understand the mean-field dynamics through, e.g., a bifurcation analysis.

This work has been submitted for publication in the Journal of Mathematical Neuroscience [55].

5.3.3. Three applications of GPU computing in neuroscience

Participants: Javier Baladron, Olivier Faugeras.

GPUs are low cost highly parallel devices that are now not only used for graphics but also for numerical simulation. We present three applications of a computer system with multiple GPUs to the domain of theoretical neuroscience. The first application is to a continuous model of the primary visual area, the second to the simulation of a stochastic neural network, and the third to the computation of the probability distribution on the possible states of a network. In all three cases we show that the speed-up obtained by the use of GPUs has considerably helped answering a scientific or technological question.

This work has been accepted for publication in Computing in Science and Engineering [63].

5.4. Neural Fields

5.4.1. Modelling the dynamics of contextual motion integration in the primate

Participants: Heiko Neumann [Institute of Neural Information Processing, Ulm University, Ulm, Germany], Pierre Kornprobst, Guillaume Masson [Institut de Neurosciences de la Timone, UMR 6193, CNRS, Marseille, France], Emilien Tlapale.

The dynamics of motion integration show striking similarities when observed at neuronal, psychophysical, and oculomotor levels. Based on the inter-relation and complementary insights given by those dynamics, our goal is to investigate how basic mechanisms of dynamical cortical processing can be incorporated in a dynamical model to solve several aspects of 2D motion integration and segmentation.

Thanks to Emilien Tlapale PhD [13] (see also [16]), we have obtained the following results:

- We proposed a recurrent model of motion integration. Proposing a simple readout mechanism, we reproduced not only motion perception but also the dynamics of smooth pursuit eye movements on various line figures and gratings viewed through different apertures. Our model can also solve various contextual problems where extrinsic junctions should be eliminated, without relying on complex junction detectors or depth computation [71]. Finally, we have also shown how our model can be rewritten in the neural fields formalism (see [52] and the Software MotionLib), which has opened new perspectives as detailed in Section 5.4.2.
• We confronted our results to artificial and biological vision. To formalize the comparison against visual performance, we proposed a new evaluation methodology based on human visual performance by establishing a database of image sequences taken from biology and psychophysics literature [70], [69], [67]. We compared our results against the state of the art computer vision approaches and we found that our model also gives results comparable to recent computer vision approaches of motion estimation.

5.4.2. Neural fields models for motion integration: Characterising the dynamics of multi-stable visual motion stimuli

Participants: Olivier Faugeras, Pierre Kornprobst, Guillaume Masson [Institut de Neurosciences de la Timone, UMR 6193, CNRS, Marseille, France], Andrew Meso [Institut de Neurosciences de la Timone, UMR 6193, CNRS, Marseille, France], James Rankin, Emilien Tlapale, Romain Veltz.

In [57] we investigated the temporal dynamics of the neural processing of a multi-stable visual motion stimulus with two complementary approaches: psychophysical experiments and mathematical modelling. The so-called “barber pole” stimulus is considered with an aperture configuration that supports horizontal (H), diagonal (D) or vertical (V) perceived directions for the same input. The phenomenon demonstrates an interesting variable and dynamic competition for perceptual dominance between underlying neural representations of the three directions. We probe the early processing from stimulus presentation to initial perceived direction (before perceptual reversals). Starting from a simplified neural fields model inspired from [13], we constructed a model of the necessary motion integration that shows a shift in perceptual dominance from D to either H or V with increasing duration. Further, the timing of this shift is shown to be controlled by a stimulus gain parameter analogous to contrast. In psychophysics experiments with concurrent eye movement recordings, observers report their perceived direction of motion for presentation durations between 0.1s and 0.5s. There is a also consistent transition in perceptual dominance from D to H/V as duration is increased. This trend, seen in both perceived direction decisions and eye movement patterns, is consistent with previous experiments using similar stimuli with an aperture configured for two (D/H) rather than three (D/H/V) states. The basic dynamic properties of the early transition from D to H/V are well predicted by the model. The experimental work additionally reveals asymmetric data patterns that guide adjustments to the model’s input equations. Observers have an H bias over V, which is also reflected in faster reaction times for H. In order to capture the bias between H and V a separate weighting is attributed to the local input corresponding to each state. The work presented forms a solid foundation for future experimental and modelling work investigating the longer term dynamics for which perceptual reversals are known to occur.

5.4.3. Analysis of a hyperbolic geometric model for visual texture perception

Participants: Pascal Chossat, Grégory Faye, Olivier Faugeras.

We study the neural field equations introduced by Chossat and Faugeras in [64] to model the representation and the processing of image edges and textures in the hypercolumns of the cortical area V1. The key entity, the structure tensor, intrinsically lives in a non-Euclidean, in effect hyperbolic, space. Its spatio-temporal behaviour is governed by nonlinear integro-differential equations defined on the Poincaré disc model of the two-dimensional hyperbolic space. Using methods from the theory of functional analysis we show the existence and uniqueness of a solution of these equations. In the case of stationary, i.e. time independent, solutions we perform a stability analysis which yields important results on their behavior. We also present an original study, based on non-Euclidean, hyperbolic, analysis, of a spatially localised bump solution in a limiting case. We illustrate our theoretical results with numerical simulations.

This work has been published in the Journal of Mathematical Neuroscience [21].

5.4.4. Bifurcation of Hyperbolic Planforms

Participants: Pascal Chossat, Grégory Faye, Olivier Faugeras.
Motivated by a model for the perception of textures by the visual cortex in primates, we analyze the bifurcation of periodic patterns for nonlinear equations describing the state of a system defined on the space of structure tensors, when these equations are further invariant with respect to the isometries of this space. We show that the problem reduces to a bifurcation problem in the hyperbolic plane \( D \) (Poincaré disc). We make use of the concept of a periodic lattice in \( D \) to further reduce the problem to one on a compact Riemann surface \( D/\Gamma \), where \( \Gamma \) is a cocompact, torsion-free Fuchsian group. The knowledge of the symmetry group of this surface allows us to use the machinery of equivariant bifurcation theory. Solutions which generically bifurcate are called “H-planforms”, by analogy with the “planforms” introduced for pattern formation in Euclidean space. This concept is applied to the case of an octagonal periodic pattern, where we are able to classify all possible H-planforms satisfying the hypotheses of the Equivariant Branching Lemma. These patterns are, however, not straightforward to compute, even numerically, and in the last section we describe a method for computation illustrated with a selection of images of octagonal H-planforms.

This work has been published in the Journal of Nonlinear Science [19].

5.4.5. Bifurcation diagrams and heteroclinic networks of octagonal H-planforms

**Participants:** Grégory Faye, Pascal Chossat [correspondent].

This paper completes the classification of bifurcation diagrams for H-planforms in the Poincaré disc \( D \) whose fundamental domain is a regular octagon. An H-planform is a steady solution of a PDE or integro-differential equation in \( D \), which is invariant under the action of a lattice subgroup \( \Gamma \) of \( U(1,1) \), the group of isometries of \( D \). In our case \( \Gamma \) generates a tiling of \( D \) with regular octagons. This problem was introduced in [19] as an example of spontaneous pattern formation in a model of image feature detection by the visual cortex where the features are assumed to be represented in the space of structure tensors. Under “generic” assumptions the bifurcation problem reduces to an ODE which is invariant by an irreducible representation of the group of automorphisms \( \mathcal{G} \) of the compact Riemann surface \( D/\Gamma \). The irreducible representations of \( \mathcal{G} \) have dimension one, two, three and four. The bifurcation diagrams for the representations of dimension less than four have already been described and correspond to already well known group actions. In the present work we compute the bifurcation diagrams for the remaining three irreducible representations of dimension four, thus completing the classification. In one of these cases, there is generic bifurcation of a heteroclinic network connecting equilibria with two different orbit types.

This work has been accepted for publication in the Journal of Nonlinear Science [22].

5.4.6. Hopf bifurcation curves in neural field networks with space-dependent delays

**Participant:** Romain Veltz.

We give an analytical parametrization of the curves of purely imaginary eigenvalues in the delay-parameter plane of the linearized neural field network equations with space-dependent delays. In order to determine if the rightmost eigenvalue is purely imaginary, we have to compute a finite number of such curves; the number of curves is bounded by a constant for which we give an expression. The Hopf bifurcation curve lies on these curves.

This work has appeared in the Comptes Rendus Mathématiques de l’Académie des Sciences [30].

5.4.7. Stability of the stationary solutions of neural field equations with propagation delays

**Participants:** Olivier Faugeras, Romain Veltz.

We consider neural field equations with space-dependent delays. Neural fields are continuous assemblies of mesoscopic models arising when modeling macroscopic parts of the brain. They are modeled by nonlinear integro-differential equations. We rigorously prove, for the first time to our knowledge, sufficient conditions for the stability of their stationary solutions. We use two methods 1) the computation of the eigenvalues of the linear operator defined by the linearized equations and 2) the formulation of the problem as a fixed point problem. The first method involves tools of functional analysis and yields a new estimate of the semigroup of the previous linear operator using the eigenvalues of its infinitesimal generator. It yields a sufficient condition for stability which is independent of the characteristics of the delays. The second method allows us to find
new sufficient conditions for the stability of stationary solutions which depend upon the values of the delays. These conditions are very easy to evaluate numerically. We illustrate the conservativeness of the bounds with a comparison with numerical simulation.

This work has appeared in the Journal of Mathematical Neuroscience [29].

5.4.8. Neural Mass Activity, Bifurcations and Epilepsy

Participants: Patrick Chauvel [INSERM U751, Marseille, Assistance Publique-Hopitaux de Marseille Timone, and Universite Aix-Marseille, Marseille], Olivier Faugeras, Jonathan Touboul, Fabrice Wendling [INSERM, U642, Rennes].

We propose a general framework for studying neural mass models defined by ordinary differential equations. By studying the bifurcations of the solutions to these equations and their sensitivity to noise we establish an important relation, similar to a dictionary, between their behaviors and normal and pathological, especially epileptic, cortical patterns of activity. We then apply this framework to the analysis of two models that feature most phenomena of interest, the Jansen and Rit model, and the slightly more complex model recently proposed by Wendling and Chauvel. This model-based approach allows to test various neurophysiological hypotheses on the origin of pathological cortical behaviors and to investigate the effect of medication. We also study the effects of the stochastic nature of the inputs which gives us clues about the origins of such important phenomena as interictal spikes, inter-ical bursts and fast onset activity, that are of particular relevance in epilepsy.

This work has appeared in Neural Computation [27].
5. New Results

5.1. Analysis of structures resulting from meristem activity

5.1.1. Acquisition and design of plant geometry

Participants: Chakkrit Preuksakarn, Frédéric Boudon, Christophe Pradal, Christophe Godin.

This research theme is supported by RTRA and ARC projects named PlantScan3D.

Virtual 3D model of plants are required in many areas of plant modeling. They can be used for instance to simulate physical interaction of real plant structures with their environment (light, rain, wind, pests, ...), to set up initial conditions of growth models or to assess their output against real data. In the past decade, methods have been developed to digitize plant architectures in 3D [42], [33]. These methods are based on direct measurements of position and shape of every plant organ in space. Although they provide accurate results, they are particularly time consuming. More rapid and automated methods are now required in order to collect plant architecture data of various types and sizes in a systematic way. In this aim, we explore the use of pictures, video, laser scanner and direct sketching.

- **Reconstruction of plant architecture from 3D laser scanner data.** (Chakkrit Preuksakarn, Mathilde Balduzzi, Frédéric Boudon, Jean-Baptiste Durand, Christophe Godin, Xinghua Song [INRIA, Galaad], Bernard Mourrain [INRIA, Galaad], Dobrina Boltcheva [INRIA, Imagine], Franck Herty [INRIA, Morpheus], Marie-Paule Cani [Inria, Imagine], Pascal Ferraro [Labri, Bordeaux])

  We investigate the possibility to use 3D laser scanners to automate plant digitizing. We are developing algorithms to reconstruct branching systems without leaves or foliage from scanner data or from scan simulated on plant mock-up obtained using different digitizing method. For this we collaborate with the EPI Galaad from Sophia-Antipolis, the EPI Imagine from Grenoble, different INRA teams, UMR PIAF in Clermont Ferrand, UMR LEPSE and AFEF team in Montpellier and Lusignan, the University of Helsinki, Finland and the CFCC in England.

  We developed a reconstruction pipeline composed of several procedures. A contraction procedure, first aggregates points at the center of the point cloud. The team proposed a simple adaptive scheme to contract points. In parallel, the Galaad team explored uses of detection of circular patterns to be contracted toward their center. Comparison of these approaches has been carried out. In a second step, a skeleton procedure uses a Space Colonization Algorithm [41] to build the skeleton of the shape from the contracted point set. This method is adaptive to the local density of the point set. Then a pipe-model based procedure makes it possible to estimate locally diameters of the branches. Finally, an evaluation procedure has been designed to assess the accuracy of the reconstruction. Results Publication of this work is in progress.

- **Sketching of plants.** (Frédéric Boudon, Christophe Godin, Steven Longuay [University of Calgary, Canada], Przemyslaw Prusinkiewicz [University of Calgary, Canada])

  Modeling natural elements such as trees in a plausible way, while offering simple and rapid user control, is a challenge. In a first collaboration with the EPI Evasion we developed a method based on the design of plants from silhouettes [43]. This sketching paradigm allows quick and intuitive specification of foliage at multiple scales. On this topic, we started a collaboration with S. Longuay and P. Prusinkiewicz who develop iPad tools to design plants based on SCA. Combination of multitouch interface, sketching paradigm and powerful adaptive procedural model that generate realistic trees offer intuitive and flexible design tools. This work is part of the INRIA associated team with the University of Calgary.

- **Reconstruction of vineyards from video.** (Frédéric Boudon, Jerome Guenard [IRIT, Toulouse], Geraldine Morin [IRIT, Toulouse], Pierre Gurdjos [IRIT, Toulouse], Vincent Charvillat [IRIT, Toulouse])
In this work, we investigate the reconstruction of constrained plant geometry of a vineyard from a set of pictures coming from video. Pictures are segmented to identify the different trees of a same row in the yard. From this segmentation, a number of parameters are estimated, which makes it possible to instantiate a virtual model of a vine tree. In particular, paths of main branches and leaf volumes and densities are estimated. A preliminary version of this work has been presented to the AFIG conference [21].

- **Reconstruction of virtual fruits from pictures.** (Mik Cieslak, Nadia Bertin [Inra, Avignon], Frédéric Boudon, Christophe Godin, Christophe Pradal, Michel Genard [Inra, Avignon], Christophe Goz-Bac [Université Montpellier 2])

This research theme is supported by the Agropolis project Fruit3D.

Understanding the controlling factors of fruit quality development is challenging, because fruit quality is the result of the interplay between physical and physiological processes that are under the control of genes and the environment. Although process-based models have been used to make significant progress in understanding these factors, they to a large extent ignore the shape and internal structure of the fruit. To help characterizing effects of fruit shape and internal structure on quality, the creation of a 3D virtual fruit model that integrates fruit structure and function with growth governed by environmental inputs has been investigated. For this purpose, a modelling pipeline has been created that includes the following steps: creation of a 3D volumetric mesh of the internal and external fruit structure, calculation of the fruit’s physical properties from the resulting mesh, and integration of aspects of fruit physiology into the 3D structure. This pipeline has been applied to study tomato fruit (Solanum lycopersicum) by constructing 3D volumetric meshes from two images of perpendicular fruit slices and from MRI data, and integrating water and carbon transport processes into one of these meshes. To illustrate the tomato model, a simulation of one season’s of the fruit’s growth has been performed and its results compared with an already published process-based tomato fruit model. The results of the two models were in general agreement, but our model provided additional information on the internal properties of the fruit, such as a gradient in sugar concentration. Once the model is calibrated and evaluated, our approach will be suitable for studying the effects of internal fruit heterogeneity and overall shape on fruit quality development [18].

5.1.2. Modeling the plant ontogenic programme

**Participants:** Christophe Godin, Yann Guédon, Evelyne Costes, Jean-Baptiste Durand, Pascal Ferraro, Anaëlle ambreville, Christophe Pradal, Catherine Trottier, Jean Peyhardi, Yassin Refahi, Etienne Farcot.
This research theme is supported by two PhD programmes.

The remarkable organization of plants at macroscopic scales may be used to infer particular aspects of meristem functioning. The fact that plants are made up of the repetition of many similar components at different scales, and the presence of morphological gradients, e.g., [23], [35], [36], [32], provides macroscopic evidence for the existence of regularities and identities in processes that drive meristem activity at microscopic scales. Different concepts have been proposed to explain these specific organizations such as “morphogenetic programme” [38], “age state” [31] or “physiological age” [24]. All these concepts state that meristem fate changes according to position within the plant structure and during its development. Even though these changes in meristem fate are specific to each species and lead to the differentiation of axes, general rules can be highlighted [31], [24]. Here we develop computational methods to decipher these rules.

- **Branching and axillary flowering structures of fruit tree shoots.** (Yann Guédon, Evelyne Costes, David Da Silva [UC Davis], Anna Davidson [UC Davis], Ted DeJong [UC Davis], Claudia Negron [UC Davis]).

  In the context of a collaboration with Claudia Negron, Anna Davidson, David Da Silva and Ted DeJong, stochastic models (hidden semi-Markov chains) for the branching and axillary flowering structures of different categories of peach and almond shoots corresponding to different genetic backgrounds, environment conditions and horticultural practices were built. These stochastic models have been integrated in simulation systems which combine stochastic models with different mechanistic models of biological function, in particular carbon partitioning models. This collaboration extends the work initiated on apple trees [40], [3]; see 5.1.3.

- **Genetic determinisms of the alternation of flowering in apple tree progenies.** (Jean-Baptiste Durand, Jean Peyhardi, Baptiste Guitton [DAP, AFEF team], Catherine Trottier, Evelyne Costes, Yann Guédon)

  Previous approaches for a statistical quantification of the effect of factors on tree architecture were mainly oriented toward the structure of main axes, and environmental explanatory variables (see [34], [27]). To characterize genetic determinisms of the alternation of flowering in apple tree progenies at annual shoot (AS) scale, a model of the transitions between ASs was built. The ASs were of two types: flowering or vegetative. Two replications of each genotype were available. Our model operated on tree-structured data and relied on a second-order Markov tree. Generalized Linear Mixed Models (GLMMs) were used to model the effect of year, replications and genotypes (with their interactions with year or memories of the Markov model) on the transition probabilities. This work was the continuation of the Master 2 internship of Jean Peyhardi (Bordeaux 2 University) and was carried out in the context of the PhD thesis of Baptiste Guitton.

  This PhD thesis also comprised the study of alternation in flowering at individual scale, with annual time step. To relate alternation of flowering at AS and individual scales, indices were proposed to characterize alternation at individual scale. The difficulty is related to early detection of alternating genotypes, in a context where alternation is often concealed by a substantial increase of the number of flowers over consecutive years. To separate correctly base effect from alternation in flowering, our model relied on a parametric hypothesis on the base effect (random slopes specific to genotype and replications), which translated into mixed effect modeling. Different indices of alternation were then computed on the residuals. Clusters of individuals with contrasted patterns of bearing habits were identified. Our models highlighted significant correlations between indices of alternation at AS and individual scales. The roles of local alternation and asynchronism in regularity of flowering were assessed using an entropy-based criterion, which characterized asynchronism.

  As a perspective of this work, patterns in the production of children ASs (numbers of flowering and vegetative children) depending on the type of the parent AS must be analyzed using branching processes and different types of Markov trees, in the context of Pierre Fernique’s PhD Thesis.

- **Modeling branching patterns in fruit tree shoots through the characterization of their demographic properties** (Pierre Fernique, Jean-Baptiste Durand, Yann Guédon).
To test the effect of some properties of a given parent shoot on the properties of its children shoots, statistical models based on multitype branching processes were developed. This kind of dependence between parent and children shoots is frequently at stake in fruit trees, for which the number of flowering or vegetative children of a parent shoot depends on its nature, with potential interactions with other factors. Thus, controlling demographic patterns of the shoots (through varietal selection or crop management strategies) is expected to bring substantial improvements in the quantity and quality of yields.

Formally, the shoot properties are summed up using the notion of shoot state. The number of children shoots in each state is modeled through discrete multivariate distributions. Model selection procedures are necessary to specify parsimonious distributions. We developed an approach based on probabilistic graphical models to identify and exploit properties of conditional independence between numbers of children in different states, so as to simplify the specification of their joint distribution. The graph building stage was based on a Poissonian Generalized Linear Model for the contingency tables of the counts of joint children state configurations. Then, parametric families of distributions were implemented and compared statistically to provide probabilistic models compatible with the estimated independence graph.

This work was carried out in the context of Pierre Fernique’s Master 2 internship (Montpellier 2 University and AgroParisTech). It was applied to model dependencies between short or long, vegetative or flowering shoots in apple trees. The results highlighted contrasted patterns related to the parent shoot state, with interpretation in terms of alternation of flowering. This work will be continued during Pierre Fernique’s PhD thesis, with extensions to other fruit tree species (mango trees) and other strategies to build probabilistic graphical models and parametric discrete multivariate distributions including covariates and mixed effects.

- **Analyzing fruit tree phenology** (Anaëlle Dambreville, Jean-Baptiste Durand, Yann Guédon, Christophe Pradal, Pierre-Eric Lauri [UMR AGAP], Frédéric Normand [UPR HortSys], Catherine Trottier)
  Mango is a tropical tree characterized by strong asynchronisms within and between trees. Causation networks explaining the vegetative and reproductive growths within and between growing cycles were studied on the basis of generalized linear models. We highlighted in this way marked interplays between structural and temporal components of tree structure development at three scales. At growth unit scale, a growth unit appeared early in the growing cycle had higher rate of burst compared to late appeared growth units. At growing cycle scale, a growth unit which flowered delayed its future vegetative growth compared to a vegetative growth unit. At tree scale, a fruiting tree delayed further vegetative growth and flowering compared to a non-fruiting tree. These results evidenced that tree phenology is strongly affected by structural components and not only by the environment.

- **Self-nested structure of plants.** (Christophe Godin, Pascal Ferraro)
  To study the redundancy of structures embedded at various levels in tree architectures, we investigated the problem of approximating trees by trees with particular self-nested structures. Self-nested trees are such that all their subtrees of a given height are isomorphic. We show that these trees present remarkable compression properties, with high compression rates. In order to measure how far a tree is from being a self-nested tree, we introduced a quantitative measure of the degree of self-nestedness for any tree. For this, we want to find a self-nested tree that minimizes the distance of the original tree to the set of self-nested trees that embed the initial tree:

\[ NEST(T) = \arg \min_{S \in S^+(T)} D(T, S), \]

where \( T \) is a tree, \( D(\cdot, \cdot) \) is a distance on the set of trees (chosen so as to preserve certain structural properties between the compared trees) and \( S^+(T) \) is the set of self-nested trees that contain \( T \), i.e. that can be obtained from \( T \) by inserting nodes only. In a previous work we showed that this problem can be solved in polynomial time and gave the corresponding algorithm [6].
Now, we continue this work along different directions:

- **Approximate compression including geometry** (Anne-Laure Gaillard, Pascal Ferraro, Frédéric Boudon, Christophe Godin)

  We now investigate how to include branch geometry in the compression schemes. This problem constitutes a part of the PhD thesis of Anne-Laure Gaillard co-supervised with P. Ferraro. First results show that any tree architecture can be compressed at different degrees with a varying loss in the geometric information. A publication on this topic is currently in progress.

- **Search for the NST** (Farah Ben-Naoum, Christophe Godin, Pascal Ferraro)

  The NEST algorithm constructs a closest self-nested tree that embeds a given tree $T$. This means in particular that $T \in \text{NEST}(T)$. For some trees this definition might be too restrictive and one may want to get rid of the latter constraint. In this case, the closest self-nested tree (NST) to a tree $T$ would be defined by:

  $$\text{NST}(T) = \arg \min_{S \in S(T)} D(T, S),$$

  where $S(T)$ is now the set of all self-nested trees. Although it has not yet been demonstrated, the computation of $\text{NST}(T)$ is likely to be a NP-complete problem. Therefore we decided to develop a heuristic algorithm to carry out this optimization. This algorithm is based on evolutionary optimization algorithms. First results obtained at the University of Sidi Bel Abbès by Farah Ben Naoum are very encouraging. They should contribute to define a general method to compress tree-like structures based on self-similarity.

- **Stochastic NESTs** (Jean-Baptiste Durand, Christophe Godin, Pascal Ferraro, Yann Guédon)

  One of the limits of the original NEST algorithm in defining a compressed structure is that it does not take into account possible small variations in the identification of similar tree patterns in the global tree structure. In real trees, it is however seldom the case that similar tree parts are exactly isomorphic. They may differ by small details while keeping globally similar. A general compression technique should account for such variations from a theoretical point of view. This is what we intend to do here by developing a stochastic NEST approach. The idea is to extend the definition of tree compression to include the notion of tree distribution in compressed representations of trees. A first formalism has been designed for stochastic tree compression.

- **Statistical characterization of the branching properties of trees at individual scale – application to the quantification of approximate self-nestedness.** (Jean-Baptiste Durand, Yann Guédon, Christophe Godin)

  To test different hypotheses related to the role of apical control on local branching properties within tree architecture, statistical models of the fates of the apical meristem and its production at a given scale were specified. They extend the hidden Markov tree models [29] to variable number of children of a given entity, and provide a model for their dependencies. Moreover, they allow for the comparison of different hypotheses regarding the relevance of the children ordering (absence of ordering, partial or total ordering). In the next years, these models will be used on mango and apple tree cultivars. Particularly, they will found new approaches to compare the effect of various strategies of culture.

  These models are based on a notion of state that extends the concept of equivalence classes for tree isomorphism used in the above paragraph, to the notion of classes of approximate isomorphism (i.e. isomorphism as a stochastic process). As a consequence, the hidden Markov tree models offer new insight for lossy compression of trees, which will be investigated in future work.

- **Analyzing perturbations in Arabidopsis thaliana phyllotaxis.** (Yann Guédon, Yassin Refahi, Etienne Faricot, Christophe Godin, Fabrice Besnard [RDP, Lyon], Teva Vernoux [RDP, Lyon])
The cytokinin hormones are known to play a significant role in the regulation of phyllotaxis. To investigate this, Fabrice Besnard and Teva Vernoux are studying Arabidopsis thaliana ahp6 mutants, AHP6 being a protein known for its inhibitory effect in the cytokinin signaling pathway. At the macroscopic scale, this mutation induces perturbations of the phyllotaxis, barely sensible on single plants. In order to characterize these perturbations, we designed a pipeline of models and methods which decompose into three steps: (i) identification of perturbation patterns, (ii) characterization of perturbation patterns using hidden variable-order Markov chains and combinatorial mixture models both with von Mises observation distributions (Gaussian-like periodic distribution for circular variables), (iii) classification of plant phyllotaxis among wild-type and ahp6 mutant phyllotaxies. Using this pipeline of methods, we have shown that the perturbation patterns in both wild-type and mutant plants can be explained by permutations in the order of insertion along the stem of 2 or 3 consecutive organs. The number of successive synchronized organisms between two permutations reveals unexpected patterns that depend on the nature of the preceding permutation (2- or 3-permutation). We identified significant individual deviations of the level of baseline segments with reference to 137.5°, which confirms theoretical model predictions. Finally, we highlighted a marked relationship between permutation of organs and defects in the elongation of the internodes in between these organs. All these results can be explained by the absence of a strict coupling between the timing of organ development and their angular and longitudinal position on the stem. A paper about these results is in revision for the journal Science. Another more methodological paper is in progress.

5.1.3. Analyzing the influence of the environment on the plant ontogenic programme

Participants: Jean-Baptiste Durand, Damien Fumey, Frédéric Boudon, Christophe Godin, Yann Guédon, Jean Peyhardi, Pierre Fernique, Christian Cilas, Evelyne Costes, Pascal Ferraro, Catherine Trottier.

This research theme is supported by a CIFRE contract and two PhD programmes.

The ontogenetic programme of a plant is actually sensitive to environmental changes. If, in particular cases, we can make the assumption that the environment is a fixed control variable (see section 5.1.2), in general the structure produced by meristem results from a tight interaction between the plant and its environment, throughout its lifetime. Based on observations, we thus aim to trace back to the different components of the growth (ontogenetic development and its modulation by the environment). This is made using two types of approaches. On the one hand, we develop a statistical approach in which stochastic models are augmented with additional time-varying explanatory variables that represent the environment variations. The design of estimation procedures for these models make it possible to separate the plant ontogenetic programme from its modulation by the environment. On the other hand, we build reactive models that make it possible to simulate in a mechanistic way the interaction between the plant development and its environment.

- Analyzing growth components in trees. (Yann Guédon, Jean-Baptiste Durand, Jean Peyhardi, Yves Caraglio [AMAP], Emilie Lebarbier [AgroParisTech], Catherine Trottier, Olivier Taugourdeau [AMAP])

Observed growth, as given for instance by the length of successive annual shoots along a forest tree trunk, is assumed to be mainly the result of three components: (i) an endogenous component assumed to be structured as a succession of roughly stationary phases separated by marked change points that are asynchronous between individuals [34], (ii) a time-varying environmental component assumed to take the form of fluctuations that are synchronous between individuals, (iii) an individual component corresponding to the local environment of each tree. This environmental component is thus assumed to be a "population" component as opposed to the individual component. In order to identify and characterize these three components, we proposed to use Markov and semi-Markov switching linear mixed models [27] [2]. The underlying Markov or semi-Markov chain represents the succession of growth phases (endogenous component) while the linear mixed model attached to each state of the underlying Markov or semi-Markov chain represents -in the corresponding growth phase- both the influence of time-varying climatic explanatory variables (environmental component) as fixed effects, and inter-individual heterogeneity (individual component) as random effects. We
investigated the estimation of Markov and semi-Markov switching linear mixed models in a general framework using MCEM-like algorithms. These integrative statistical models were in particular applied in a forest ecology context to characterize the opportunistic development of understory samplings in relation to light environment [15]. Concerning the application to forest trees, the proposed statistical modeling approach relies on the availability of climatic data. In the case where climatic data are not available, we are studying Markov and semi-Markov switching linear mixed models with year random effects common to all the trees to model the synchronous part of the growth fluctuations. With Markov and semi-Markov switching linear mixed models, the response variable is constrained to be approximately normally distributed. We are now studying the statistical methodology for Markov switching linear mixed models to take into account non-normally distributed response variables (e.g., number of growth units, apex death/life, non-flowering/flowering character). It should be noted that the estimation algorithms proposed for Markov switching linear mixed models can be directly transposed to other families of hidden Markov models such as, for instance, hidden Markov tree models; see Section 5.1.2.

- Coupling stochastic models with mechanistic models for plant development simulation. (Damien Fumey, Yann Guédon, Christophe Godin, Thomas Cokelaer, Evelyne Costes, Pierre-Eric Lauri [UMR AGAP])

Arboricultural practices such as pruning, artificial bending or fruit thinning are crucial interventions in orchard management and are used for controlling tree size, penetration of light into the canopy and the equilibrium between vegetative and reproductive growth. In the PhD of Damien Fumey we explored the possibility of integrating such practices in a model of apple tree development. To this end, a field experiment was designed to study the effects of pruning (thinning or heading cuts) on two apple cultivars with contrasted architecture, ‘Fuji’ and ‘Braeburn’. Results of this experiment [30] showed that thinning cuts of laterals tended to be compensated by an increase in lateral branching. Based on these field experiments, a model is currently being developed to account for pruning practices on fruit trees. This model relies on a formalization of the competition of meristems by combining a carbon allocation strategy and a competition mechanism in a stochastic manner. The resulting model is reactive to human interventions and should enable us to capture plant reactions to pruning practices in a robust way [28]. A paper describing this model is currently in progress.

5.2. Meristem functioning and development

In axis 2 work focuses on the creation of a virtual meristem, at cell resolution, able to integrate the recent results in developmental biology and to simulate the feedback loops between physiology and growth. The approach is subdivided into several sub-areas of research.

5.2.1. Data acquisition and design of meristem models

Participants: Frédéric Boudon, Christophe Godin, Vincent Mirabet, Jan Traas, Grégoire Malandain, Jean-Luc Verdeil.

This research theme is supported by the ATP CIRAD Meristem and the ANR GeneShape and FlowerModel projects.

Studies on plant development require the detailed observation of the tissue structure with cellular resolution. In this context it is important to develop methods that enable us to observe the inner parts of the organs, in order to analyze and simulate their behavior. Here we focus on the apical meristems, that have been extensively studied using live imaging techniques and confocal microscopy. An important limitation of the confocal microscope lies in the data anisotropy. To overcome this limitation, we designed new protocols to achieve an accurate segmentation of the cells. Using these segmentations, a geometrical and topological representation of the meristem is built. Such representations may be used to analyze the meristem structure at cell level, to support the description of gene expression patterns and to initiate and assess virtual meristem simulations.

- Microscopy image reconstruction and automatic lineage tracking of the growing meristem cells
Participants: Romain Fernandez, Christophe Godin, Grégoire Malandain, Jan Traas, Pradeep Das, Vincent Mirabet.

In previous work [5], we studied the tracking of meristem cells using time-lapse confocal microscopy acquisition on early stages flowers of Arabidopsis shoot apical meristems. We designed a reconstruction method (MARS, Figure 4) and a tracking algorithm (ALT) in order to map the segmentations of the same meristem at different times, based on a network flow representation in order to solve the cell assignment problem. We validated the MARS-ALT pipeline on a four-steps time course of an early stage floral bud. In 2011, we worked to improve the robustness of the MARS-ALT pipeline. The software pipeline was completely re-engineered so as to be easily available in OpenAlea with documentation. We also designed new tests to improve the method and get high-quality results on different types of organisms.

Figure 4. Surface view of a flower meristem automatically segmented using MARS at cell resolution and a transversal cut showing the inner segmented tissues

- Design of 3D virtual maps for specifying gene expression patterns (Jérôme Chopard, Christophe Godin, Jan Traas, Françoise Monéger [ENS Lyon])

This research theme is supported the ANR GeneShape and FlowerModel projects.

To organize the various genetic, physiological, physical, temporal and positional informations, we build a spatialized and dynamic database. This database makes it possible to store all the collected information on a virtual 3D structure representing a typical organ. Each piece of information has to be located spatially and temporally in the database. Tools to visually retrieve and manipulate the information, quantitatively through space and time are being developed. For this, the 3D structure of a typical organ has been created at the different stages of development of the flower bud. This virtual structure contains spatial and temporal information on mean cell numbers, cell size, cell lineages, possible cell polarization (transporters, microtubules), and gene expression patterns. Such 3D virtual map is mainly descriptive. However, like for classical databases, specific tools make it possible to explore the virtual map according to main index keys, in particular spatial and temporal keys. Both a dedicated language and a 3D user interface are being designed to investigate and query the 3D virtual map.
A prototype version of this 3D virtual map is currently being built and is integrated in V-Plants, see figure 5 where a cell-based volumic tissue that can contain different types of information (cell lineage, cell size, cell identity, etc...).

![Virtual Plant Map](image)

**Figure 5. A 3D database represented as a cell-based map that can contain different types of information (cell lineage, cell size, cell identity, ...), from [14].**

Using this 3D virtual map prototype, along with piecewise-linear models of gene networks, and optimization techniques, our colleagues from ENS/Lyon have been able to develop a model of the gene network regulating sepal polarization in *A. thaliana*. This model, which is consistent with a very comprehensive literature data set, also include new predictions for certain gene interactions, and will appear next year in Plant Cell.[14]

### 5.2.2. Shape analysis of meristems

(Jean-Baptiste Durand, Frederic Boudon, Yann Guedon, Francois Mankessi [BURST, DAP], Olivier Monteuius [BURST, DAP], Jean Luc Verdeil [PHIV, DAP])

Plants that grow several forms or type of leaves along a shoot, depending on age or shoot length, are called heteroblastic. The influence of heteroblastic on morphological and histocytological characteristics of Acacia mangium shoot apical meristems (SAMs) was assessed comparing materials with mature and juvenile leaf morphology in natural and in vitro conditions. For this we introduced a workflow for characterizing dome shape with few parameters (SAM dome heighth (H), basal diameter (D) and shape factor (S)) and their joint
statistical analysis to assess influence of conditions on SAM shape. In particular, a new statistical test is introduced here for multivariate analysis. This is a generalization of univariate ANOVA that takes into account statistical dependencies between the shape parameters. As a result, we found that SAM dome height (H) and basal diameter (D) were highly correlated. The joint analysis revealed that H, D, and shape (S) varied significantly according to the four plant origins investigated, with the higher scores for the outdoor mature source “Mat”. Overall, heteroblasty induced more conspicuous differences of SAM characteristics for the outdoor than for the in vitro materials.

5.2.3. Transport models

Participant: Michael Walker.

This research theme is supported by the ANR GeneShape and ERASysBio+ iSAM projects.

Active transport of the plant hormone auxin has been shown to play a key role in the initiation of organs at the shoot apex, and vein formation in both leaves and the shoot apical meristem. Polar localized membrane proteins of the PIN1 and AUX/LAX family facilitate this transport and observations and models suggest that the coherent organization of these proteins in the L1 layer is responsible for the creation of auxin maxima (surrounded by a depletion zone), which in turn triggers organ initiation close to the meristem center [39] [1]. Furthermore, canalized PIN allocations are thought to play a crucial role in vein formation in the leaf and in the L2.

Previous studies have typically modeled the L1 and L2 with different models to explain their different PIN allocations. For example [25] used the so-called ‘up-the-gradient’ model in the L1 to recreate the depletion zone around a maximum, and a ‘flux-mediated’ model to produce the midvein growing through the L2. A more unified approach was that of Stoma et.al. [10] who used the flux-mediated model in both tissues, but with different powers of the auxin flux, i.e. with a linear function in the L1 and a quadratic one in the L2.

We seek a completely unified model with no difference in the PIN allocation model. Our approach is based on inherent differences between the L1 and L2, specifically their dimensionality and the distribution of sources and sinks. We also use a flux-feedback function whose power is intermediate between one and two. For powers close to but slightly higher than one, the diffuse PIN distribution is retained in the L1 but a canalized one is formed in the L2.

We are also comparing two and three dimensional simulations to test the common approximation of three-dimensional tissues being modeled as two-dimensional. We find that source-sink models are well-approximated in two dimensions while sink- and source-driven systems show qualitatively different behavior.

A paper on this unified model is being written. This work is done in the context of the GeneShape and iSAM projects.

5.2.4. Mechanical model

Participants: Jérôme Chopard, Christophe Godin, Frédéric Boudon, Jan Traas, Olivier Hamant [ENS-Lyon], Arezki boudaoud [ENS-Lyon].

This research theme is supported by the ANR VirtualFlower and GeneShape projects.

The rigid cell walls that surround plant cells is responsible for their shape. These structures are under constraint due to turgor pressure inside the cell. To study the overall shape of a plant tissue and morphogenesis, its evolution through time, we therefore need a mechanical model of cells. We developed such a model, in which walls are characterized by their mechanical properties like the Young modulus which describes the elasticity of the material. Wall deformation results from forces due to turgor pressure. Growth results from an increase in cell wall synthesis when this deformation is too high. The final shape of the tissue integrates mechanically all the local deformations of each cell.
To model this process, we used a tensorial approach to describe both tissue deformation and stresses. Deformations were decomposed into elementary transformations that can be related to underlying biological processes. However, we showed that the observed deformations does not map directly local growth instructions given by genes and physiology in each cell. Instead, the growth is a two-stage process where genes are specifying by their activity a targeted shape for each cell (or small homogeneous region) and the final cell shape results from the confrontation between this specified shape and the physical constraints imposed by the cell neighbors. Hence the final shape of the tissue results from the integration of all these local rules and constraints at organ level. This work is being described in a paper which will be submitted for publication at the beginning of 2011.

5.2.5. Gene regulatory networks

Modeling gene activities within cells is of primary importance since cell identities correspond to stable combination of gene level activity.

- The auxin signaling pathway (Etienne Farcot, Yann Guédon, Christophe Godin, Yassin Refahi, Jonathan Legrand, Jan Traas, Teva Vernoux)

The auxin signalling network involves about 50 potentially interacting factors. We applied a graph clustering method [16] that relies on 0/1 interactions between factors deduced from yeast two-hybrid (Y2H) data. The Y2H analysis involves two independent tests (X-gal and HIS3 tests). Each possible interaction was tested in the two possible configurations, where each protein was alternatively the bait and the prey protein. A binary interaction is thus a summary of the four outputs of the X-gal and HIS3 tests. In order to limit the loss of information, we designed a standardization procedure to summarize the outputs of the X-gal and HIS3 tests as a distance defined on a continuous scale. This opens the possibility to studies the influence of phylogenetic distances between factors on their interactions using an extension of the mixture model for random graphs that incorporate explanatory variables. This work is the object of a collaboration with Jean-Benoist Leger and Stéphane Robin (MIA, AgroParisTech/INRA).

As an output to this interaction extraction, it was possible to apply a clustering procedure (work of Y. Guédon and J. Legrand), based on a mixture model of random graphs. This allowed us to determine a well-founded simplified network, which was then used to develop a differential equation model. This model showed that the meristem could present well differentiated buffering abilities in its role of auxin perception, a prediction that was corroborated by experiments using a newly developed auxin sensor. The combination of computational techniques and novel experimental tools and data that were developed in this project have led to a new view of the meristem behavior, where auxin signaling, and not only its transport, plays a crucial role. These results have appeared in a systems biology journal with wide readership [16].

As a follow up, extensions of the differential equation model are under study, with the aim of better understanding this system in more general contexts than the shoot apical meristem development.

- Complex dynamics and spatial interactions in gene networks (Yassin Refahi, Etienne Farcot, Christophe Godin)

Complex computational and mathematical questions arise in the study of gene networks at two levels: (i) the single cell level, due to complex, nonlinear interactions, (ii) the tissue level, where multiple cells interact through molecular signals and growth, so that even simple local rules can challenge our intuition at higher scales.

At the single cell level, new results were obtained in the framework of piecewise-linear models. Since their introduction in the late 1960’s, these models have been believed to present chaotic behavior in some parameter regimes. However, this was mostly observed numerically, based on intensive generation of random networks. In a collaboration between E. Farcot and R. Edwards (Univ. Victoria, Canada), with recent input from one of his students, E. Foxall, we have introduced a method to explicitly build piecewise affine models having a return map which is conjugate to a topological horseshoe. A paper presenting these results is currently in revision.
At a higher scale, we have also continued the study of gene regulation in meristematic tissues. Numerical tools for the simulation of gene network models have been further developed, in the context of Y. Refahi’s thesis. Using realistic 3D and 4D structures, reconstructed from microscopy data using the methods in [5], we have defined putative gene expression patterns in collaboration with biologist colleagues in Lyon (P. Das). We then have constructed an adimensional model with only 3 parameters, based on the existing literature [37]. Using standard optimization procedures to fit these parameters we were able to reproduce the in silico patterns with great accuracy. Moreover, alternative patterns were found, based on hypotheses of continuous deformation of gene patterns under tissue growth. These predictions remain to be tested, and this will constitute a large part of Y. Refahi’s post-doctoral project (INRA/INRIA CJS fund), in collaboration with ENS Lyon and the team of Henrik Jönsson in Lund, Sweden.

A related study has been initiated during a visit of E. Farcot at the Memorial University of Newfoundland, funded by the INRIA "Explorateur" program (see international relations section). This visit has allowed to initiate a mathematical study of gene regulated patterning in plants, with the aid of tools from equivariant dynamical systems theory, and more generally by relying on symmetry properties. Some first results have been obtained concerning homogeneous patterns in regular tissues, and their bifurcations.

5.2.6. Model integration

Participants: Mikael Lucas, Michael Walker, Jerôme Chopard, Frédéric Boudon, Christophe Godin, Laurent Laplaze, Jan Traas, François Parcy.

This research theme is supported by the ANR/BBSRC project iSam.

Our approach consists in building a programmable tissue which is able to accept different modeling components. This includes a central data structure representing the tissue in either 2-D or 3-D, which is able to grow in time, models of gene activity and regulation, models of signal exchange (physical and chemical) between cells and models of cell cycle (which includes cell division). For each modeling component, one or several approaches are investigated in depth, possibly at different temporal and spatial scales, using the data available from the partners (imaging, gene networks, and expression patterns). Approaches are compared and assessed on the same data. The objective of each submodel component will be to provide plugin components, corresponding to simplified versions of their models if necessary, that can be injected in the programmable tissue platform.

- Development of a computer platform for the 'programmable tissue'. (Jérôme Chopard, Michael Walker, Frédéric Boudon, Etienne Farcot, Christophe Godin)
  
  One key aspect of our approach is the development of a computer platform dedicated to programming virtual tissue development. This platform will be used to carry out integration of the different models developed in this research axis. The platform is based on OpenAlea. Partner models can be integrated in the platform in a non-intrusive way (the code of their model need not be rewritten). In this context, model integration will i) consist of designing adequate data-structures at different levels that will be exchanged and reused among the different plug-in models and ii) defining control flows at adequate levels to avoid the burden of excessive interaction between components.

- Design of a genetic model of inflorescence development. (Etienne Farcot, Christophe Godin, François Parcy)
  
  At a scale involving organs and their geometric arrangement, we have developed a first model of the control of floral initiation by genes, and in particular the situation of cauliflower mutants, in which the meristem fails in making a complete transition to the flower. This work couples models at different scales, since gene regulation is described by a minimal gene network, which is used as a decision module in an L-system model of the inflorescence architecture. This coupled model has led us to hypothesize some interactions between genes and a particular plant hormone, and experiments are currently being made to verify this prediction.
6. New Results

6.1. Contracts and sessions

Participants: Ugo Dal Lago, Ornella Dardha, Maurizio Gabbrielli, Elena Giachino, Claudio Guidi, Jacopo Mauro, Fabrizio Montesi.

Contracts are descriptions of the functionalities offered by a component or a service, and of the way these functionalities may be accessed by clients. A contract may include a description of the component capabilities, place constraints on their usage, as well as declare preferences, entitlements and credentials. When a client wants to use one of the functionalities offered, it engages a dialogue (e.g., a sequence of interactions) with the servers; this is usually called a session.

The expected dialogue in a session can be specified by means of types, the *session types*. We have studied [11] the integration of union and session types in a class-based language for building network applications, which amalgamates sessions and methods in order to flexibly exchange data according to communication protocols. We have established type safety properties guaranteeing that, after a session has started, computation cannot get stuck on a communication deadlock, and studied type inference. On a similar topic is the paper [23].

We have used types [32] to guarantee bounds on sessions. These are polynomial bounds, on both time and space needed by the interacting processes to carry out the interactions in their sessions. This is the first example of a refinement of session types guaranteeing quantitative properties beyond the usual safety property, and builds on earlier work on soft linear logic.

In service-oriented architectures, the mechanism that allows to manage sessions and, in particular, to assign incoming messages to the correct sessions, is critical for efficiency and performance. A well-known solution to this problem, first introduced by BPEL, makes use of *correlation sets*. Intuitively these distinguish different sessions by means of the values for some specific variables which are present also in messages, thus allowing for their routing to sessions on the basis of these values. We have studied [36] a typed language for programming services based on correlation sets, that takes into account key aspects of service-oriented systems, such as distribution, loose coupling, open-endedness and integration. We have provided an implementation of the language as an extension of the Jolie language and applied it to a nontrivial real-world example of a fully-functional distributed user authentication system.

In current SOC languages based on correlation sets, a message can be assigned to a unique session. In another line of work on correlation sets [35], we have studied the possibility – useful in many practical examples – of broadcasting messages to more than one session. We have investigated a data structure, based on radix trees, and an algorithm for managing a correlation mechanism, that support the broadcast primitive without degrading performances.

6.2. Fault Handling, compensations and transactions

Participants: Mila Dalla Preda, Maurizio Gabbrielli, Ivan Lanese, Jacopo Mauro, Gianluigi Zavattaro.

One of the predominant properties of CBUS is the loose coupling among the components. In fact, components can dynamically connect/disconnect and can be modified/updated at run time. It is thus important to support unexpected events, called faults.

In [30] we have studied the problem of fault handling in the kind of object-oriented languages developed in the EU Hats project; notably these languages have asynchronous method calls whose results are returned inside futures. We present an extension for those languages where futures are used to return fault notifications and to coordinate error recovery between the caller and callee. This can be exploited to ensure that invariants involving many objects are restored after faults.
Traditional fault handling mechanisms, including those based on try-catch operators, do not seem sufficient to deal with the non-local errors and failures of distributed systems. At the application level, more advanced transactional models and primitives are needed to guarantee integrity and continuity of the whole system. We study approaches based on long running transactions and compensations. A long running transaction is a computation that either successfully terminates, or it aborts. In case of abort, a compensation is executed to take the system to a consistent state. In [53], extending work started last year, we make a thorough comparison among different approaches to the specification of compensations, in particular static forms of recovery where the compensation is statically defined together with the transaction, and dynamic forms where the compensation is progressively built along with a computation.

We have also continued our study on faults and compensations in Service Oriented Computing. The approach to the interplay between bi-directional request-response interaction and faults, proposed in our past works on the Jolie language, supported the idea that the bi-directional pattern should not be interrupted in case of faults. However, this may cause long delays or even deadlocks if the communicating partner disappears. On the contrary, the approach of WS-BPEL causes no delay, but it does not allow to compensate the remote activity. We have investigated [38] an intermediate approach in which it is not necessary for the fault handler to wait for the response, but it is still possible on response arrival to gracefully close the conversation with the remote service.

A related work, but mainly developed in 2010, is [21].

### 6.3. Service orchestration and choreography

**Participants:** Mila Dalla Preda, Maurizio Gabbielli, Claudio Guidi, Ivan Lanese, Jacopo Mauro, Fabrizio Montesi, Marco Pistore, Gianluigi Zavattaro.

Orchestration has to do with the definition of services that should obey given behaviours. The services may be realised by composing services already available. Orchestration is often discussed in relationship with “choreography”, which refers to global descriptions of the intended behaviour of a system of components, stating the role of each participant and the set of coordination requirements.

In [55] we have studied the basic linguistic constructs and a reference implementation for aggregation, a mechanism for composing services that abstracts from the order of their communications. Aggregation is widely used in practice. However, since it is not natively supported by service-oriented languages, it is mostly implemented by means of ad-hoc solutions which typically exploit middleware technology.

A critical aspect for pervasive computing is the possibility to discover and use process knowledge at run time depending on the specific context. This can be achieved by using an underlying service-based application and exploiting its features in terms of dynamic service discovery, selection, and composition. Pervasive process fragments represent a service-based tool that allows to model incomplete and contextual knowledge. In [40] we provide a solution to automatically compose such fragments into complete processes, according to a specific context and specific goals. We compute the solution by encoding process knowledge, domain knowledge and goals into an AI planning problem.

Concerning choreography languages, two main approaches have been followed in their design: the interaction-oriented approach at the basis of WS-CDL and the process-oriented approach of BPEL4Chor. In [52] we investigate the relationships between the two approaches. In particular, we point out several possible interpretations for interaction-oriented choreographies: one synchronous and various asynchronous. Under each of these possible interpretations we characterize the class of interaction-oriented choreographies which have a direct process-oriented counterpart, and we formalize the corresponding notion of equivalence between the initial interaction-oriented choreography and the corresponding process-oriented counterpart.

In [50] we study the issue of checking a multiparty choreography against formal protocol specifications, and then projecting it onto a description of the individual service orchestrators. Contributions are also the definition of a multiparty choreography model, and the correctness proof for the projection.
6.4. Primitives for adaptable and evolvable components

**Participants:** Mario Bravetti, Ivan Lanese, Michael Lienhardt, Jacopo Mauro, Marco Pistore, Davide Sangiorgi, Gianluigi Zavattaro.

In Focus we study linguistic primitives for components, and models for them following the process calculus approach. A special emphasis is given to the adaptability and evolvability of the components — important issue in complex software systems. Components indeed are often used in contexts that had not been predicted at the time when the components were built. Moreover, the needs and the requirements on a system may change over time: one may find that the original specification was incomplete or ambiguous, or new needs may arise that had not been predicted at design time. As designing and deploying a system is costly, it is important that the system be capable of evolving and adapting itself to changes in the surrounding environment.

Models and linguistic constructs for adaptability and evolvability of components are studied in [34] and [19]. The key features of the component model in [34] are: a hierarchical structure of components; the capacity to move, update, wrap components; method interfaces for components; and capacities to isolate and distribute components. In the model in [19], adaptable processes have a location and can be subject to dynamic update actions at runtime (related to this paper is also [20]).

In [22] we provide an adaptation approach that can automatically adapt business processes to run-time context changes that impede achievement of a business goal. We define a formal framework that adopts planning techniques to automatically derive necessary adaptation activities on demand. The adaptation consists in identifying recovery activities that guarantee that the execution of a business process can be successfully resumed and, as a consequence, the business goals are achieved. The solution proposed is evaluated on a real-world scenario from the logistics domain.

Adaptability and evolvability are major concerns in Software Product Lines. The EU Hats project has developed the idea of delta-oriented programming (DOP) as a technique for implementing Software Product Lines based on modifications (add, remove, modify) to a core object-oriented program. Such modifications can introduce errors into a program, when type signatures of classes are modified. To overcome this problem we have introduced [54] a type system for delta-oriented programs. The system is based on row polymorphism, a well-known method in type systems for records.

6.5. Resource Control

**Participants:** Ugo Dal Lago, Marco Gaboardi, Daniel Hirschkoff, Simone Martini, Paolo Parisen Toldin, Giulio Pellitta, Davide Sangiorgi.

In Focus, we study both foundations and methodologies for controlling the amount of resources programs and processes make use of. The employed techniques mainly come from the world of type theory and proof theory, and as such have been used extensively in the context of sequential computation. Interesting results have been obtained recently indicating that those techniques can be quite useful in the concurrent context too, thus being potentially interesting for CBUS.

We have continued our work on techniques for ensuring termination of programs. On the one hand we have refined [25] previous techniques, enhancing them by taking into account input/output capabilities of channels and subtyping. On the other hand we have studied [28] how to transport techniques initially devised for processes onto sequential higher-order languages with imperative features (e.g., λ-calculi with references). The method employed makes it possible to combine term rewriting measure-based techniques for termination of imperative languages with traditional approaches to termination in purely functional languages, such as logical relations.

In [31], a type system of linear and dependent types, called dIPCF, has been proved to be a sound, but also relatively complete, way to prove intensional (but also extensional) properties of PCF programs. In other words, not only all properties of programs proved by typing are operationally valid, but all true properties of programs can be proved so by way of dIPCF. This holds not only for terms of base type, but also for (first-order) functions: this makes the type system more expressive than intersection type disciplines.
A characterization of probabilistic polynomial time classes by way of linear typing systems for a variant of Godel’s T called RSLR, has been proposed \[ 51 \]. Classes like BPP and PP are characterized by RSLR once appropriate constraints on the probability of error are imposed.

A unifying methodology for the study of resource consumption of processes has been presented in \[ 13 \]: it is a refinement of realizability, in which not only termination but also concrete resource bounds can be obtained by showing a function to be realized by a program.

On the same topic is the paper \[ 26 \], that polishes previous work.

### 6.6. Verification of extensional properties

**Participants:** Elena Giachino, Cosimo Laneve, Tudor Alexandru Lascu, Davide Sangiorgi, Gianluigi Zavattaro.

Extensional refers to properties that have to do with behavioral descriptions of a system (i.e., how a system looks like from the outside). Examples of such properties include classical functional correctness and deadlock freedom. We mainly employ techniques based on behavioral equivalences (and preorders), and on types and logics. Type systems offer a good trade-off between expressiveness and efficiency of the techniques. A substantial amount of the work carried out this year has to do with the transfer of techniques between the areas of concurrency theory and object-oriented languages.

We have developed \[ 29 \] a technique for the deadlock analysis of systems of concurrent object groups. The technique makes use of types in the form of contracts, that is, abstract descriptions of method’s behaviours. Object groups are collections of objects that perform collective work. Within a group, there can be only one running thread at a time; the scheduling of threads is cooperative.

We have studied \[ 37 \] the concept of ownership types, originally introduced for (sequential) object-oriented languages, in the setting of pure message-passing concurrency. Ownership types have the effect of statically preventing certain communications, and can block the accidental or malicious leakage of secrets. Intuitively, a channel defines a boundary and forbids access to its inside from outer channels, thus preserving the secrecy of the inner names from malicious outsiders.

In a different line of work, we have analyzed ad hoc networks, intended as networks of devices connected by wireless links and communicating via broadcast. We have considered \[ 27 \], \[ 18 \] models in which the communication topology of a network is represented as a graph. Nodes represent states of individual processes, and adjacent nodes represent single-hop neighbors. Processes are finite state automata that communicate via selective broadcast messages. Reception of a broadcast is restricted to single-hop neighbors. In these systems we have studied various forms of reachability (example: the existence of an initial topology in which the execution of the protocol can lead to a configuration with at least one node in a certain state).

Induction is a pervasive tool in Computer Science and Mathematics for defining structures and reasoning on them. Coinduction is the dual of induction, and as such it brings in tools that are quite different from those provided by induction. The best known instance of coinduction is bisimulation, mainly employed to define and prove equalities among potentially infinite objects: processes, streams, non-well-founded sets, and so on. Sangiorgi has completed \[ 47 \], \[ 49 \] two comprehensive textbooks on bisimulation and coinduction (in \[ 49 \], Sangiorgi is an editor, and author of two chapter contributions \[ 48 \], \[ 46 \]). The books explain the fundamental concepts and techniques, and the duality with induction. A special emphasis is put on bisimulation as a behavioural equivalence for processes. Thus the books also serve as an introduction to models for expressing processes, and to the associated techniques of operational and algebraic analysis.

### 6.7. Tutorial papers on Service-Oriented Computing

**Participants:** Mario Bravetti, Ivan Lanese, Davide Sangiorgi, Gianluigi Zavattaro.

We have contributed to a few tutorial papers that summarise the work on languages and tools for Service-Oriented Computing that has been carried out within the EU project Sensoria in Focus and elsewhere. The papers appear as chapters of a book dedicated to the topic.
The chapters [45] and [43] present and contrast the primitives and the behavioural theories of the main process calculi designed for modeling services.

Languages and models for service-oriented applications usually include primitives and constructs for exception and compensation handling. Exception handling is used to react to unexpected events while compensation handling is used to undo previously completed activities. In [44] we investigate the impact of exception and compensation handling in service-oriented process calculi.

The chapter [42] deals with contracts: descriptions intended to provide support for the automatic on-demand discovery of functionalities offered by a service. The approach followed is to describe such contracts as process calculi expressions. We show how, in certain cases, service contracts can be automatically extracted out of service behaviour, and how they can be used to formally check compliance among the communication protocols followed by the interacting services.

Finally, in [41] we present different tools that have been developed for verifying properties of service implementations with respect to their formal specifications in an automated, or semi-automated, way.

### 6.8. Expressiveness of computational models

**Participants:** Mila Dalla Preda, Ugo Dal Lago, Ivan Lanese, Cosimo Laneve, Davide Sangiorgi.

Expressiveness refers to the study of the expressive power of computational models. In 2011 we have addressed four main aspects.

First, we have continued our investigation of reversible computations. Reversibility is a main ingredient in the study of programming abstractions for reliable systems, e.g. for exception handling. In fact, reversibility can be used for going back to some consistent state after an exception has occurred. In previous years we had defined $\rho\pi$, a higher-order calculus where processes can both go forward and backward in the computation. This year we have studied fine-grained rollback primitives to control reversibility. The definition of a proper semantics for such a primitive is a surprisingly delicate matter because of the potential interferences between concurrent rollbacks. We have also considered lower-level distributed semantics, which are closer to an actual implementation of the rollback primitives, and their relationship with the high-level semantics.

A thread of research close to that of $\rho\pi$ is the study of the properties and the expressive power of a simple calculus with reversible transitions, called reversible structures. In [24], we have demonstrated a standardization theorem for these structures. When terms in reversible structures have unique id, the standardization theorem may be strengthened in a form that bears a quadratic algorithm for reachability, a problem that is EXPSPACE-complete for generic structures (as in Petri Nets). The expressive power of reversible structures has been studied in [17], [12] and a compilation of asynchronous Reversible CCS has been provided.

A second aspect has been motivated by the analogy between malicious software and biological infections [39]. In the paper, we have used a formalism originally developed for the analysis of biological systems — the kappa calculus by Danos and Laneve — for the formalization and analysis of malicious software. In particular we have modeled the different actors involved in a malicious code attack in the kappa-calculus. Then, by simulating the behavior, we have shown how to extract relevant information that can drive the choice of the defense technique to apply.

A third aspect has been the refinement [14] of some previous work on the expressiveness and decidability of higher-order concurrent languages. — formalisms for concurrency in which processes can be passed around in communications.

A fourth aspect has been the study of properties of a simple calculus for quantum computation. In [16], we have demonstrated a confluence property both for finite and infinite computations using a novel technique.
6. New Results

6.1. Security

Participants: Ilaria Castellani, Zhengqin Luo, Tamara Rezk [correspondant], José Santos, Manuel Serrano.

6.1.1. Secure session calculi

We have pursued our work on controlling information flow in session calculi, started in previous years in collaboration with colleagues from the university of Torino. We also started investigating a notion of (objective) reputation for principals participating in sessions. The reputation of a principal is based on her previous behaviour as a user of a service. A principal's reputation can be checked both by the service itself, before admitting again the principal as a user, or by other principals to evaluate the reputation of the current users before they join a service (we consider multi-user services). We plan to apply this idea to refine our previous work on information flow control in multiparty sessions, by considering reputations built on the "security behaviour" of principals.

In the work “Information flow safety in multiparty sessions” [11], we consider a calculus for multiparty sessions enriched with security levels for messages. We propose a monitored semantics for this calculus, which blocks the execution of processes as soon as they attempt to leak information. We illustrate the use of our monitored semantics with various examples, and show that the induced safety property implies the security property studied previously for the same calculus. This work was presented at the 18th International Workshop on Expressiveness in Concurrency (EXPRESS’11).

In the work “A Reputation System for Multirole Sessions” [10], we extend role-based multiparty sessions with reputations and policies associated with principals. The reputation associated with a principal in a service is built by collecting her relevant behaviour as a participant in sessions of the service. The service checks the reputation of principals before allowing them to take part in a session, and decides whether to accept them or not depending on their reputation and on the role they want to play. Furthermore, principals can declare policies that must be fulfilled by the other participants of the same service. These policies are used by principals to check the reputation of the current participants and to decide accordingly whether or not to join the service. Our approach is illustrated by an example describing a real-world protocol. This work was presented at the 6th International Symposium on Trustworthy Global Computing (TGC’11).

Both [11] and [10] were partially funded by the ANR-08-EMER-010 grant PARTOUT.

6.1.2. Automatic Code Injection Prevention for Web Applications

We propose a new technique based on multitier compilation for preventing code injection in web applications. It consists in adding an extra stage to the client code generator which compares the dynamically generated code with the specification obtained from the syntax of the source program. No intervention from the programmer is needed. No plugin or modification of the web browser is required. The soundness and validity of the approach are proved formally by showing that the client compiler can be fully abstract. The practical interest of the approach is proved by showing the actual implementation in the Hop environment.

This work was presented in TOSCA ’11 and appeared in the LNCS series [13]. See also software section.

6.1.3. A Certified Lightweight Non-Interference Java Bytecode Verifier

We propose a type system to verify the non-interference property in the Java Virtual Machine. We verify the system in the Coq theorem prover.

This work will appear in the journal of Mathematical Structures in Computer Science [8].
6.1.4. Information-flow types for homomorphic encryptions

We develop a flexible information-flow type system for a range of encryption primitives, precisely reflecting their diverse functional and security features. Our rules enable encryption, blinding, homomorphic computation, and decryption, with selective key re-use for different types of payloads. We show that, under standard cryptographic assumptions, any well-typed probabilistic program using encryptions is secure (that is, computationally non-interferent) against active adversaries, both for confidentiality and integrity. We illustrate our approach using ElGamal and Paillier encryption.

We present two applications of cryptographic verification by typing: (1) private search on data streams; and (2) the bootstrapping part of Gentry’s fully homomorphic encryption.

We provide a prototype typechecker for our system.

This work appeared in CCS’11 [12]. See also software section.

6.1.5. The Mashic compiler

Mashups are a prevailing kind of web applications integrating external gadget APIs often written in the Javascript programming language. Writing secure mashups is a challenging task due to the heterogeneity of existing gadget APIs, the privileges granted to gadgets during mashup executions, and Javascript’s highly dynamic environment.

We propose a new compiler, called Mashic, for the automatic generation of secure Javascript-based mashups from existing mashup code. The Mashic compiler can effortlessly be applied to existing mashups based on a wide-range of gadget APIs. It offers security and correctness guarantees. Security is achieved by using the Same Origin Policy. Correctness is ensured in the presence of benign gadgets, that satisfy confidentiality and integrity constrains with regard to the integrator code. The compiler has been successfully applied to real world mashups based on Google maps, Bing maps, YouTube, and Zwibbler APIs.

See also software section.

6.1.6. Secure Information Flow by Self-Composition

Information flow policies are confidentiality policies that control information leakage through program execution. A common means to enforce secure information flow is through information flow type systems. Although type systems are compositional and usually enjoy decidable type checking or inference, their extensibility is very poor: type systems need to be redefined and proven sound for each new single variation of security policy and programming language for which secure information flow verification is desired. In contrast, program logics offer a general mechanism to enforce a variety of safety policies, and for this reason are favored in Proof Carrying Code, a promising security architecture for mobile code. However, the encoding of information flow policies in program logics is not straightforward, because they refer to a relation between two program executions. The purpose of this work is to investigate logical formulations of secure information flow based on the idea of self-composition, that reduces the problem of secure information flow of a program P to a safety property for program P composed with itself.

This work appeared in the special issue of MSCS of PLID [7].

6.1.7. Secure Information flow enforcement techniques for dynamic security policies

We performed a comprehensive investigation of alternative static mechanisms to enforce information flow policies considering a setting in which programs run under an authority that is only known at runtime and that yields a relaxation of the base security policy. The devised method aims at eliminating the need to reanalyse a program each time its authority changes. The soundness of the proposed approach was established for a concurrent higher-order imperative lambda calculus with reference creation. This work resulted in the report “Typing Illegal Information Flows as Program Effects” available at http://www-sop.inria.fr/members/Jose.Santos/reportInfFlow.pdf.
6.2. Models, semantics, and languages

Participants: Pejman Attar, Gérard Berry [correspondant], Gérard Boudol, Frédéric Boussinot, Johan Grande, Cyprien Nicolas, Manuel Serrano.

6.2.1. HipHop

HipHop is a new Domain Specific Language for HOp dedicated to request and event orchestration. HipHop follows the synchronous reactive model of the Esterel and ReactiveC languages, originally developed for embedded systems programming. It is based on synchronous concurrency and preemption primitives, which are known to be key components for the modular design of complex temporal behaviors. Although the language is concurrent, the generated code is purely sequential and thread-free. HipHop is translated to Hop code to be interpreted by the Hop runtime, either on server or client sides. HipHop has been described in a paper [9] accepted at the new International Workshop on Programming Language And Systems Technologies for Internet Clients (Plastic 2011).

6.2.2. Tesard

Tesard is a programming language of the Caml family, designed to offer simple constructs for shared memory concurrency and a deadlock-free semantics.

It features in particular the 2 following constructs:

- `thread e` which launches a thread that will execute expression `e`, and returns immediately;
- `lock x in e` which takes a lock on mutable value `x` (possibly waiting to be able to do so), executes expression `e`, releases the lock on `x` and returns the result of the execution of `e`.

A type and effect system is used at compile-time to:

- associate a mutex with every mutable value
- make the `lock ... in ...` construct implement deadlock avoidance.

The language is implemented as an interpreter and a bytecode compiler. It is a fork of Llama Light, which is itself derived from Caml. Llama Light is functionnaly roughly equivalent to Caml Light, but a large part of its code comes from OCaml, along with the threads library and runtime that we ported ourselves.

Several key parts of the language have been implemented: the runtime and most of the type and effect system, including inference of the creation of mutexes (`let region r in ...`) but not region polymorphism yet. The project is not in a usable state (no release has been made yet). Development versions are publicly available via GitHub at https://github.com/nahoj/llama.

6.2.3. Synchronous orchestration and beyond

We studied DSL, an orchestration language based on the synchronous/reactive model. In DSL, systems are composed of several sites executed asynchronously. Within each site, scripts are run in a synchronous parallel way. Scripts may call functions that are treated in an abstract way: their effect on the memory is not considered, but only their “orchestration”, *i.e.*, the organisation of their calls in time and in place (the site where they are called). The mapping of sites onto cores allows one to benefit from multicore architectures. Two properties are required by DSL: reactivity of sites and absence of interferences between scripts run by distinct sites. We also introduced DSLM, which adds a memory level to DSL and a way to automatically adapt the execution to get a maximal use of the available cores. This work, presented respectively in [18] and [15], was funded by the ANR-08-EMER-010 grant PARTOUT.

6.3. Web programming

Participants: Zhengqin Luo, Cyprien Nicolas, Tamara Rezk, Bernard Serpette, Manuel Serrano [correspondant].
6.3.1. A new evaluator for Hop

At the time where Hop programs were basic scripts, the performance of the server-side interpreter was not a concern. An inefficient interpreter was acceptable. As Hop expanded, Hop programs got larger and more complex. A more efficient interpreter was necessary. Therefore, this year we have designed and implemented a new interpreter for Hop. It is compact, its whole implementation counting no more than 2.5 KLOC. It is more than twice faster than the old interpreter and consumes less than a third of its memory. The architecture and the performance of the new interpreter are described in [14].

6.3.2. Abstraction of Hop with a bicolored lambda-calculus

We have studied an extension of the $\lambda$-calculus where each expression has a color. These colors designate the sites where expressions are evaluated, i.e., in the server or in the client. Colors are similar to the $\$\$ and ~ annotations of Hop. With this, we have defined a transformation, using $\beta$-expansion, which groups together expressions with the same color. Correction, confluence and termination of the transformation was verified using the Coq system and its description was described in paper to appear in 2012 [16].

Following Hop’s syntax, $\$\$ mentions that the followed expression is evaluated on the server while the ~ character introduce client code. Inside the client code it is allowed to reintroduce some server expressions by reusing a $\$\$. We can imagine, for example, that the action associated to a client’s button is dependent on some server’s data (order number, proxies, databases, ...). This kind of example is depicted with the abstract syntax tree in the upper left part of figure 1.

![Figure 1. Example of Hop colored tree](image)

From the server’s point of view, the client’s code following a ~ is ignored, but the environment in which this client code is activated must be preserved for all $\$\$ inside this code. Therefore, there exists a relation between a node ~ and its including $\$\$. These relations are depicted by the dashed arrows in the upper right tree of Figure 1. The proposed transformation highlights this relation and is shown in the bottom of the figure where it can be observed that the transformation groups together expressions of the same color.

6.3.3. HopTeX, an Hop application for authoring documents

HOPTeX is a Hop application for authoring HTML and LaTeX documents. The content of the document is either expressed in HTML or in a blending of HTML and a dedicated wiki syntax, for the sake of conciseness and readability. The rendering of the document is expressed by a set of CSS rules. The main originality of HOPTeX is to consider LaTeX as a new media type for HTML and to express the compilation from HTML to LaTeX by the means of dedicated style sheet rules. HOPTeX can then be used to generate high quality documents for both paper printed version and electronic version.
HOPTeX is implemented in HOP, a multi-tier programming language for the Web 2.0. This implementation extensively relies on two facilities generally only available on the client-side that HOP also supports on the server-side of the application: DOM manipulations and CSS server-side resolutions.

The online version of the paper describing HOPTeX [17] is available at the HOPTeX web page (http://hop.inria.fr/hop/weblets/homepage?weblet=hoptex). Contrary to the PDF version published in the proceeding of the workshop, the online version is convenient for both desktop computers and Smartphones.
5. New Results

5.1. IP networks

Participants: Eitan Altman, Konstantin Avrachenkov.

5.1.1. Interdisciplinary study of the Internet access and of network neutrality

In our previous research we have identified large inefficiencies that occur when one allows one type of provider (e.g. access provider) to impose costs on another type of provider (e.g. content provider). This part in which E. Altman collaborated with P. Bernhard (INRIA project-team BIOCORE), S. Caron and G. Kesidis (both from Pennsylvania State Univ., USA), J. Rojas-Mora (Univ. of Barcelona, Spain), and S. Wong (Univ. of A Coruña, Spain) has now appeared in [ 96 ].

This investigation has been pursued in various directions. In [ 42 ], E. Altman, A. Legout (INRIA project-team PLANETE) and Y. Xu (Univ. Avignon/LIA) have studied a hierarchical structure of ISPs, the economic impact of some caching placement policies, and more complex demand functions (the demands of users for content). In [ 33 ], E. Altman and the law specialist S. Wong (Univ. of A Coruña, Spain) analyze in cooperation with the economist J. Rojas-Mora (Univ. of Barcelona, Spain) the impact of legislation related to network neutrality on the quality of service for the end users.

5.1.2. Adaptive monitoring system for IP networks

The remarkable growth of the Internet infrastructure and the increasing heterogeneity of applications and users’ behavior make more complex the manageability and monitoring of ISP networks and raises the cost of any new deployment. The main consequence of this trend is an inherent disagreement between existing monitoring solutions and the increasing needs of management applications. In this context, in [ 62 ] K. Avrachenkov, I. Lassoued, A. Krifa and C. Barakat (all three from INRIA project-team PLANETE) present the design of an adaptive centralized architecture that provides visibility over the entire network through a network-wide cognitive monitoring system. Practically, given a measurement task and a constraint on the volume of collected information, the proposed architecture drives the sampling rates on the interface of network routers to achieve the maximum possible accuracy, while adapting itself to any change in network traffic conditions. The authors tune the system parameters with the help of FAST sensitivity test.

5.1.3. Size based scheduling

Size-based scheduling is a promising solution to improve the response time of small flows (mice) that have to share bandwidth with large flows (elephants). To do this, one important task is to track the size of the ongoing flows at the router. However, most of the proposed size-based schedulers either employ the trivial way of tracking the size information of all flows, or require changes at end-hosts. Hence, either they are not scalable or they increase complexity. In [ 55 ], E. Altman, D. Mon Divakaran (IIT Mandi, India) and P. Vicat-Blanc Primet (Lyatiss, France) have proposed a new way of performing size-based scheduling in a practical and scalable fashion, by identifying and ‘de-prioritizing’ elephants only at times of high load. They exploit TCP’s behavior by using a mechanism that detects a window of packets - called spikes - when the buffer length exceeds a certain threshold. This spike-detection is used to identify elephant flows and thereafter de-prioritize them. Two-level processor-sharing (TLPS) scheduling is employed to schedule flows in two queues, one with the high-priority flows, and the other with the de-prioritized flows. They show that the proposed mechanism not only improves the response time of mice flows in a scalable way, but also gives better response times to other flows by treating them preferentially as long as they do not overload the high-priority queue.
5.1.4. Accuracy of fluid models for bandwidth-sharing networks

Optimal control of stochastic bandwidth-sharing networks is typically difficult. In order to facilitate the analysis, deterministic analogues of stochastic bandwidth-sharing networks, the so-called fluid models, are often chosen for analysis, as their optimal control can be found more easily. The tracking policy translates the fluid optimal control policy back into a control policy for the stochastic model, so that the fluid optimality can be achieved asymptotically when the stochastic model is scaled properly. In [20] K. Avrachenkov, A. Piunovsky and Y. Zhang (both from the University of Liverpool, UK) study the efficiency of the tracking policy, that is, how fast the fluid optimality can be achieved in the stochastic model with respect to the scaling parameter. In particular, the result of [20] shows that, under certain conditions, the tracking policy can be as efficient as feedback policies.

5.1.5. Bootstrap method for simulating bandwidth sharing

In [71], E. Altman, T. Jimenez and J. Rojas-Mora (both from Univ. Avignon/LIA) identify difficulties in evaluating through simulations the expected transfer time of a file when several TCP connections share a common bottleneck buffer. The main difficulties are due to the fact that the file size over the Internet has been reported to have a Pareto distribution with parameter smaller than 1.5. This implies that the number of ongoing connections as well as the sojourn times have infinite variance. This has two implications: one cannot estimate the confidence intervals for simulation based on the CLT (central Limit Theory approach), and the duration of the simulations needed to get to steady state is very long. The authors show how to solve both problems by the use of the bootstrap approach.

5.2. Wireless communications

Participants: Sara Alouf, Eitan Altman, Konstantin Avrachenkov, Nicaise Choungmo Fofack, Mahmoud El Chamie, Majed Haddad, Manjesh Kumar Hanawal, Philippe Nain, Giovanni Neglia, Manoj Panda, Sreenath Ramanath.

5.2.1. Green networking

Green networking is a new trend in network design that is more aware of the impact of technology on the environment and on humans. Reducing energy has been so far the main concern in that approach, and much of the research has been devoted to understanding the tradeoffs between reducing energy and other performance measures such as coverage and delay.

For several years we have been contributing to this research effort, many of which have been summarized in the survey [67] by E. Altman in collaboration with G. S. Paschos (Center for Research and Technology, Hellas, Greece), P. Mannersalo (VTT, Finland), S. Stanczak (HH-Fraunhofer, Berlin, Germany) and L. Tassiulas (Univ. of Thessaly, Greece). In particular, much of the work involving members of MAESTRO that had appeared in previous years in conferences concerning energy saving in WiMax has now appeared in a journal publication [21] by A. P. Azad, S. Alouf, E. Altman, in cooperation with V. Borkar (TATA Inst. of Fundamental Research, Mumbai, India) and G. S. Paschos (Center for Research and Technology, Hellas, Greece).

In 2011 we started investigating policies for switching off base stations using two new tools: multimodularity and stochastic geometry. The latter has been used by E. Altman and M. K. Hanawal (also with Univ. Avignon/LIA), in cooperation with R. El-Azouzi (Univ. Avignon/LIA) and S. Shamai (Technion, Israel) in [98] to study the tradeoffs related to the uplink, and by E. Altman in cooperation with C. Hasan and J. M. Gorce (both from INSA-Lyon and INRIA project-team SWING) in [89] for the downlink. Optimal policies were obtained within the class of policies that switch off base stations with some fixed probability but independently of each other. To relax this restriction of independence and thus obtain even better policies, S. Ramanath and E. Altman, in collaboration with V. Kavitha (Univ. Avignon/LIA), have used in [69] the theory of multimodularity, which is the discrete counterpart of convexity. Among the most striking points in this research has been the observation in [98] that the conventional energy saving approach can have the opposite effect on the humans in the uplink: when the base station closest to a mobile phone is switched off (for energy...
saving) then the mobile phone has to transmit with a larger power so as to reach a more remote base station. It turns out that the main source of radiation to the human brain is indeed the uplink transmission, which implies that switching off base stations could cause more exposure to radiation. This is of particular concern in view of the announcement by the World Health Organization (May 31, 2011) that cell phones cause cancer.

5.2.2. Cellular networks with continuous connectivity

In [65], S. Alouf and V. Mancuso (Institute IMDEA Networks, Madrid, Spain) analyze the power save and its impact on web traffic performance when customers adopt the continuous connectivity paradigm. Considering realistic http traffic, they evaluate the user access delay, the download time and the expected economy of energy in the cell. The model, validated through packet-level simulations, shows that dramatic energy save can be achieved by both mobile users and base stations. In case of Poisson arrivals, the aggregate behavior of a base station’s users is studied by means of a processor-shared queueing system [105]. The model can be used to maximize the base station energy savings under a given set of QoS performance constraints. With the participation of N. Choungmo Fofack, the work in [65] has been complemented with a sensitivity analysis [95]. The impact of model parameters on the performance and cost metrics is thoroughly assessed.

5.2.3. Power allocation in multicell networks

Power allocation to satisfy user demands in the presence of large number of interferers in a multicellular network is a challenging task. Further, power to be allocated depends upon the system architecture, for example, upon components like coding, modulation, transmit precoder, rate allocation algorithms, available knowledge of the interfering channels, etc. This calls for an algorithm via which each base station in the network can simultaneously allocate power to their respective users so as to meet their demands (when they are within the achievable limits), using whatever information regarding the other users is available. In [70], S. Ramanath, V. Kavitha (Univ. Avignon/LIA) and M. Debbah (SUPELEC) devise such an algorithm which is in fact universal: the proposed algorithm works from a fully cooperative setting to almost no cooperation and or for any configuration of modulation, rate allocation, etc. schemes. The algorithm asymptotically satisfies the user demands, running simultaneously and independently within a given total power budget at each base station. Further, it requires minimal information to achieve this: every base station needs to know its own users demands, its total power constraint and the transmission rates allocated to its users in every time slot. The authors formulate the power allocation problem in a system specific game theoretic setting, define system specific capacity region and analyze the proposed algorithm using ordinary differential equation (ODE) framework. Simulations confirm the effectiveness of the proposed algorithm.

5.2.4. Small cell networks

In [28], S. Ramanath and E. Altman, in collaboration with V. Kavitha (Univ. Avignon/LIA), characterize the performance of Picocell networks in the presence of moving users. They model various traffic types between base-stations and mobiles as different types of queues. They derive explicit expressions for the expected waiting time, service time and drop/block probabilities for both fixed and random velocity of mobiles. They obtain (approximate) closed-form expressions for optimal cell size when the velocity variations of the mobiles is small for both non-elastic and elastic traffic. They conclude from the study that, if the expected call duration is long enough, the optimal cell size depends mainly on the velocity profile of the mobiles, its mean and variance. It is independent of the traffic type or duration of the calls. Further, for any fixed power of transmission, there exists a maximum velocity beyond which successful communication is not possible. This maximum possible velocity increases with the power of transmission. Also, for any given power, the optimal cell size increases when either the mean or the variance of the mobile velocity increases.

5.2.5. New concepts in fair resource allocation

Fair resource allocation is usually studied in a static context, in which a fixed amount of resources is to be shared. In dynamic resource allocation one usually tries to assign resources instantaneously so that the average share of each user is split fairly. The exact definition of the average share may depend on the application, as different applications may require averaging over different time periods or time scales. Our main contribution is to introduce new refined definitions of fairness that take into account the time over which one averages the
performance measures. In [39] E. Altman, K. Avrachenkov and S. Ramanath examine how the constraints on the averaging durations impact the amount of resources that each user gets. The authors apply this new concept in [68] to spectrum allocation and indoor-outdoor femtocells.

5.2.6. Self organization in cellular networks

Time-slots and frequencies are contended in cellular networks and their allocation is determined by base stations. For scalability purposes the resource allocation is decentralized, so that base stations do not share their information with each other. Actions are often taken based on partial information on the system. In particular, the statistics of the channels are often not available. Scheduling decisions of a base station concerning mobiles in its cell cause interference in other cells and there is thus a need to dynamically adjust to interference and to converge to a satisfactory operation point. This has motivated a large amount of work on self-organization in cellular networks based on OFDMA. E. Altman, Z. Altman, R. Combes (both from Orange Labs, Issy les Moulineaux) and M. Haddad have written a series of papers on self-organization. In [53] and [51] self-organization in interference coordination is studied. In [50], R. Combes, Z. Altman and E. Altman, further propose and analyze a self-optimization method for coverage-capacity optimization in OFDMA networks with MIMO. Moreover, they study in [52] self-organization when adding relays so as to increase coverage. Static and dynamic resource sharing mechanisms are investigated. In the static case they use a queuing model to calculate the optimal resource sharing strategy and the maximal capacity of the network analytically. The influence of relay planning and number of deployed relays is investigated, and gains resulting from good planning are evaluated analytically. Self-optimizing dynamic resource allocation is tackled using a Markov Decision Process (MDP) model.


Self-organization has also been used in the past to obtain opportunistic scheduling in a way that achieves proportional fair resource sharing. In [23], R. Combes, Z. Altman (both from Orange Labs, Issy les Moulineaux) and E. Altman, have extended this to the general \( \alpha \)-fair concept. A dynamic choice of the factor \( \alpha \) is proposed, which has the interpretation of trading optimality with fairness in a dynamic way.

5.2.7. Dynamic networks

In source routing, a complete path is chosen for a packet to travel from source to destination. While computing the time to traverse such a path may be straightforward in a fixed, static graph, doing so becomes much more challenging in dynamic graphs, in which the state of an edge in one timeslot (i.e., its presence or absence) is random, and may depend on its state in the previous time step. The traversal time is due to both time spent waiting for edges to appear and time spent crossing them once they become available. In [99], P. Nain in collaboration with A. Bar-Noy (City University of New York), P. Basu (Raytheon BBN Technologies), M. P. Johnson (Pennsylvania State University), F. Yu (City University of New York) and D. Towsley (University of Massachusetts at Amherst) computes the expected traversal time (ETT) for a routing path in a number of special cases of stochastic edge dynamics models, and for three edge failure models, culminating in a surprisingly challenging yet realistic setting in which the initial configuration of edge states for the entire path is known. We show that the ETT for this “initial configuration” setting can be computed in quadratic time, by an algorithm based on probability generating functions. The authors also give several linear-time upper and lower bounds on the ETT.

5.2.8. Sensor networks

In many application scenarios sensors need to calculate the average of some local values, e.g. of local measurements. A possible solution is to rely on consensus algorithms. In this case each sensor maintains a local estimate of the global average, and keeps improving it by performing a weighted sum of the estimates of all its neighbors. The number of iterations needed to reach an accurate estimate depends on the weights used at each sensor. K. Avrachenkov, G. Neglia and M. El Chamie have proposed a new average consensus algorithm, where each sensor selects its own weights on the basis of some local information about its neighborhood [45]. In realistic sensor network topologies, the algorithm shows faster convergence than other existing consensus protocols.
5.2.9. Delay and disruption-tolerant networks (DTNs)

5.2.9.1. Applying risk sensitive control to delay tolerant networks

When controlling the propagation of a message in DTNs, the objective is often to maximize the successful delivery probability of a message within a given deadline. It takes often the form of the expectation of the exponent of some integral cost. So far, models involving such costs have been solved by interchanging the order of expectation and the exponential function. While reducing the problem to a standard optimal control problem, this interchange is only tight in the mean-field limit obtained as the population tends to infinity. In [41] E. Altman, V. Kavitha (Univ. Avignon/LIA), F. De Pellegrini (Create-Net, Trento, Italy), V. Kamble (UC Berkeley, CA, USA) and V. Borkar (TATA Inst., Mumbai, India), identify a general framework from optimal control in finance, known as risk sensitive control, which allows handling the original (multiplicative) cost and obtaining solutions to several novel control problems in DTNs. New optimal control problems which consider the effect of wireless propagation path loss factor and the power constraints at the source and or the destination are proposed for DTNs within this framework. Optimal policies of non-threshold type are found.

5.2.9.2. Multiple destinations

In [73], C. Singh, A. Kumar and R. Sundaresan (all three from IISc Bangalore, India) in collaboration with E. Altman, use Markov Decision Processes to study optimal policies for propagation of contents in DTNs to multiple destinations. They obtain structural properties for a discretized system which allows them to derive the structure of optimal policies to the original problem.

5.2.9.3. Reliable unicast and multicast

In case the DTN does not deliver a packet within some time $T$, it has to be retransmitted. In [36] E. Altman and M. Panda, in collaboration with T. Chahed and A. Ali, (both from Telecom SudParis) and L. Sassatelli (Univ. Nice Sophia Antipolis/I3S), propose protocols for unicast and for multicast that render the connection reliable. These protocols include ACKs and retransmissions. The authors compute the value of $T$ that optimizes the throughput and address energy consumptions aspects.

5.2.9.4. Network coding

In [18], E. Altman studies, in cooperation with F. De Pellegrini (Create-Net, Trento, Italy), how to improve the performance of DTNs by adding network coding. The latter has the effect of efficiently adding spatial redundancy to the network. They identify the structure of optimal policies, which are shown not always to be of a threshold type.

G. Neglia, in collaboration with X. Zhang (Fordham University, New York) and J. Kurose (University of Massachusetts at Amherst), has published a survey on the application of network coding to DTNs [78].

5.2.9.5. Ferry based local area networks

Polling systems are used to model the Ferry assisted Wireless LANs and thereby to obtain the stationary workload performance. Not much theory is available for calculating the stationary workload of polling systems with arrivals in a continuum. In [103], V. Kavitha (Univ. Avignon/LIA) and E. Altman propose a discretization approach, by which the so-called “pseudo conservation law” of the discrete polling systems is utilized to derive the stationary performance of continuous polling systems. The continuous polling results are used in deriving optimal ferry routes.

5.2.9.6. Adaptive epidemic routing in DTNs

G. Neglia and R. Masiero (University of Padua, Italy) have explored a recently proposed optimization framework that relies on local sub-gradient methods and consensus algorithms. The research is described in MAESTRO 2010 activity report and has appeared in [66].

5.2.9.7. Routing in quasi-deterministic networks

G. Neglia, U. Acer (Bell labs Antwerp), P. Giaccone and S. Tarapiah (both from Politecnico di Torino, Italy) and D. Hay (Hebrew University of Jerusalem), have investigated routing in DTNs where the underlying node mobility is known in advance but can be modified by random effects. The research is described in MAESTRO 2010 activity report and has appeared in [35] and [94].
5.3. Information systems

Participants: Eitan Altman, Konstantin Avrachenkov, Nicaise Choungmo Fofack, Majed Haddad, Alain Jean-Marie, Dorian Mazauric, Philippe Nain, Marina Sokol.

5.3.1. Web crawler optimization

A typical web search engine consists of three principal parts: crawling engine, indexing engine, and searching engine. The work [19] by K. Avrachenkov and P. Nain, together with A. Dudin, V. Klimenok, and O. Semenova (all three from Belarusian State University, Belarus), aims to optimize the performance of the crawling engine. The crawling engine finds new web pages and updates existing web pages in the database of the web search engine. The crawling engine has several robots collecting information from the Internet. The authors first calculate various performance measures of the system (e.g., probability of arbitrary page loss due to the buffer overflow, probability of starvation of the system, average time waiting in the buffer). Intuitively, one would like to avoid system starvation and at the same time to minimize the information loss. The authors formulate the problem as a multi-criteria optimization problem and solve it in the class of threshold policies. The authors consider a very general web page arrival process modeled by Batch Marked Markov Arrival Process and a very general service time modeled by Phase-type distribution. The model has been applied to the performance evaluation and optimization of the crawler designed by INRIA-MAESTRO team in the framework of the RIAM INRIA-Canon research project (see MAESTRO 2006 and 2007 activity reports).

5.3.2. PageRank node centrality

In [48] K. Avrachenkov and M. Sokol, together with D. Nemirovsky (former MAESTRO team member), E. Smirnova (INRIA project-team AXIS) and N. Litvak (University of Twente, The Netherlands), study a problem of quick detection of top-k Personalized PageRank (PPR) lists. This problem has a number of important applications such as finding local cuts in large graphs, estimation of similarity distance and person name disambiguation. The authors suggest that two observations are important when finding top-k PPR lists. Firstly, it is crucial that one detects fast the top-k most important neighbors of a node, while the exact order in the top-k list and the exact values of PPR are by far not so crucial. Secondly, by allowing a small number of “wrong” elements in top-k lists, one achieves great computational savings, in fact, without degrading the quality of the results. Based on these ideas, the authors propose Monte Carlo methods for quick detection of top-k PPR lists. We demonstrate the effectiveness of these methods on the Web and Wikipedia graphs, provide performance evaluation and supply stopping criteria.

5.3.3. Analysis of YouTube

E. Altman and M. Haddad, in collaboration with S.-E. Elayoubi (Orange Labs, Issy les Moulineaux), R. El-Azouzi, T. Jimenez and Y. Xu (all three from Univ. Avignon/LIA) have been investigating streaming protocols similar to the one used by YouTube. After preparing a survey on the state-of-the-art in [57], they used Ballot theorems in [106] in order to compute the starvation probabilities (these are the probability that the queue empties before completing to send a streaming application).

This work is carried out in the framework of the Grant with Orange Labs (see Section 6.3) on “Quality of Service and Quality of Experience”.

5.3.4. Peer-to-peer networks

5.3.4.1. Real-time control of contents download

In the course of the VOODDO project, the question of assessing the theoretical limits of prefetching information in real-time arose. Given a network bandwidth and a graph of documents, is it possible to download documents in advance, so that the document surfer is never blocked because of missing information? The problem is modeled using a “cops-and-robbers” game and some of its algorithmic properties are derived. This work of A. Jean-Marie and D. Mazauric is joint with F. Fomin (Univ. Bergen) and F. Giroire and N. Nisse (both from INRIA project-team MASCOTTE) [86].
5.3.4.2. P2P traffic classification

P2P downloads still represent a large portion of today’s Internet traffic. More than 100 million users operate BitTorrent and generate more than 30% of the total Internet traffic. According to the Wikipedia article about BitTorrent, the traffic generated by BitTorrent is greater than the traffic generated by Netflix and Hulu combined. Recently, a significant research effort has been done to develop tools for automatic classification of Internet traffic by application. The purpose of the work [47] by K. Avrachenkov and M. Sokol, together with A. Legout (INRIA project-team PLANETE) and P. Gonçalves (INRIA project-team RESO), is to provide a framework for subclassification of P2P traffic generated by the BitTorrent protocol. The general intuition is that users with similar interests download similar contents. This intuition can be rigorously formalized with the help of graph based semi-supervised learning approach. In particular, the authors propose to work with PageRank based semi-supervised learning method, which scales well with very large volumes of data.

5.3.4.3. BitTyrant

The success of BitTorrent has fostered the development of variants to its basic components. Some of the variants adopt greedy approaches aiming at exploiting the intrinsic altruism of the original version of BitTorrent in order to maximize the benefit of participating to a torrent. G. Neglia, D. Carra (University of Verona, Italy), P. Michiardi and F. Albanese (both from INSTITUT EURECOM) have studied BitTyrant, a recently proposed strategic client. The research is described in MAESTRO 2008 activity report. Results have been extended and supported by PlanetLab experiments in [22].

5.3.5. Content-centric networks

In [100] N. Choungmo Fofack, P. Nain and G. Neglia, together with D. Towsley (University of Massachusetts at Amherst), provide building blocks for the performance evaluation of Content Centric-like Networks (CCNs). In CCNs if a cache receives a request for a content it does not store, it forwards the request to a higher-level cache, if any, or to the server. When located, the document is routed on the reverse-path and a copy is placed in each cache along the path. In this work the authors consider a cache replacement policy based on Time-to-Lives (TTLs) like in a DNS network. A local TTL is set when the content is first stored at the cache and is renewed every time the cache can satisfy a request for this content (at each hit). The content is removed when the TTL expires. Under the assumption that requests follow a renewal process and the TTLs are exponential random variables, we determine exact formulas for the performance metrics of interest (average cache occupancy, hit and miss probabilities/rates) for some specific architectures (a linear network and a tree network with one root node and \(N\) leaf nodes). For more general topologies and general TTL distributions, an approximate solution is proposed. Numerical results show the approximations to be accurate, with relative errors smaller than \(10^{-3}\) and \(10^{-2}\) respectively for exponentially distributed and constant TTLs.

This work is carried out in the framework of the Grant with Orange Labs on “Content-centric networks” (Section 6.2).

5.4. Game theory applied to networking

Participants: Eitan Altman, Konstantin Avrachenkov, Majed Haddad, Manoj Panda, Giovanni Neglia.

5.4.1. Resource allocation in wireless networks

5.4.1.1. Power control

In [14] E. Altman, K. Avrachenkov and A. Garnaev (St. Petersburg State University, Russia), study power control for Gaussian interference channel in optimization and game frameworks. In the optimization framework there is a single decision maker who assigns network resources and in the game framework users share the network resources according to Nash equilibrium. The authors enhance the water-filling technique with explicit analytic solutions. The authors also provide an alternative simple proof of the convergence of the Iterative Water Filling Algorithm. Finally, the authors compare the non-cooperative approach with the cooperative approach and show that the non-cooperative approach results in a more fair resource distribution.
There has been a debate between those proposing protocols based on a centralized controller and those favoring decentralized protocols based on non-cooperative game theory. In [26] E. Altman and M. Haddad, together with S.-E. Elayoubi and Z. Altman (both from Orange Labs, Issy les Moulineaux), consider a situation where a base station lets mobiles take power control decisions in some system states and imposes actions in other states. The authors study how best to choose what information to make available and how mobiles should react.

5.4.1.2. Joint power and rate allocation

In [63] X. Lei and L. Cottatellucci (both from INSTITUT EURECOM) and K. Avrachenkov consider a block fading interference channels with partial channel state information and address the issue of joint power and rate allocation in a game theoretic framework. Resource allocation algorithms based on Bayesian games are proposed. The existence, uniqueness, and some stability properties of Nash equilibria are analyzed. For some asymptotic setting, closed-form expressions of Nash equilibria are also provided.

5.4.1.3. Jamming in wireless networks

Jamming is a form of a denial of service attack in which an adversary can degrade the quality of the reception by creating interference. One can study jamming both in the purpose of protecting a wireless network against such attack or, on the contrary, in order to efficiently disrupt the communications of some adversary. In both cases jamming is part of a conflict for which game theory is an appropriate tool. In [15] E. Altman, K. Avrachenkov and A. Garnaev (St. Petersburg State University), investigate the effect of partially available information in which the user does not even know whether or not the jammer is indeed present. The problem is formulated as a zero-sum game. The authors find the equilibrium strategies in closed-form and specify the range of sub-carriers where the user can expect the jamming attack.

5.4.1.4. Channel access

In WiFi networks, mobile nodes compete for accessing the shared channel by means of a random access protocol called Distributed Coordination Function (DCF), which is long term fair. Selfish nodes could benefit from violating the protocol and increasing their transmission probability. G. Neglia, I. Tinnirello and L. Giarré (University of Palermo, Italy) have been studying the interaction of selfish nodes in the last two years (the research activity is described in MAESTRO 2009 and MAESTRO 2010 activity reports). [31], [74] further extend the results to a heterogeneous scenario, where nodes have different requirements in terms of uplink/downlink ratios.

5.4.2. Network formation games

The continued growth of computer networks such as the Internet has raised the interest in understanding how networks get formed. The design of such networks is generally carried out by a large number of self-interested actors (users, Internet Service Providers ...), all of whom seek to optimize the quality and cost of their own operation. Previous works have addressed the “Network Formation” problem considering almost exclusively networks designed by selfish users, which can be consistently suboptimal. In [46] K. Avrachenkov and G. Neglia, together with J. Elias (University Paris Descartes), F. Martignon (University Paris-Sud 11), and L. Petrosyan (St. Petersburg State University, Russia), address the network formation issue using cooperative game theory, which permits to study ways to enforce and sustain cooperation among agents. Both the Nash bargaining solution and the Shapley value are investigated. After the comparison of these two approaches, the authors conclude that the Nash bargaining solution is more suitable to enforce cooperation in the network formation game in terms of cost allocation to users and computation time to get the solution.

5.4.2.1. Network design with socially-aware users

In many scenarios network design is not enforced by a central authority, but arises from the interactions of several self-interested agents. This is the case of the Internet itself. K. Avrachenkov and G. Neglia, in collaboration with J. Elias (University Paris Descartes) and F. Martignon (University Paris-Sud 11), have proposed two novel socially-aware network design games. The research has been described in MAESTRO 2010 activity reports. [24] extends the results for the case when users’ utility functions incorporate a socially-aware component.
5.4.2.2. Stochastic games for cooperative network routing

In [64] K. Avrachenkov, L. Maggi and L. Cottatellucci (both from INSTITUT EURECOM) consider a system where several providers share the same network and control the routing in disjoint sets of nodes. They provide connection toward a unique server (destination) to their customers. The objective is to facilitate the design of the available network links and their costs such that all network providers are interested in cooperating and none of them withdraw from the coalition. More specifically, the authors establish the framework of a coalition game by providing an algorithm to compute the transferable coalition values. As by-product, the authors apply the proposed algorithm to two-player games both in networks subject to hacker attacks and in epidemic networks.

5.4.2.3. Association games

Using tools from coalition game theory, E. Altman, in cooperation with C. Singh (IISc Bangalore, India), considers in [72] a wireless framework in which several mobile terminals can receive and decode the same signal of the base station, and where the cost for broadcasting is taken to be the transmission power. They begin by proposing various schemes to share the cost and study their properties. Then, they study the association with partial information: an arriving user knowing its location has to decide without knowledge of the location of the other users and their number whether to join the multicast tree and pay according to a given cost sharing scheme, or to have a unicast connection at a given cost. The unicast alternative that each mobile has, results in a limitation on the coverage (area covered by the multicast session) and on the capacity (number of mobiles connected to the multicast session). The authors derive the expected capacity and coverage as a function of the cost sharing policy. This work is extended in [58] to the case of several base stations by E. Altman in collaboration with C. Hasan and J. M. Gorce (both from INSA Lyon and INRIA project-team SWING).

5.4.3. Routing games

In [40], E. Altman, M. Panda and A. Estanisla (Master student at UPMC) study ring networks extensively used in both road traffic and telecommunications (in local area networks) in which each source with a given origin and destination on the ring, can split its traffic and send some part in one direction of the ring and some other part in the other direction. They compute the equilibria and find out that due to non-cooperation, much traffic is sent at equilibrium along long paths.

In [16], E. Altman, O. Pourtallier (INRIA project-team COPRIN), T. Jimenez (Univ. Avignon/LIA) and H. Kameda (Univ. Tsukuba, Japan) study a load balancing processor sharing problem. The classical framework of routing games turned out not to apply here. Indeed, it had been used to model situations where the flow from each class of users is split among paths without any information on the realization of the sizes of each packet. In contrast, in this paper, each individual knows its size. The authors have succeeded in computing the equilibrium within the new setting.

Collusion is the situation where several players decide to cooperate and to choose their actions as if they were a single player - each player maximizes the sum of utilities of that group instead of only its own utility. In [90], E. Altman in collaboration with Y. Hayel (Univ. Avignon/LIA) and H. Kameda (Univ. Tsukuba, Japan), has proposed various concepts that evaluate the impact of collusions. The authors have further studied collisions in routing games and identified situations where collusions are bad for all players; both those that collide loose in performance as well as those who remain independent.

5.5. Stochastic processes, queueing, control theory and game theory

Participants: Eitan Altman, Julien Gaillard, Majed Haddad, Alain Jean-Marie.

5.5.1. Convergence of rolling horizon control

In collaboration with E. Della Vecchia and S. Di Marco (both from National Univ. Rosario, Argentina), A. Jean-Marie has investigated the performance (convergence and error bounds) of the Rolling Horizon heuristic for optimal stochastic control and stochastic games in different modeling situations.
In the case of the long-term average expected gain, they have shown [85] that convergence occurs whenever the value iteration algorithm converges. They have then considered zero-sum semi-Markov games with discounted payoff [54], [76], for which they have proved geometric convergence under the usual assumptions of the literature.

5.5.2. Impulse control versus continuous control

Impulse control is a modeling framework of optimal control theory, in which the control actions can provoke instantaneous changes in the value of the state. For modelers, it has the features of both continuous-time and discrete-time models, and it can help understand which one to choose in a given optimization situation. A. Jean-Marie has studied the question in conjunction with K. Erdlenbruch (CEMAGREF), M. Tidball (INRA) and M. Moreaux (Univ. Toulouse 1). In a quite generic single-dimensional model, they show that the optimality of impulse policies with respect to “smooth” control policies is strongly related to a submodularity property of the instantaneous cost function [101].

5.5.3. Routing games

Several fundamental results have been obtained in routing games that model finite number of sources of traffic (players) who decide how to split the traffic among various paths. When the number of players is large, the Wardrop equilibrium concept is often used, where the problem is modeled as one with a continuum of decision makers where each has a negligible impact (non atomic game) on other’s performance. E. Altman and his co-workers have studied the question of whether Wardrop equilibrium is a good approximation for a problem with finitely many players for which the Nash equilibrium is the solution concept. In [38], E. Altman, in collaboration with Z. Altman, R. Combes (both from Orange Labs, Issy les Moulineaux) and S. Sorin (Univ. Pierre and Marie Curie (UPMC)) establishes the convergence under mild convexity assumptions on the link costs (or delays). The proof is based on yet another fundamental result derived in that reference and that was later extended in [44] by E. Altman in collaboration with O. Pourtallier (INRIA project-team COPRIN), T. Jimenez (Univ. Avignon/LIA) and H. Kameda (Univ. Tsukuba, Japan), that states that if there is some symmetry in a network then any Nash equilibrium will inherit the symmetric properties (for example, if two users have the same source and destination and the same demand then at equilibrium, they will send the same amount of traffic over each link).

In all the above work there is an assumption that the link cost (or delay) per packet is class independent (it depends on the flows through the link only through their sum). In the case of Wardrop equilibrium this assumption implies that the game has an equivalent global optimization problem whose solution coincides with the equilibrium. The link cost evaluated at some \( x \) in the equivalent problem is the integral of the original link cost (from 0 to \( x \)) and is in fact a potential. In the case of class dependent cost, that is, when the cost depends in other ways on the traffic of each class then the result of the integration may depend on the path and one cannot transform the problem to an equivalent optimization one. H. Kameda and J. Li (both from Univ. Tsukuba, Japan) in collaboration with E. Altman, identify in [27] other class-dependent cost that have the property of a field, that is, it can be expressed as the gradient of a potential. They obtain the Wardrop equilibrium and study its properties.

Another difficulty occurs in rouging games when the paths available are not the same for all users. This is the case, in particular, when there are priorities. This problem is addressed in [25] by J. Elias (University Paris Descartes), F. Martignion (University Paris-Sud 11), A. Capone (Politecnico di Milano, Italy) and E. Altman within an application to non-cooperative spectrum access in cognitive radio networks.

5.5.4. Bio-inspired paradigms

5.5.4.1. Epidemiology

For several years now, E. Altman has been developing techniques for dynamic optimal control and games in cooperation with S. Sarkar’s group from the University of Pennsylvania (which used to be part of the DAWN associated team with MAESTRO). This year this collaboration has resulted in three additional publications co-authored by M.H.R. Khouzani and S. Sarkar (both from Univ. of Pennsylvania, PA, USA) and E. Altman [61], [60], [104]. All three papers use the Pontriagin maximal principle to derive the structure
of optimal policies applied to a mean-field approximation of the problem. The first two papers do that in a context of optimal control theory while the third one does it in the context of a dynamic game.

5.5.4.2. Sequential anonymous games (SAG)

Sequential Anonymous Games (SAG) can be viewed as an extension of Markov Decision evolutionary games. In both formalisms there are many players modeled as a continuum number of players. A Markov chain is associated with each player. There are several types of players. The fraction of players in each class is called a global state and the state of the Markov chain of an individual is called the individual state. An individual chooses at some sequential decision opportunities actions. It earns some immediate reward (fitness) at each slot and moves with some probability to another individual state. In SAG, both the transition probabilities and the immediate fitness of an individual depend on its current state and action as well as on the current global system state. The latter evolves according to some function averaged over the fitness of the individuals in each class (the fraction of individuals in a class grows if they do better than those in the other classes). In [75] E. Altman investigates, in collaboration with P. Wiecek (Wroclaw Univ. of Technology, Poland), the case where the objective of an individual is to maximize either its total expected fitness during its life time or its expected average fitness. The authors establish the existence of equilibria and study its properties. Applications to power control have appeared in [32] by E. Altman, in collaboration with P. Wiecek (Wroclaw Univ. of Technology, Poland) and Y. Hayel (Univ. Avignon/LIA).

5.5.4.3. Markov decision evolutionary games (MDEGs)

Since his 2004 Infocom paper, E. Altman has been working on this novel paradigm. The model is similar to that of the previous paragraph (SAG) except that in MDEG both immediate fitness and transition probabilities depend linearly on the global state. This reflects a scenario where the interaction between a player and the rest of the population occurs through pairwise interactions: each player encounters from time to time a randomly chosen other player and it finds itself playing a matrix game with that player. The entries of the matrix corresponding to each player as well as the transition probabilities for each player depend on the individual states of the players. E. Altman has applied this model to the dynamic Hawk and Dove game, in which individuals have to choose the degree of aggressiveness in their behavior as a function of their energy state.

Below are three publications both with biological applications and applications to wireless communications (where depending on one’s remaining battery energy, one has to decide at what power to transmit). The first publication in [30], by E. Altman, H. Tembine (SUPELEC), R. El-Azouzi and Y. Hayel (both from Univ. Avignon/LIA), lays the foundations of MDEGs and presents the application to power control in which the individual state is the battery level of energy. The second publication in [59], by Y. Hayel (Univ. Avignon/LIA), E. V. Belmega (SUPELEC) and E. Altman, studies theoretical aspects that arise in case that the global state cannot represent the fractions of different populations but rather their actual size. The third publication in [97] by E. Altman, J. Gaillard, M. Haddad and P. Wiecek (Wroclaw Univ. of Technology, Poland) again studies MDEGs (as in the first paper), but restricts to policies that use static policies: the same mixed strategy is taken by a player at each state. The authors manage to compute explicitly the equilibrium in this game within this class of policies.

5.5.4.4. Delayed evolutionary games

Evolutionary game theory includes much theory on the description of the global system state as a function of the fitness of individuals. The models are often described through differential equations (e.g. the “replicator dynamics”). In many scenarios it is realistic to consider delays between the moment that one receives a given fitness till this is translated to a change in the population size. For example, if the lifetime of a computer is three years then an application that performs better with one computer may take more than a year till it is adopted by other users who do not have the same computer. In [30], H. Tembine (SUPELEC), E. Altman, R. El-Azouzi and Y. Hayel (both from Univ. Avignon/LIA) investigate instability phenomena that are introduced by the delay and derive necessary stability conditions.
6. New Results

6.1. Network Design and Optimization

Participants: Jean-Claude Bermond, Nathann Cohen, David Coudert, Frédéric Giroire, Dorian Mazauric, Joanna Moulierac, Nicolas Nisse, Ronan Pardo Soares, Issam Tahiri.

6.1.1. Backbone and Broadband Networks

Network design is a very wide subject that concerns all kinds of networks. We mainly study telecommunications networks which can be either physical networks (backbone, access, wireless, ...) or virtual (logical) ones. The objective is to design a network able to route a (given, estimated, dynamic, ...) traffic under some constraints (e.g. capacity) and with some quality of service (QoS) requirements. Usually the traffic is expressed as a family of requests with parameters attached to them. In order to satisfy these requests, we need to find one (or many) path(s) between their end nodes. The set of paths is chosen according to the technology, the protocol or the QoS constraints. For instance, optical backbones use the WDM technology to take better advantage of the capacity of the optical fibers often already installed. This is achieved through the multiplexing of several wavelength channels onto the same fiber. In that case a resource allocation is an optical channel, also called lightpath, which includes a path and wavelengths assigned to its links, one per link. If wavelength translation is performed in optical switching, then each channel may be assigned different wavelengths on the links of its path; otherwise the wavelength continuity imposes all the links to have the same wavelength. Of course, two lightpaths sharing a link must use different wavelengths on that link. The design can be done at the conception of the network (i.e. when conceiving a virtual network in MPLS where we have to establish virtual paths) or to adapt the network to changes (failures, new link, updates of routers, variation of traffic, ...). Finally there are various optimization criteria which differ according to the point of view: for a network user they are related to his/her satisfaction (minimizing delays, increasing available bandwidth, ...), while for a network operator, economics criteria like minimizing deployment and operating costs are more important.

This very wide topic is addressed by a lot of academic and industrial teams in the world. Our approach is to attack these problems with tools from Discrete Mathematics.

6.1.1.1. Traffic Grooming

In a WDM network, routing a connection request consists in assigning to this request a route in the physical network and a wavelength. When each request uses at most $1/C$ of the bandwidth of the wavelength, we say that the grooming factor is $C$. It means that on a given link of the network we can groom at most $C$ requests on the same wavelength. Under this constraint the objective can be either to minimize the number of wavelengths (related to the transmission cost) or to minimize the number of Add/Drop Multiplexers (ADM) used in the network (related to the cost of the nodes). During the last years, we have addressed this problem in various WDM network topologies with the goal of minimizing the total number of required ADMs.

This year, we considered the minimization of the number of ADMs in optical WDM bidirectional rings, considering symmetric shortest path routing and all-to-all unitary requests [24]. We formulate the problem in terms of graph decompositions, and state a general lower bound for all the values of the grooming factor $C$ and $N$, the size of the ring. We have studied exhaustively the cases $C = 1$, $C = 2$, and $C = 3$, providing improved lower bounds, optimal constructions for several infinite families, as well as asymptotically optimal constructions and approximations. We have also studied the case $C > 3$, focusing specifically on the case $C = k(k + 1)/2$ for some $k \geq 1$. We have also proposed optimal decompositions for several congruence classes of $N$ using the existence of some combinatorial designs.
6.1.1.2. Routing Reconfiguration and its Links with Graph Searching

In production networks, traffic evolution, failures and maintenance operations force to adapt regularly the current configuration of the network (virtual topology, routing of connections). The routing reconfiguration problem in WDM networks is thus to schedule the migration of established lightpaths from current routing to a new pre-computed one while minimizing service disruptions. We have shown in the past the relations between this problem and the graph searching problem (see also Section 6.4.3).

This year, we have continued studying the tradeoffs between the total number and the number of simultaneous interruptions that occurs during the reconfiguration process, proving in particular that the knowledge of one parameter does not help to optimize the other [28], [15]. We have also started investigating the influence of physical layer impairment constraints on the reconfiguration problem [74]. More precisely, using a new wavelength in a fiber of a WDM network forces to tune or recalibrate all already used wavelengths. We thus model the cost of using a new wavelength with a linear function of the number of already used wavelengths. We have then studied the problem of minimizing the cost of the reconfiguration according to this function. We have shown that this optimization problem is already NP-complete in a two-node network. We have also obtained general bounds and characterized instances for which the problem can be solved in polynomial time. We have additionally proposed and evaluated heuristics.

6.1.1.3. Green Networking

The minimization of ICT (Information and Communications Technologies) energy consumption has become a priority with the recent increase of energy cost and the new sensibility of public, governments and corporations towards energy consumption. ICT alone is responsible of 2% to 10% (depending on the estimations) of the world power consumption. For example, it is estimated that switches, hubs, routers account for 6 TWh per year in the US.

Several studies exhibit that the traffic load of the routers only has a small influence on their energy consumption. Hence, the power consumption in networks is strongly related to the number of active network elements, such as interfaces, line cards, base chassis, etc. In [78], [15], we have defined and modeled formally the problem of finding a routing that minimizes the (weighted) number of active network elements. We have proved that this problem is not in APX, that is there is no polynomial-time constant-factor approximation algorithm to solve it. We have obtained general bounds for this problem, and bounds for particular topologies such as trees, grids, and cliques. We have also proposed a heuristic algorithm offering good performance on real topologies. Last, we have analyzed the impact of energy efficient routing on the stretch factor and on fault tolerance.

We have also studied potential energy savings in fixed broadband wireless networks [77], [61]. See Section 6.1.2.1 for more details.

6.1.1.4. Xcast6 Treemap Islands

IP multicast is a protocol that deals with group communications with the aim of reducing traffic redundancy in the network. However, due to difficulty in deployment and poor scalability with a large number of multicast groups, IP multicast is still not widely deployed nor used on the Internet. Recently, Xcast6 and Xcast6 Treemap, the two network layer multicast protocols, have been proposed with complementary scaling properties to IP multicast: they support a very large number of active multicast sessions. However, the key limitation of these protocols is that they only support small multicast groups. To overcome this limitation, we have proposed the Xcast6 Treemap Island [96], a hybrid model of Application Layer Multicast (ALM) and Xcast6 that can work for large multicast groups. Our model has several advantages: ease of deployment, efficiency in bandwidth savings, no control message between end-host and router, zero multicast forwarding state at router and no need for a multicast address allocation protocol. In addition, this model is a potential service from which an ISP (Internet Service Provider) can get new revenue. We have shown the feasibility of our model by simulation and comparison with IP multicast and NICE protocols.

6.1.1.5. Time-Dependent Graphs - Applications to Transport Networks

In [70], we focus on time-dependent graphs which seem to be a good way to model transport networks. In the first part, we remind some notations and techniques related to time-dependent graphs. In the second one, we
introduce new algorithms to take into account the notion of probability related to paths in order to guarantee travelling times with a certain accuracy. We also discuss different probabilistic models and show the links between them.

Other results on multi-interface networks were obtained outside of MASCOTTE [37], [36], [65], [63], [66], [52], [20].

6.1.2. Wireless Networks

MASCOTTE has conducted an intense research effort on wireless access networks. From the technological and architectural point of view, the field is broad, from mesh (or multi-hop cellular) networks to ad-hoc and sensor networks. Nevertheless, many questions and approaches are generic from an algorithmic and structural prospect. In particular, we have considered three of the most prominent performance metrics for radio networks. Using combinatorial optimization and centralized algorithmic with a network design flavor, fast data gathering, call scheduling, transport capacity and energy consumption of the networks have been studied. Our approach is complementary with those developed in other INRIA project-teams such as PLANETE, MAESTRO, SWING, or POPS. The complementarity has been exploited through a joint Ph.D. between MAESTRO and MASCOTTE [15], through an ANR VERSO project in which MAESTRO, MASCOTTE, and SWING are involved, and through regular collaborations with POPS. At the international level, we cooperate with some groups in renowned research centers such as CTI of Patras in Greece, RWTH Aachen in Germany, Universities of Roma or Salerno in Italy, the Technion Institute in Israël, SFU in Vancouver, Canada, UFC Universidade Federal do Ceará, Fortaleza, Brazil, or the University of Sao Paulo in Brazil. We studied a wide range of issues of wireless networks, from the design of efficient cross-layer medium access, call scheduling and routing techniques to energy efficient optimization. We developed theoretical tools for integrating dynamic characteristics of the networks in the optimization models, and analyzing and evaluating dynamic networks. Some graph coloring problems motivated by channel assignment in wireless networks are detailed in Section 6.3.

6.1.2.1. Wireless Backhaul

We have investigated network optimization problems related to the design and configuration of fixed wireless microwave backhaul - the portion of the network infrastructure that provides interconnectivity between the access and the core networks. Unlike wired networks, the capacity of a microwave radio link is prone to variations, either due to external factors (e.g., weather) or by the action of the network operator. This fundamental difference raises a variety of new issues to be addressed appropriately. We concentrated on conceiving reliable fixed broadband wireless networks under outage probability constraints [60], [59]. We have developed a joint optimization of data routing and bandwidth assignment that minimizes the total renewal fees of licenses, while handling all the traffic requirements simultaneously. We have proposed a chance-constrained mathematical program taking into account unreliable channel conditions. This approach remains one of the main challenges of modern stochastic programming and it is still considered as very difficult and widely intractable. We have derived integer linear programming (ILP) counterparts for these chance-constrained programs and propose cutset-based valid inequalities to enhance the performance of ILP solvers. Computational results illustrate the price of reliability and present a comparative study on the performance of the different formulations. Moreover, we have been interested in potential energy savings in fixed broadband wireless networks by selectively turning off idle communication devices in low-demand scenarios [77], [61]. We have proposed a mathematical formulation of the problem relying on a fixed-charge capacitated network design (FCCND) problem, which is very hard to optimize. We have derived from this modeling heuristic algorithms producing feasible solutions in a short time. This work was done in collaboration with the SME 3Roam, and partially developed within the scope of the joint project RAISOM (Réseaux de collecte IP sans fil optimisés).

6.1.2.2. Wireless Mesh Networks

We have addressed the problem of computing the transport capacity of Wireless Mesh Networks (WMNs) dedicated to Internet access [26]. Routing and transmission scheduling have a major impact on the capacity provided to the clients. A cross-layer optimization of these problems allows the routing to take into account
contentions due to radio interference. We have presented a generic Mixed Integer Linear Programming (MILP) addressing gateway placement, routing, and scheduling optimizations in a WMN. We have then derived new optimization models that can take into account a large variety of radio interference models, and QoS requirements on the routing. We also provide efficient resolution methods that deal with realistic size instances. It allows to work around the combinatoric of simultaneously achievable transmissions and point out a critical region in the network bounding the network achievable capacity. Based upon strong duality arguments, it is then possible to restrict the computation to a bounded area. It allows for computing solutions very efficiently on large networks. We have then extended our models to deal with the dynamic characteristics of the network [75]. We have proposed a new robust optimization model that considers traffic demand uncertainty, in order to compute an optimal robust routing and bandwidth allocation in WMNs. We have presented a linear program efficiently solved by column generation, and we have quantified the price of robustness, i.e. the additional cost to pay in order to obtain a feasible solution for the robust scheme.

We have additionally investigated on the feasibility of providing network connectivity to vehicles over a predefined trajectory (trains, metros, urban buses, etc.) [14]. The communication between the vehicle and the infrastructure network is based only on WiFi technology. The contributions of this work are two-fold: 1) the horizontal handover (between WiFi access points) and 2) the design and analysis of an infrastructure network (backbone network plus WiFi access network) deployed along the trajectory of the vehicle.

6.1.2.3. Data Gathering

We have studied algorithmic and complexity issues originating from the problem of data gathering in wireless networks [56]. We give an algorithm to construct minimum makespan transmission schedules for data gathering when the communication graph is a tree network, the interference range is any integer \( m \geq 2 \), and no buffering is allowed at intermediate nodes. In the interesting case in which all nodes have to deliver an arbitrary non-zero number of packets, we provide a closed formula for the makespan of the optimal gathering schedule. Additionally, we consider the problem of determining the computational complexity of data gathering in general graphs and show that the problem is weakly NP-complete. On the positive side, we design a simple \( (1 + 2/m) \) factor approximation algorithm for general networks. We have also considered the data gathering process in multi-hop wireless sensor networks [76], [57]. Wireless sensors networks (WSNs) are deployed to collect huge amounts of data from the environment. This produced data has to be delivered through sensor’s wireless interface using multi-hop communications toward a sink. The position of the sink impacts the performance of the wireless sensor network regarding delay and energy consumption especially for relaying sensors. Optimizing the data gathering process in multi-hop wireless sensor networks is, therefore, a key issue. We have addressed the problem of data collection using mobile sinks in a WSN. We provide a multi-objective optimization framework that studies the trade-off between energy consumption and delay of data collection. This framework provides solutions that allow decision makers to optimally design the data gathering plan in wireless sensor networks with mobile sinks.

6.1.3. P2P Networks

6.1.3.1. Performance Analysis of Distributed Storage Systems

Distributed or peer-to-peer storage solutions rely on the introduction of redundant data to be fault-tolerant and to achieve high reliability. To ensure long-term fault tolerance, the storage system must have a self-repair service that continuously reconstructs lost fragments of redundancy. The speed of this reconstruction process is crucial for the data survival. In [93], we propose a new analytical framework, based on queuing models, to estimate the repair time and the probability of data loss. This model takes into account the correlation of concurrent repairs. The models and schemes proposed are validated by mathematical analysis, extensive set of simulations, and experimentation using the Grid’5000 test-bed platform. Recently, the Regenerating Codes were proposed as an improvement over classical replication and erasure codes to introduce redundancy. These codes make a better use of the available bandwidth when reconstructing the missing information. In [50], we propose a new code based on a hybrid approach, Double Coding, and compare it to existing codes from the point of view of availability, durability and storage space.
6.1.3.2. Well Balanced Designs for Data Placement

In collaboration with MAESTRO. The problem we consider in [88] is motivated by data placement, in particular, data replication in video on-demand systems. We are given a set $V$ of $n$ servers and $b$ files (data, documents). Each file is replicated on exactly $k$ servers. A placement consists in finding a family of $b$ subsets of $V$ (representing the files) called blocks each of size $k$. Each server has some probability to fail and we want to find a placement which minimizes the variance of the number of available files. It was conjectured that there always exists an optimal placement (with variance better than that of any other placement for any value of the probability of failure). We show that the conjecture is true if there exists a well balanced design, that is a family of blocks, such that each $j$-element subset of $V$, $1 \leq j \leq k$, belongs to the same or almost the same number of blocks (difference at most one). The existence of well balanced designs is a difficult problem as it contains as subproblem the existence of Steiner systems. We completely solve the case $k = 2$ and give bounds and constructions for $k = 3$ and some values of $n$ and $b$.

6.1.3.3. Peer-Assisted Time-shifted Streaming Systems: Design and Promises

Time-shifted streaming (or catch-up TV) allows viewers to watch their TV programs within an expanded time window. In [71], we emphasize the challenging characteristics of time-shifted TV systems that prevent known delivery systems to be used. We model time-shifted TV as multiple-interval graph, then we present a Peer-Assisted Catch-Up Streaming system, namely PACUS, where a set of end users’ computers assists the server for the content delivery. We show in particular how the PACUS tracker server can be efficiently implemented for catch-up TV. We demonstrate the benefits of PACUS by simulations. We especially highlight that PACUS reduces the traffic at the server side with the advantages of lightweight and self-adaptive unstructured peer-to-peer systems.

6.2. Simulation and Optimization Tools

Participants: Olivier Dalle, Luc Hogie, Aurélien Lancin, Emilio Mancini, Juan-Carlos Maureira, Philippe Mussi, Van Dan Nguyen, Judicaël Ribault, Issam Tahiri.

The works related to simulation and optimization tools address two kinds of issues: issues related to the development of the tools and their associated methodology, and issues related to the use of these tools in order to investigate a particular problem or assess the performances or properties of a particular system.

Since 2005, MASCOTTE has been developing a discrete event simulation architecture, named OSA, whose aim is to investigate how new software engineering techniques, such as component-based frameworks or Aspect Oriented Programming can help improving the simulation methodology, especially in terms of software reuse [16], [46], [47]. After six years of research development, OSA entered in the process of being diffused in 2011. This process is supported by a two-year INRIA “Development Action” (ADT) funding. This first year was devoted to cleaning the code base and produce a public release with a significant effort placed on user documentation and tutorial (cf http://osa.inria.fr/).

Aside our efforts on the OSA project, we are strongly involved in the USS-SimGrid ANR funded project, whose aim is at developing an efficient simulation platform geared at Grid Computing and very large scale distributed computing architectures. In this project, we worked on two tasks:

- Monitoring and characterization of the workload of large scale distributed applications [68];
- Support for modeling Peer-to-peer applications in the SimGrid simulator (originally designed for modeling grid-computing platforms).

We also pursued our involvement in the Discrete Event Systems Specification (DEVS) standardization effort [81], [82]. This formalism has reached a strong agreement amongst the community, but it still lacks implementation standard. Since OSA is aimed at providing better support for methodology, we consider necessary to support DEVS and participate to this effort. Our particular focus was on techniques and Architecture Description Languages (ADLs) for describing very large models of distributed applications [69].
Regarding our works on simulation studies and application-oriented developments, this year was the conclusion of our effort on the Internet on Rails project [14]. In this project, we studied and designed, both by means of simulations and experimentations, a low cost communication architecture based on IEEE 802.11 WiFi to provide high quality Internet access onboard high speed trains.

6.3. Graph Theory

Participants: Julio Araújo, Jean-Claude Bermond, Nathann Cohen, Frédéric Giroire, Frédéric Havet, František Kardoš, Ana Karolinna Maia, Dorian Mazauric, Remigiusz Modrzejewski, Leonardo Sampaio, Michel Syska.

MASCOTTE principally investigates applications in telecommunications via Graph Theory (see other objectives). However it also studies a number of theoretical problems of general interest. Our research mainly focused on graph coloring and some other problems arising from networks problems.

6.3.1. Graph Coloring

Coloring and edge-coloring are two central concepts in Graph Theory. There are many important and long-standing conjectures in these areas. We are trying to make advances towards such conjectures, in particular Steinberg’s conjecture, the List coloring Conjecture and the Acyclic Edge-Coloring Conjecture.

We are also interested in coloring problems arising from some practical problems: improper coloring, L(p,q)-labeling, directed star arboricity and good edge-labelling. The first two are both motivated by channel assignment and the last two by problems arising in WDM networks. For many practical problems are posed in a dynamic setting, we study on-line coloring and list coloring.

We also study some other variants of coloring like non-repetitive coloring or frugal coloring.

For all the coloring problems, we also consider the associated algorithmic problem, which consists in designing algorithms for finding the minimum number of colors of a coloring of a given graph. Algorithmic results on graph coloring are presented in Section 6.4.

The most classical notion of coloring (of edges or vertices) is the one of proper coloring, in which we insist on two adjacent elements to have distinct colors. However, it is usual to consider additional constraints, as well as relaxed constraints. For each variant of coloring, one can consider, its list version in which every element \( x \) is given a list \( L(x) \) of prescribed colors. A graph is said to be \( L \)-\( \text{colorable} \) if it has an \( L \)-coloring (fulfilling the constraints) such that \( x \in L(x) \) for all element \( x \). The \( \text{choosability} \) of a graph \( G \) is the smallest integer \( k \) for which \( G \) has an \( L \)-coloring whenever \( |L(x)| \geq k \) for all elements \( x \).

6.3.1.1. Coloring Graphs with Few Crossings

The famous Four Color Theorem states that every planar graph can be properly colored with 4 colors and Thomassen Five Color Theorem states that the choosability of every planar graph is at most 5. Hence, a natural question is to ask about the chromatic number and choosability of graphs with few crossings. In [38], we disprove a conjecture of Oporowski and Zhao stating that every graph with crossing number at most 5 and clique number at most 5 is 5-colorable. However, we show that every graph with crossing number at most 4 and clique number at most 5 is 5-colorable. We also show some colorability results on graphs that can be made planar by removing few edges. In particular, we show that if there exists three edges whose removal leaves the graph planar then it is 5-colorable. In [90], we show that every graph with two crossings is 5-choosable. We also prove that every graph which can be made planar by removing one edge is 5-choosable.

Another famous theorem on planar graphs is the one of Grötzsch, which says that every planar graph with no cycle of length 3 can be properly 3-colored. Steinberg’s Conjecture (1976) asserts that a graph with no cycles of length 4 or 5 is 3-colorable. Many approaches have been used towards this conjecture. We considered the following one in which, we relax the constraints on the color classes. Instead of insisting on them being independent sets, we allow them to induce a graph with some bounded degree. A graph \( G = (V, E) \) is said to be \((i, j, k)\)-\( \text{colorable} \) if its vertex set can be partitioned into three sets \( V_1, V_2, V_3 \) such that the graphs \( G[V_1], G[V_2], G[V_3] \) induced by the sets \( V_1, V_2, V_3 \) have maximum degree at most \( i, j, k \) respectively. Under
6.3.1.2. Acyclic, Linear and Frugal Colorings

A classical constraint added to a proper coloring is that at least three colors appears on each cycle, in which case we speak about acyclic coloring. In other words, the graph induced by the elements of any two color classes is a forest. The acyclic chromatic index of a graph $G$, denoted $\chi'_a(G)$ is the minimum $k$ such that $G$ admits an acyclic edge-coloring with $k$ colors. The famous Acyclic Edge-Coloring Conjecture asserts that $\chi'_a(G) = \Delta(G) + 2$, where $\Delta(G)$ is the maximum degree of the graph. In [21], we conjecture that if $G$ is planar and $\Delta(G)$ is large enough then $\chi'_a(G) = \Delta(G)$. We settle this conjecture for planar graphs with girth at least 5. We also show that $\chi'_a(G) \leq \Delta(G) + 12$ for all planar $G$.

Even stronger constraints are the following: a proper coloring of a graph is 2-frugal (resp. linear) if the graph induced by the elements of any two color classes is of maximum degree 2 (resp. is a forest of paths). In [29], we improve some bounds on the 2-frugal choosability and linear choosability of graphs with small maximum average degree.

6.3.1.3. Coloring of Plane Graphs with Constraints on the Faces

We studied several variants of vertex and edge colorings of plane graphs insisting on some constraints on the faces.

A face of a vertex colored plane graph is called loose if the number of colors used on its vertices is at least three. The looseness of a plane graph $G$ is the minimum $k$ such that any surjective $k$-coloring involves a loose face. In [35], we prove that the looseness of a connected plane graph $G$ equals the maximum number of vertex disjoint cycles in a dual graph $G^*$ increased by 2. We also show upper and lower bounds on the looseness of graphs based on the number of vertices, the edge connectivity, and the girth of the dual graph. These bounds improve the result of Negami for the looseness of plane triangulations. We also present infinite classes of graphs where the equalities are attained.

A vertex coloring of a 2-connected plane graph $G$ is a strong parity vertex coloring if for every face $f$ and each color $c$, the number of vertices incident with $f$ colored by $c$ is either zero or odd. Czap et al. [Discrete Math. 311 (2011) 512–520] proved that every 2-connected plane graph has a proper strong parity vertex coloring with at most 118 colors. In [34], we improve this upper bound for some classes of plane graphs.

A facial parity edge coloring of a connected bridgeless plane graph is such an edge coloring in which no two face-adjacent edges (consecutive edges of a facial walk of some face) receive the same color, in addition, for each face $f$ and each color $c$, either no edge or an odd number of edges incident with $f$ is colored with $c$. From Vizing’s theorem it follows that every 3-connected plane graph has a such coloring with at most $\Delta^* + 1$ colors, where $\Delta^*$ is the size of the largest face. In [33] we prove that any connected bridgeless plane graph has a facial parity edge coloring with at most 92 colors.

A sequence $r_1, r_2, \cdots, r_{2n}$ such that $r_i = r_{n+i}$ for all $1 \leq i \leq n$, is called a repetition. A sequence $S$ is called non-repetitive if no block (i.e. subsequence of consecutive terms of $S$) is a repetition. Let $G$ be a graph whose edges are colored. A trail is called non-repetitive if the sequence of colors of its edges is non-repetitive. If $G$ is a plane graph, a facial non-repetitive edge-coloring of $G$ is an edge-coloring such that any facial trail (i.e. trail of consecutive edges on the boundary walk of a face) is non-repetitive. We denote $\pi_f^r(G)$ the minimum number of colors of a facial non-repetitive edge-coloring of $G$. In [41], we show that $\pi_f^r(G) \leq 8$ for any plane graph $G$. We also get better upper bounds for $\pi_f^r(G)$ in the cases when $G$ is a tree, a plane triangulation, a simple 3-connected plane graph, a Hamiltonian plane graph, an outerplanar graph or a Halin graph. The bound 4 for trees is tight.

6.3.1.4. Improper Coloring

In [85] and [48], we study a coloring problem motivated by a practical frequency assignment problem and up to our best knowledge new. In wireless networks, a node interferes with the other nodes the level of interference depending on numerous parameters: distance between the nodes, geographical topography, obstacles, etc. We model this with a weighted graph $G$ where the weights on the edges represent the noise (interference) between
the two end-nodes. The total interference in a node is then the sum of all the noises of the nodes emitting on the same frequency. A weighted $t$-improper $k$-coloring of $G$ is a $k$-coloring of the nodes of $G$ (assignment of $k$ frequencies) such that the interference at each node does not exceed some threshold $t$. The Weighted Improper Coloring problem, that we consider here consists in determining the weighted $t$-improper chromatic number defined as the minimum integer $k$ such that $G$ admits a weighted $t$-improper $k$-coloring. We also consider the dual problem, denoted the Threshold Improper Coloring problem, where given a number $k$ of colors (frequencies) we want to determine the minimum real $t$ such that $G$ admits a weighted $t$-improper $k$-coloring. We show that both problems are NP-hard and first present general upper bounds; in particular we show a generalization of Lovász’s Theorem for the weighted $t$-improper chromatic number. We then show how to transform an instance of the Threshold Improper Coloring problem into another equivalent one where the weights are either 1 or $M$, for a sufficient big value $M$. Motivated by the original application, we study a special interference model on various grids (square, triangular, hexagonal) where a node produces a noise of intensity 1 for its neighbors and a noise of intensity 1/2 for the nodes that are at distance 2. Consequently, the problem consists of determining the weighted $t$-improper chromatic number when $G$ is the square of a grid and the weights of the edges are 1, if their end nodes are adjacent in the grid, and 1/2 otherwise. Finally, we model the problem using linear integer programming, propose and test heuristic and exact Branch-and-Bound algorithms on random cell-like graphs, namely the Poisson-Voronoi tessellations.

### 6.3.1.5. On-line Coloring

Several on-line algorithms producing colorings have been designed. The most basic and most widespread one is the greedy algorithm. The largest number of colours that can be given by the greedy algorithm on some graph, is called its Grundy number. Determining the Grundy number of a graph is NP-hard even for $P_4$-free graphs, while it is polynomial-time solvable for $P_4$-free graphs. In [19], we define a new class of graphs, namely the fat-extended $P_4$-laden graphs, which intersects the class of $P_5$-free graphs and strictly contains the one of $P_4$-free. We show a polynomial-time algorithm to determine the Grundy number of such graphs. It implies that the Grundy number can be computed in polynomial time for most graph classes defined in terms of containing few $P_4$’s: $P_4$-reducible, extended $P_4$-reducible, $P_4$-sparse, extended $P_4$-sparse, ...

In [94], we study a game version of greedy coloring. Given a graph $G = (V, E)$, two players, Alice and Bob, alternate their turns in choosing uncolored vertices to be colored. Whenever an uncolored vertex is chosen, it is colored by the least positive integer not used by any of its colored neighbors. Alice’s goal is to minimize the total number of colors used in the game, and Bob’s goal is to maximize it. The game Grundy number of $G$ is the number of colors used in the game when both players use optimal strategies. It is proved in this paper that the maximum game Grundy number of forests is 3, and the game Grundy number of any partial 2-tree is at most 7. We also gave some complexity results on $b$-colorings, which is a manner of improving colorings on-line [43].

### 6.3.1.6. Other Results on Graph Coloring

In [18], we aim at characterizing the class of graphs that admit a good edge-labelling. Such graphs are interesting, as they correspond to set of requests in UPP-digraphs (those in which there is at most one dipath from a vertex to another) for which the minimum number of wavelengths is equal to the maximum load. This implies that the problem can be solved efficiently. First, we exhibit infinite families of graphs for which no good edge-labelling can be found. We then show that deciding if a graph admits a good edge-labelling is NP-complete. Finally, we give large classes of graphs admitting a good edge-labelling: $C_3$-free outerplanar graphs, planar graphs of girth at least 6, subcubic $C_3$-K2, 3-free graphs.

A wheel is a graph formed by a chordless cycle and a vertex that has at least three neighbors in the cycle. We prove in [83] that every 3-connected graph that does not contain a wheel as a subgraph is in fact minimally 3-connected. We prove that every graph that does not contain a wheel as a subgraph is 3-colorable. We were then told that this result was already proved by Thomassen, though with a different proof. Gallai-Hasse-Roy-Vitaver Theorem states that every $n$-chromatic digraph contains a directed path of order $n$. Let $f(k)$ be the smallest integer such that every $f(k)$-chromatic digraph contains every oriented tree of order $k$. Burr proved that $f(k) \leq (k - 1)^2$ and conjectured $f(k) = 2n - 2$. In [84], we give some sufficient conditions
for an \( n \)-chromatic digraphs to contains some oriented tree. In particular, we show that every acyclic \( n \)-chromatic digraph contains every oriented tree of order \( n \). We also show that \( f(k) \leq k^2/2 - k/2 + 1 \). Finally, we consider the existence of antidirected trees in digraphs. We prove that every antidirected tree of order \( k \) is contained in every \( (5k - 9) \)-chromatic digraph. We conjecture that if \( |E(D)| > (k - 2)|V(D)| \), then the digraph \( D \) contains every antidirected tree of order \( k \). This generalizes Burr’s conjecture for antidirected trees and the celebrated Erdős-Sós Conjecture. We give some evidences for our conjecture to be true.

6.3.3. Hypergraphs

Hypergraphs, also called set systems, are a natural generalization of graphs. In a graph an edge is set of two vertices, while in a hypergraph an edge is a set of any size. It turns out to be an important notion in database theory. A digraph is \( \alpha \)-acyclic if it can be reduced to the null hypergraph by successively removing either a vertex which in at most one edge or an edge included in another. It is one of the possible generalizations of forest to hypergraphs. The \( \alpha \)-arboricity of a hypergraph \( H \) is the minimum number of \( \alpha \)-acyclic hypergraphs that partition the edge set of \( H \). In \cite{23}, the \( \alpha \)-arboricity of the complete 3-uniform (every edge is a set of 3 vertices) hypergraph is determined completely.

In \cite{45}, we show that every subcubic \( n \)-vertex graph with sufficiently large girth has fractional chromatic number at most 2.2978 which implies that it contains an independent set of size at least 0.4352\( n \). Our bound on the independence number is valid to random cubic graphs as well as it improves existing lower bounds on the maximum cut in cubic graphs with large girth.

In \cite{39}, we show that every cubic bridgeless graph \( G \) has at least \( 2|V(G)|/3656 \) perfect matchings. This confirms an old and celebrated conjecture of Lovász and Plummer in the 1970’s. This improves the first superlinear bound given in \cite{40}.

6.3.4. Miscellaneous

6.3.4.1. Zagreb Indices

The first and second Zagreb indices of a graph are defined by \( M_1 = \sum_{v \in V(G)} d(v)^2 \) and \( M_2 = \sum_{u \in V(G)} d(u)d(v) \), respectively. They are used in chemistry where it represents properties of molecules. In \cite{17}, we present some classes of graphs with prescribed degrees that satisfy \( M_1/n \leq M_2/m \), where \( M_1 \) and \( M_2 \) are the first and second Zagreb indices. We also prove that for any \( \Delta \geq 5 \), there is an infinite family of graphs of maximum degree \( \Delta \) such that the inequality is false. Moreover, we give alternative and slightly shorter proof of this inequality for trees and unicyclic graphs.

6.3.4.2. Induced Decomposition

An induced \( H \)-decomposition of a graph \( G \) is a partition \( (E_1, \cdots, E_k) \) of its edge set \( E(G) \), such that the graph induced by each \( E_i \), \( 1 \leq i \leq k \), is a copy of \( H \). Bondy and Szwarcflter asked for the maximum number \( ex(n,H) \) of edges in \( n \) vertices which admits an induced \( H \)-decomposition. In \cite{13}, we prove that for every non-empty graph \( H \), \( ex(n,H) = n(n - 1)/2 - o(n^2) \).

6.4. Algorithms

Participants: Julio Araújo, Janna Burman, Nathann Cohen, David Coudert, Gianlorenzo d’Angelo, Frédéric Giroire, František Kardoš, Dorian Mazauric, Nicolas Nisse, Stéphane Pérennes, Ronan Pardo Soares, Leonardo Sampaio.
MASCOTTE is also interested in the algorithmic aspects of Graph Theory. In general we try to find the most efficient algorithms to solve various problems of Graph Theory and telecommunication networks.

6.4.1. Coloring Graphs

Almost all graph coloring problems are NP-hard and most of them are even hard to approximate. Hence, to solve them efficiently, we aim at designing general exponential-time algorithms as well as polynomial-time algorithms for special classes. This is exemplified by the following results.

6.4.1.1. L(p, q)-labeling

An L(p, q)-labeling of $G$ is an integer assignment $f$ to the vertex set $V(G)$ such that $|f(u) - f(v)| \geq p$, if $u$ and $v$ are adjacent, and $|f(u) - f(v)| \geq q$, if $u$ and $v$ have a common neighbor. Such a concept is a modeling of a simple channel assignment, in which the separation between channels depends on the distance. The goal is to find an $L(p, q)$-labeling $f$ of $G$ with minimum span (i.e. $\max \{f(u) - f(v), u, v \in V(G)\}$). It is well known that for all $k \geq 4$, deciding if a graph has an $L(p, 1)$-labeling with minimum span $k$ is NP-complete. In [42], we present exact exponential time algorithms that are faster than existing ones.

6.4.1.2. Counting and Enumerating Total and Edge Colorings

In [89], we are interested in computing the number of edge colorings and total colorings of a graph. We prove that the maximum number of $k$-edge-colorings of a $k$-regular graph on $n$ vertices is $k \cdot (k - 1)^{n/2}$. Our proof is constructible and leads to a branching algorithm enumerating all the $k$-edge-colorings of a $k$-regular graph using a time $O^*((k - 1)^{n/2})$ and polynomial space. In particular, we obtain a algorithm on time $O^*(2^{n/2}) = O^*(1.4143^n)$ and polynomial space to enumerate all the 3-edge colorings of a cubic graph, improving the running time of $O^*(1.5423^n)$ of the algorithm due to Golovach et al. We also show that the number of 4-total-colorings of a connected cubic graph is at most $3.2^{3n/2}$. Again, our proof yields a branching algorithm to enumerate all the 4-total-colorings of a connected cubic graph.

6.4.1.3. Coloring Graphs of Special Classes

For some coloring problems that are known to be NP-hard for general graphs, we give some polynomial-time algorithms for the restriction to some graph classes. These graph classes defined in terms of forbidden induced subgraphs. In [95], [79], we provide linear algorithms for coloring $P_3$-free graphs. In [58], we obtain polynomial time algorithms to determine the acyclic chromatic number, the star chromatic number and the harmonious chromatic number of $(q, q - 4)$-graphs. Such graphs are those such that no set of at most $q$ vertices induces more than $q - 4$ distinct $P_4$’s.

6.4.2. Complexity and Computation of Graph Parameters

We used graph theory to model various networks’ problems. In general we study their complexity and then we investigate the structural properties of graphs that make these problems hard or easy. In particular, we try to find the most efficient algorithms to solve the problems, sometimes focusing on specific graph classes where the problems are polynomial-time solvable.

6.4.2.1. Path Vertex Cover

A subset $S$ of vertices of a graph $G$ is called a $k$-path vertex cover if every path of order $k$ in $G$ contains at least one vertex from $S$. The $k$-path vertex cover problem consists in finding such a set with minimum cardinality in $G$. In [25], it is shown that this problem is NP-complete for each $k \geq 2$ while it can be solved in linear-time in trees. The particular case of $k = 3$ is studied in [44], where an exact algorithm is given with running time $O^*(1.5171^n)$ in $n$-node graphs. In [44], we also design a polynomial time randomized approximation algorithm with an expected approximation ratio of $\frac{17}{11}$ for the minimum 3-path vertex cover.
6.4.2.2. Convexity in Graphs

The geodesic convexity of graphs naturally extends the notion of convexity in euclidean metric spaces. A set $S$ of vertices of a graph $G = (V, E)$ is convex if any vertex on a shortest path between two vertices of $S$ also belongs to $S$. The convex hull of $S \subseteq V$ is the smallest convex set containing $S$. Finally, a hull set of a graph is a set of vertices the convex hull of it is $V$. The hull number of a graph $G$ is the minimum size of a hull set in $G$. In [86], [49], we prove that computing the hull number is NP-complete in bipartite graphs. We also provide bounds and design various polynomial-time algorithms for this problem in different graph classes as co-bipartite graphs, $P_4$-sparse graphs, etc.

6.4.2.3. Induced Subdivision in Digraphs

In [51], we consider the following problem for oriented graphs and digraphs: Given an oriented graph (digraph) $G$, does it contain an induced subdivision of a prescribed digraph $D$? The complexity of this problem depends on $D$ and on whether $H$ must be an oriented graph or is allowed to contain 2-cycles. We give a number of examples of polynomial instances as well as several NP-completeness proofs.

6.4.2.4. Circuits in Grids

A circuit in a simple undirected graph $G = (V, E)$ is a sequence of vertices $\{v_1, v_2, \ldots, v_{k+1}\}$ such that $v_1 = v_{k+1}$ and $\{v_i, v_{i+1}\} \in E$ for $i = 1, \ldots, k$. A circuit $C$ is said to be edge-simple if no edge of $G$ is used twice in $C$. In [30], we study the following problem: which is the largest integer $k$ such that, given any subset of $k$ ordered vertices of an infinite square grid, there exists an edge-simple circuit visiting the $k$ vertices in the prescribed order? We prove that $k = 10$. To this end, we first provide a counterexample implying that $k < 11$. To show that $k \geq 10$, we introduce a methodology, based on the notion of core graph, to reduce drastically the number of possible vertex configurations, and then we test each one of the resulting configurations with an ILP solver.

6.4.3. Graph Searching, Cops and Robber Games

Pursuit-evasion encompasses a wide variety of combinatorial problems related to the capture of a fugitive residing in a network by a team of searchers. The goal consists in minimizing the number of searchers required to capture the fugitive in a network and in computing the corresponding capture strategy. This can also be viewed as cleaning the edges of a contaminated graph. We investigated several variants of these games.

6.4.3.1. Process Number and Routing Reconfiguration in WDM Networks

Graph searching, where the fugitive is arbitrary fast and moves simultaneously to the searchers, has been widely studied for its close relationship with graph decompositions. More recently, a variant of graph searching, namely the graph processing game, has been widely studied as a model for the routing reconfiguration in WDM networks (see Section 6.1.1.2). In [32], we give a linear time (resp., polynomial-time) algorithm to recognize graphs (resp., digraphs) with process number at most 2, along with a characterization in terms of forbidden minors, and a structural description. In [31], we give a polynomial (both in terms of time complexity and in the number of exchanged messages) distributed algorithm to compute the process number of trees. By slightly modifying the initial parameter of the algorithm, it also allows to compute various parameters of trees as pathwidth, search number, etc.

6.4.3.2. Cops and Robber Games

The “Cops and Robber” games are turn-by-turn games where a team of cops purchase a robber in a graph. We investigated two generalizations of the game introduced by Quilliot, Nowakoski and Winkler in 1983. We provided structural characterizations of graphs where one cop is sufficient to capture a fast fugitive able to hide [27]. In particular, one of these characterizations relies on hyperbolicity of the considered graph.

A surprising application of “Cops and Robber”-like games is the problem for a web-browser to download documents in advance while an internaut is surfing on the Web. In [92], we provide a modelling of the prefetching problem in terms of Cops and Robber games. The parameter to be optimized is then the download-speed necessary for the Internaut only accesses to already download webpages. This allows us to provide several complexity results and polynomial-time algorithms in some graph classes.
6.4.4. Distributed Algorithms

We investigated algorithmic problems arising in complex networks like the Internet or social networks. In this kind of networks, problems are becoming harder or impracticable because of the size and the dynamicity of these networks. One way to handle the dynamicity is to provide (distributed) fault tolerant algorithms. Studying the mobile agents paradigm seems to be a promising approach (somehow related to Cops and Robber in Section 6.4.3) to address some models of distributed computing. We considered self-stabilizing algorithms for the gathering problem, and algorithms for updating routing tables.

Besides, the more an algorithm uses local information, the easier it is to update/correct the behaviour of the algorithm. In this direction, we investigated communication problems through game theory. We also studied the power of a communication model using only localized information, i.e., we study what can be computed using this communication model.

6.4.4.1. Mobile Agents and Self-stabilization

In [64], we consider a recent model of robot-based computing which makes use of identical, memoryless mobile robots placed on nodes of anonymous graphs. The robots operate in Look-Compute-Move cycles that are performed asynchronously for each robot. In particular, we consider the case of gathering robots on an anonymous ring. We provide a new distributed approach which turns out to be very interesting as it neither completely falls into symmetry-breaking nor into symmetry-preserving techniques.

We address dynamic large scale emerging networks, e.g., mobile sensor (agent) networks. The agents are resource limited and prone to failures. They move almost unpredictably and communicate in pairs. Population Protocol model is a communication model suited for such networks. We use a recently proposed version of this model where every agent is associated with a parameter called Cover Time. Cover Times abstract the interaction characteristics of mobile agents and allow the design of fast converging protocols and the evaluation of their convergence times (this is impossible in the original model). We take advantage of this model and perform first analytical analysis of a data collection protocol used in the ZebraNet project for the wild-life tracking of zebras. We propose alternative data collection protocols for ZebraNet and we analyze their time complexities [72], [53], [54]. To achieve fault-tolerance in population protocols, we develop a generic self-stabilizing transformer [22]. This is an automatic technique to convert a protocol to its self-stabilizing version.

In addition, we address important problems of coordination and synchronization. We present and prove correct two self-stabilizing deterministic protocols solving the classical mutual exclusion problem and the group mutual exclusion one [54].

6.4.4.2. Distributed Update of Routing Tables

In [62], we propose a simple and practical distributed algorithm for computing and updating routing tables for shortest path routing. This algorithm can be combined with every distance vector shortest paths routing algorithm, and allows to reduce the total number of messages sent. We give experimental evidence that it leads to an important gain in terms of the number of messages sent at the price of a little increase in terms of space occupancy per node.

Arc-Flags is a data structure used to speed-up the shortest paths computation in a graph. In [67], we introduce a new data structure, named Road-Signs, which allows us to efficiently update the Arc-Flags of a graph in a dynamic scenario. Road-Signs can be used to compute Arc-Flags, can be efficiently updated and do not require large space consumption for many real-world graphs.

6.4.4.3. Models of Distributed Computation

Since, we need to face both locality and dynamicity issues, we are developing new techniques allowing to obtain global structural information from local (partial) views of the network. In [55], [73], we have investigated the question of determining which graph properties can or cannot be computed using only local information. We consider the following model: each of the $n$ nodes of a graph which only knows its own ID and the IDs of its neighbours is allowed to send a message of $O(\log n)$ bits to some central entity, called the referee. We then investigate whether the referee is able to decide some basic structural properties of the
network topology $G$ or not. We show that simple questions like, "does $G$ contain a square?", "does $G$ contain a triangle?" or "Is the diameter of $G$ at most 3?" cannot be solved in general [55], [73]. On the other hand, the referee can decode the messages in order to have full knowledge of $G$ when $G$ belongs to many graph classes such as planar graphs, bounded treewidth graphs and, more generally, bounded degeneracy graphs [55], [73]. Following our framework, we are able to simulate asynchronicity of the network. In particular, we have exhibited a hierarchy of problems and distributed models of computation [87].
6. New Results

6.1. Distributed Programming Models

6.1.1. Multi-active Objects

Participants: L. Henrio, I. Zsolt, F. Huet.

The active object programming model is particularly adapted to easily program distributed objects: it separates objects into several activities, each manipulated by a single thread, preventing data races. However, this programming model has its limitations in terms of expressiveness – risk of deadlocks – and of efficiency on multicore machines. We proposed extends active objects with local multi-threading. We rely on declarative annotations for expressing potential concurrency between requests, allowing easy and high-level expression of concurrency. This year contribution includes:

- publication of the basic principles of the new model [25]
- refinement of the proposal and adding dynamic compatibility
- operational semantic for multi-active objects
- extensive experiments

6.2. Component-oriented Distributed and Large-Scale Programming

6.2.1. Behavioural models for Distributed Components

Participants: E. Madelaine, R. Halalai, A. Savu, M. Alexe, L. Henrio.

In the past, we defined the behavioural semantics of active objects and components, in [3]; we extended last year this work to take group communications. On the practical side, this work contributes to the Vercors platform; the overall picture being to provide tools to the programmer for defining his application, including its behavioural specification. Then some generic properties like absence of deadlocks, but also application specific properties can be validated on the composed model using an existing model-checker. We mainly use CADP model-checker, that also supports distributed generation of state-space. This year our main achievements are the following:

- We provided model for one-to-many component communication
- We provided a model for Byzantine failures, specified a component application supporting some Byzantine faults, and proved its correctness;
- We conducted heavy experiments on distributed model-checking in this context;
- We worked on the formal specification of the behavioural model generation for component systems.

Most of those results were published in [22] and [35].

6.2.2. Enacting large-scale service orchestration using a component-based approach

Participants: F. Baude, V. Legrand.
The distribution of business processes encompasses the inclusion of external service providers in the overall process as well as the usage of external infrastructures like clouds. Both of these approaches lead to decentralization and outsourcing of a part of the global workflow, resulting in a complexified management of the global orchestration. As a matter of fact, the overall data are decentralized among different domains and must, most of the time, be gathered manually. To this extent, we continue our work on agile and distributed orchestration, showing that the framework we develop eases multidomain orchestration management. Our approach extracts, gathers and digests data from the decentralized processes in order to provide an unified and global view of a distributed orchestration. This year we focussed in particular on:

- The specification of the execution framework extending the SCA specification by adding temporal dependencies
- The definition of a use-case allowing the provisioning of a large set of OSGi gateways.

This work resulted in the PhD thesis of Virginie Legrand [12].

**6.2.3. Autonomic Monitoring and Management of Components**

**Participants:** F. Baude, C. Ruz, B. Sauvan, R. Dib.

We have completed the design of a framework for autonomic monitoring and management of component-based applications. We have provided an implementation using GCM/ProActive taking advantage of the possibility of adding components in the membrane, and we have tested it in simple applications. Our implementation allows the designer to describe in a separate way each phase of the MAPE autonomic control loop (Monitoring, Analysis, Planning, and Execution), and to plug them or unplug them dynamically [16].

- We presented the general description of the framework and its capability to support autonomic behaviour in [30]. This work takes advantage of the componentized membrane of GCM/ProActive, and of the PAGCMSScript reconfiguration extensions made in our team.

- We used our proposition to design an integrated framework to cover the life-cycle of a service application from business and design level, to deployment and execution concerns in a Cloud environment, in a work done in conjunction with Adrian Mos and Alain Boulze formerly leading the INRIA ADT Galaxy from INRIA Rhône-Alpes. This work was presented in the SoEA4EE workshop [29].

- We experimented with the use of our autonomic framework to integrate autonomic behaviour into skeletons. This work was taken during the engineering internship of Rima Dib, and included the collaboration with Marco Danelutto from Università di Pisa.

**6.3. Middleware for Grid and Cloud computing**

**6.3.1. RDF Data Storage and Retrieval In P2P Systems**

**Participants:** I. Filali, F. Huet, F. Baude, F. Bongiovanni, L. Pellegrino, B. Sauvan, I. Alshabani, A. Bourdin.

We have proposed in the context of the SOA4ALL FP7-IP project (8.3.1.1) the design and the implementation of a hierarchical Semantic Space infrastructure based on Structured Overlay Networks (SONS) [46],[10]. It aims at the storage and the retrieval of the semantic description of services at the Web scale [47]. This infrastructure combines the strengths of both P2P paradigm at the architectural level and the Resource Description Framework (RDF) data model at the knowledge representation level. As it is designed, the proposed infrastructure enables the processing of simple and complex queries. This year, the following achievements have been realised.

- A thorough survey of the existing works that have adapted the combination of RDF data model and the P2P communication model to build distributed infrastructures for RDF data storage and retrieval has been performed. This effort was published in a journal [34]. A previous but more complete version of this work can be found in a research report [45] and was used extensively in [39],[38],[36].
• We provided and presented in [23] an implementation of a three dimensional CAN overlay network for storing and retrieving RDF triples. At the implementation level, a modular and flexible architecture for the Semantic Space infrastructure has been proposed. The implementation relies on the ProActive Grid middleware and provides a clear separation between its sub-components (overlay, storage, query engine, etc.). The modularity of the architecture is combined with the decentralized aspect of the infrastructure enabling the RDF data storage and retrieval at large scale. The evaluation of the infrastructure through micro-benchmarks experiments on clusters and grids shows the impact of the architecture and data distribution on the performance of the storage and processing mechanisms.

In the context of the FP7 Strep PLAY (8.3.1.2) and French ANR SocEDA (8.2.2) research projects, we have extended the aforementioned work with a content-based Publish/Subscribe abstraction in order to support asynchronous queries for RDF-based events in large scale settings. In order to support these queries efficiently, we worked on an efficient broadcast primitive on top of CAN which we formalized and implemented (see section 6.3.2). We are also working towards a generalization of this broadcast algorithm to a multicast one, and contribute intensively to the general integration effort for offering such innovative semantic described event marketplace platform at cloud scale [41].

6.3.2. An algorithm for efficient broadcast over CAN-like P2P networks

Participants: L. Henrio, F. Bongiovani, A. Craciun.

The nature of some large-scale applications such as content delivery systems or publish/subscribe systems, built on top of structured overlay networks, demands application-level dissemination primitives which do not overwhelm the overlay and which are also reliable. Building such communication primitives in a reliable manner on top of such networks would increase the confidence regarding their behavior prior to deploying them in real settings. In order to come up with real efficient primitives, we take advantage of the underlying geometric topology of the overlay network and we also model the way peers communicate with one another. For this our objective is to design and prove an efficient and reliable broadcast algorithm for CAN-like P2P networks. To this aim, this year we:

• Formalised in Isabelle/HOL a CAN-like P2P system, devised formalised tools to reason on CAN topologies, and on communication protocols on top of CANs. We proved first completeness and correctness properties on some classes of broadcast algorithms.
• Designed an efficient broadcast algorithm on top of CAN and implemented it.

Preliminary results were presented at CFSE and published as a research report [37]; another publication is under way.

6.3.3. Matlab/Scilab parallel programming

Participant: F. Viale.

Matlab & Scilab, with millions of users around the world, are industry standards for numerical computing. They both lack a powerful and modern parallel computing framework to meet the industry’s growing demand in terms of parallel processing. This activity is intended to integrate into both softwares a toolbox for parallel processing, based on ProActive.

• This year, we implemented a ProActive Scilab toolbox with the same functionalities as the ProActive Matlab toolbox we built last year.
• We added in the Matlab toolbox a disconnected mode to allow closing the Matlab session while uncomplete Matlab jobs are still running on the scheduler side, and retrieving the job results at the next connection.
• We reorganized both Matlab and Scilab toolboxes with a cleaner and more intuitive API, a stronger and stabler implementation and an extensive documentation. We designed as well unit-tests to make the toolboxes fully usable for production standards.
• The Scilab toolbox is now deployed on PACAGrid cluster and used extensively by other partners of the OMD2 ANR.
6.3.4. Network Aware Cloud Computing

Participants: S. Malik, F. Huet.

We have proposed a cloud scheduler module named Network Awareness Module (NAM), which helps the scheduler to take the more efficient scheduling decisions on the basis of resource features, such as network latency, reliability, environment compatibility and monetary cost issues.

- Currently we are working on Reliability Assessment based Scheduling on Cloud Infrastructure. We are building a model, which enables a scheduler to schedule the tasks on cloud infrastructure, on the basis of adaptive reliability of nodes (virtual machine). The core of this model is a reliability assessment algorithm, which computes the reliability for individual resources and for the group of resources.

- We have proposed, designed and implemented an algorithm for the grouping of nodes on the basis of inter-node latencies. This algorithm can do the dynamic grouping and work with the incomplete latency information available. It groups the nodes on the basis of node latency instead of neighbor count. It produces mutually exclusive groups and can perform group reconfiguration.

- We have designed a model of the Virtual Cloud [27]. The concept of Virtual cloud revolves around the concept, “Rent Out the Rented Resources”. In this model, cloud vendors offer low cost cloud services by acquiring underutilized resources from some big third-party enterprise. The cloud vendor then further rents out those resources/services to the cloud users. The upfront and administrative costs for the Virtual cloud vendor are lower, and the cloud users access services at a cheaper rate.

- We have proposed a fault tolerance model for real time cloud computing [27]. In this model, the system tolerates the faults and makes the decision on the basis of reliability of the virtual machines. The reliability of the virtual machines is adaptive, which changes after every computing cycle. A metric model is given for the reliability assessment. The system provides both the forward and backward recovery mechanisms.
6. New Results

6.1. Towards Data-Centric Networking


- Disruption Tolerant Networking

We designed an efficient message delivery framework, called MeDeHa, which enables communication in an internet connecting heterogeneous networks that is prone to disruptions in connectivity[24]. MeDeHa is complementary to the IRTF’s Bundle Architecture: besides its ability to store messages for unavailable destinations, MeDeHa can bridge the connectivity gap between infrastructure-based and multi-hop infrastructure-less networks. It benefits from network heterogeneity (e.g., nodes supporting more than one network and nodes having diverse resources) to improve message delivery. For example, in IEEE 802.11 networks, participating nodes may use both infrastructure- and ad-hoc modes to deliver data to otherwise unavailable destinations. It also employs opportunistic routing to support nodes with episodic connectivity. One of MeDeHa’s key features is that any MeDeHa node can relay data to any destination and can act as a gateway to make two networks inter-operate or to connect to the backbone network. The network is able to store data destined to temporarily unavailable nodes till the time of their expiry. This time period depends upon current storage availability as well as quality-of-service needs (e.g., delivery delay bounds) imposed by the application. We showcase MeDeHa’s ability to operate in environments consisting of a diverse set of interconnected networks and evaluate its performance through extensive simulations using a variety of scenarios with realistic synthetic and real mobility traces. Our results show significant improvement in average delivery ratio and a significant decrease in average delivery delay in the face of episodic connectivity. We also demonstrate that MeDeHa supports different levels of quality-of-service through traffic differentiation and message prioritization.

Then, we have extended the MeDeHa framework to support multihop mobile ad-hoc networks (or MANETs). Integrating MANETs to infrastructure-based networks (wired or wireless) allows network coverage to be extended to regions where infrastructure deployment is sparse or nonexistent as well as a way to cope with intermittent connectivity. Indeed, to date there are no comprehensive solutions that integrate MANETs to infrastructure-based networks. We have proposed a message delivery framework that is able to bridge together infrastructure-based and infrastructure-less networks. Through extensive simulations, we have demonstrated the benefits of the extended MeDeHa architecture especially in terms of the extended coverage it provides as well as its ability to cope with arbitrarily long-lived connectivity disruptions. Another important contribution of this work is to deploy and evaluate our message delivery framework on a real network testbed as well as conduct experiments in "hybrid" scenarios running partly on simulation and partly on real nodes[32].

Finally, we have proposed a naming scheme for heterogeneous networks composed of infrastructure-based and infrastructure-less networks where nodes may be subject to intermittent connectivity. The proposed scheme, called Henna, aims at decoupling object identification from location and is designed to operate with status-quo Internet routing. We evaluated the proposed naming scheme using the ns-3 network simulator and demonstrated that nodes were able to receive messages in both infrastructure-based and infrastructure-less networks despite frequent disconnections and changing location identifiers (i.e., IP address), while visiting different networks[31].

Another important contribution of this work is to deploy and evaluate our message delivery framework on a real network testbed as well as conduct experiments in “hybrid” scenarios running partly on simulation and partly on real nodes. This was demonstrated at the ACM Sigcomm conference in Toronto on August 2011[74].
These different works are the result of collaborations with Katia Obrazcka and Marc Mendonca from University of California Santa Cruz (UCSC) in the context of the COMMUNITY Associated Team, see URL http://inrg.cse.ucsc.edu/community/.

Another activity in the same domain relates to efficient scheduling and drop policies in DTNs. We remind that Delay Tolerant Networks are wireless networks where disconnections may occur frequently. In order to achieve data delivery in such challenging environments, researchers have proposed the use of store-carry-and-forward protocols: there, a node may store a message in its buffer and carry it along for long periods of time, until an appropriate forwarding opportunity arises. Multiple message replicas are often propagated to increase delivery probability. This combination of long-term storage and replication imposes a high storage and bandwidth overhead. Thus, efficient scheduling and drop policies are necessary to:

(i) decide on the order by which messages should be replicated when contact durations are limited, and
(ii) which messages should be discarded when nodes' buffers operate close to their capacity.

We worked on an optimal scheduling and drop policy that can optimize different performance metrics, such as the average delivery rate and the average delivery delay. First, we derived an optimal policy using global knowledge about the network, then we introduced a distributed algorithm that collects statistics about network history and uses appropriate estimators for the global knowledge required by the optimal policy, in practice. At the end, we are able to associate to each message inside the network a utility value that can be calculated locally, and that allows to compare it to other messages upon scheduling and buffer congestion. Our solution called HBSD (History Based Scheduling and Drop) integrates methods to reduce the overhead of the history-collection plane and to adapt to network conditions. The first version of HBSD and the theory behind have been published in 2008. A recent paper [27] provides an extension to a heterogenous mobility scenario in addition to refinements to the history collection algorithm. An implementation is proposed for the DTN2 architecture as an external router and experiments have been carried out by both real trace driven simulations and experiments over the SCORPION testbed at the University of California Santa Cruz. We refer to the web page of HBSD for more details http://planete.inria.fr/HBSD_DTN2/.

HBSD in its current version is for point-to-point communications. Another interesting schema is to consider one-to-many communications, where requesters for content express their interests to the network, which looks for the content on their behalf and delivers it back to them. We are working on this extension within a new framework called MobiTrade, which provides a utility driven trading system for efficient content dissemination on top of a disruption tolerant network. While simple tit-for-tat (TFT) mechanisms can force nodes to give one to get one, dealing with the inherent tendency of peers to take much but give back little, they can quickly lead to deadlocks when some (or most) of interesting content must be somehow fetched across the network. To resolve this, MobiTrade proposes a trading mechanism that allows a node (merchant) to buy, store, and carry content for other nodes (its clients) so that it can later trade it for content it is personally interested in. To exploit this extra degree of freedom, MobiTrade nodes continuously profile the type of content requested and the collaboration level of encountered devices. An appropriate utility function is then used to collect an optimal inventory that maximizes the expected value of stored content for future encounters, matched to the observed mobility patterns, interest patterns, and collaboration levels of encountered nodes. Using ns-3 simulations based on synthetic and real mobility traces, we show that MobiTrade achieves up to 2 times higher query success rates compared to other content dissemination schemes. Furthermore, we show that MobiTrade successfully isolates selfish devices. For further details on MobiTrade, we refer to [41] and to the web page of the project where the code can be downloaded for both the ns-3 simulator and Android devices.

- Naming and Routing in Content Centric Networks

1 http://planete.inria.fr/MobiTrade/
Content distribution prevails in today's Internet and content-oriented networking proposes to access data directly by their content name instead of their location, changing so the way routing must be conceived. We worked a routing mechanism that faces the new challenge of interconnecting content-oriented networks. Our solution relies on a naming resolution infrastructure that provides the binding between the content name and the content networks that can provide it. Content-oriented messages are sent encapsulated in IP packets between the content-oriented networks. In order to allow scalability and policy management, as well as traffic popularity independence, binding requests are always transmitted to the content owner. The content owner can then dynamically learn the caches in the network and adapt its binding to leverage the cache use.

The work done so far is related to routing between content-oriented networks. We are starting an activity on how to provide routing inside a content network. To that aim, we are investigating on the one hand probabilistic routing and, on the other hand, deterministic routing and possible extension to Bellman-Ford techniques. In addition to routing, we are investigating the problem of congestion in content-oriented networks. Indeed, in this new paradigm, congestion must be controlled on a per-hop basis, as opposed to the end-to-end congestion control that prevails today. We think that we can combine routing and congestion control to optimize resource consumption. Finally, we are studying the implications of using CCN from an economical perspective. This activity was started in October 2011 by Damien Saucez.

- **Application-Level Forward Error Correction Codes (AL-FEC) and their Applications to Broadcast/Multicast Systems**

With the advent of broadcast/multicast systems (e.g., DVB-H/SH), large scale content broadcasting is becoming a key technology. This type of data distribution scheme largely relies on the use of Application Level Forward Error Correction codes (AL-FEC), not only to recover from erasures but also to improve the content broadcasting scheme itself (e.g., with FLUTE/ALC).

Our recent activities, in the context of the PhD of F. Mattoussi, included the design, analysis and improvement of GLDPC-Staircase codes, a "Generalized" extension to LDPC-Staircase codes. We have shown in particular that these codes: (1) offer small rate capabilities, i.e. can produce a large number of repair symbols 'on-the-fly', when needed; (2) feature high erasure recovery capabilities, close to that of ideal codes. Therefore they offer a nice opportunity to extend the field of application of existing LDPC-Staircase codes, while keeping backward compatibility (LDPC-Staircase "codewords" can be decoded with a GLDPC-Staircase codec).

Our LDPC-Staircase codes, that offer a good balance in terms of performance, have been included as the primary AL-FEC solution for ISDB-Tmm (Integrated Services Digital Broadcasting, Terrestrial Mobile Multimedia), a Japanese standard for digital television (DTV) and digital radio. This is the first adoption of these codes in an international standard.

This success has been made possible, on the one hand, by major efforts in terms of standardization within IETF: the RFC 5170 (2008) defines the codes and their use in FLUTE/ALC, a protocol stack for massively scalable and reliable content delivery services, an active Internet-Draft published last year describes the use of these AL-FEC codes in FECFRAME, a framework for robust real-time streaming applications, and a recent Internet-Draft [66] defines the GOE (Generalized Object Encoding) extension of LDPC-Staircase codes for UEP (Unequal Erasure Protection) and file bundle protection services.

This success has also been made possible, on the other hand, by our efforts in terms of design and evaluation of two efficient software codecs of LDPC-Staircase codes. One of them is distributed in open-source, as part of our OpenFEC project (http://openfec.org), a unique initiative that aims at promoting open and free AL-FEC solutions. The second one, a highly optimized version with improved decoding speed and reduced memory requirements, will be commercialized in 2012.
through an industrial partner. This codec proves that LDPC-Staircase codes can offer erasure recovery performances close to ideal codes in many circumstances while keeping decoding speeds over 1Gbps.

The fact that LDPC-Staircase codes have been preferred to a major AL-FEC competitor for the ISDB-Tmm standard, is the recognition of their intrinsic qualities and of an appropriate balance between several technical and non technical criteria.

- Unequal Erasure Protection (UEP) and File bundle protection through the GOE (Generalized Object Encoding) scheme

This activity has been initiated with the PostDoc work of Rodrigue IMAD. It focuses on Unequal Erasure Protection capabilities (UEP) (when a subset of an object has more importance than the remaining) and file bundle protection capabilities (e.g. when one want to globally protect a large set of small objects).

After an in-depth understanding of the well-known PET (Priority Encoding Technique) scheme, and the UOD for RaptorQ (Universal Object Delivery) initiative of Qualcomm, which is a realization of the PET approach, we have designed the GOE FEC Scheme (Generalized Object Encoding) alternative. The idea, simple, is to decouple the FEC protection from the natural object boundaries, and to apply an independent FEC encoding to each "generalized object". The main difficulty is to find an appropriate signaling solution to synchronize the sender and receiver on the exact way FEC encoding is applied. In [65] we show this is feasible, while keeping a backward compatibility with receivers that do not support GOE FEC schemes. Two well known AL-FEC schemes have also been extended to support this new approach, with very minimal modifications, namely Reed-Solomon and LDPC-Staircase codes [66], [65].

During this work, we compared the GOE and UOD/PET schemes, both from an analytical point of view (we use an N-truncated negative binomial distribution to that purpose) and from an experimental, simulation based, point of view [67]. We have shown that the GOE approach, by the flexibility it offers, its simplicity, its backward compatibility and its good recovery capabilities (under finite or infinite length conditions), outperforms UOD/PET for practical realizations of UEP/file bundle protection systems. See also http://www.ietf.org/proceedings/81/slides/rmt-2.pdf.

- Application-Level Forward Error Correction Codes (AL-FEC) and their Applications to Robust Streaming Systems

AL-FEC codes are known to be useful to protect time-constrained flows. The goal of the IETF FECFRAME working group is to design a generic framework to enable various kinds of AL-FEC schemes to be integrated within RTP/UDP (or similar) data flows. Our contributions in the IETF context are three fold. First of all, we have contributed to the design and standardization of the FECFRAME framework, now published as a Standards Track RFC [68].

Secondly, we have proposed the use of Reed-Solomon codes (with and without RTP encapsulation of repair packets) and LDPC-Staircase codes within the FECFRAME framework: [59] [60] [61].

Finally, in parallel, we have started an implementation of the FECFRAME framework in order to gain an in-depth understanding of the system. Previous results showed the benefits of LDPC-Staircase codes when dealing with high bit-rate real-time flows.

A second type of activity, in the context of robust streaming systems, consisted in the analysis of the Tetrys approach, in [29]. Tetrys is a promising technique that features high reliability while being independent from RTT, and performs better than traditional block FEC techniques in a wide range of operational conditions.
• A new File Delivery Application for Broadcast/Multicast Systems

FLUTE has long been the one and only official file delivery application on top of the ALC reliable multicast transport protocol. However FLUTE has several limitations (essentially because the object meta-data are transmitted independently of the objects themselves, in spite of their inter-dependency), features an intrinsic complexity, and is only available for ALC.

Therefore, we started the design of FCAST, a simple, lightweight file transfer application, that works both on top of both ALC and NORM. This work is carried out as part of the IETF RMT Working Group, in collaboration with B. Adamson (NRL). This document has passed WG Last Call and is currently considered by IESG[ 56 ], [ 57 ], [ 58 ].

• Security of the Broadcast/Multicast Systems

We believe that sooner or later, broadcasting systems will require security services. This is all the more true as heterogeneous broadcasting technologies will be used, for instance hybrid satellite-based and terrestrial networks, some of them being by nature open, as wireless networks (e.g., wimax, wifi). Therefore, one of the key security services is the authentication of the packet origin, and the packet integrity check. A key point is the ability for the terminal to perform these checks easily (the terminal often has limited processing and energy capabilities), while being tolerant to packet losses.

The TESLA (Timed Efficient Stream Loss-tolerant Authentication) scheme fulfills these requirements. We are therefore standardizing the use of TESLA in the context of the ALC and NORM reliable multicast transport protocols, within the IETF MSEC working group. This document has been published as RFC 5776.

In parallel, we have specified the use of simple authentication and integrity schemes (i.e., group MAC and digital signatures) in the context of the ALC and NORM protocols in [ 62 ], [ 63 ], [ 64 ]. This activity is also carried out within the IETF RMT working group.

• High Performance Security Gateways for High Assurance Environments

This work focuses on very high performance security gateways, compatible with 10Gbps or higher IPsec tunneling throughput, while offering a high assurance thanks in particular to a clear red/black flow separation. In this context we have studied last year the feasibility of high-bandwidth, secure communications on generic machines equipped with the latest CPUs and General-Purpose Graphical Processing Units (GPGPU).

The work carried out in 2011 has consisted in setting up and evaluating the high performance platform. This platform heavily relies on the Click modular TCP/IP protocol stack implementation, which turned out to be a key enabler both in terms of specialization of the stack and parallel processing. Our activities also consisted in analyzing the PMTU discovery aspect since it is a critical factor in achieving high bandwidths. To that goal we have designed a new approach for qualifying ICMP blackholes in the Internet, since PMTUD heavily relies on ICMP.

6.2. Network Security and Privacy

Participants: Sana Ben Hamida, Claude Castelluccia, Walid Dabbous, Mohamed Ali Kaafar, Arnaud Legout, Stevens Le Blond, Daniele Perito.

• Online users tracking and profiling techniques
Usernames are ubiquitously used for identification and authentication purposes on web services and the Internet at large, ranging from the local-part of email addresses to identifiers in social networks. Usernames are generally alphanumerical strings chosen by the users and, by design, are unique within the scope of a single organization or web service. In this work, we investigate the feasibility of using usernames to trace or link multiple profiles across services that belong to the same individual. The intuition is that the probability that two usernames refer to the same physical person strongly depends on the entropy of the username string itself. Our experiments, based on usernames gathered from real web services, show that a significant portion of the users’ profiles can be linked using their usernames. In collecting the data needed for our study, we also show that users tend to choose a small number of related usernames and use them across many services. This work is the first to consider usernames as a source of information when profiling users on the Internet. It has been published in PETS 2011 [47], one of the most prestigious conference in the area of Computer Privacy, and has been awarded the Andreas Pfitzmann award for the best contribution.

**Online Privacy measurements and threats identification in online social networks**

In this work, we show how these seemingly harmless interests (e.g., Music Interests) can leak privacy-sensitive information about users. In particular, we infer their undisclosed (private) attributes using the public attributes of other users sharing similar interests. In order to compare user-defined interest names, we extract their semantics using an ontologized version of Wikipedia and measure their similarity by applying a statistical learning method. Besides self-declared interests in Music, our technique does not rely on any further information about users such as friends relationship or group belongings. Our experiments, based on more than 104K public profiles collected from Facebook and more than 2000 private profiles provided by volunteers, show that our inference technique efficiently predicts attributes that are very often hidden by users. To the best of our knowledge, this is the first time that user interests are used for profiling, and more generally, semantics-driven inference of private data is addressed. This work has been published in the prestigious Network & Distributed System Security Symposium (NDSS) 2012 [37].

**Privacy Enhancing Technologies**

The increasing amount of personal and sensitive information disseminated over the Internet prompts commensurately growing privacy concerns. Digital data often lingers indefinitely and users lose its control. This motivates the desire to restrict content availability to an expiration time set by the data owner. This work presents and formalizes the notion of Ephemeral Publishing (EphPub), to prevent the access to expired content. We propose an efficient and robust protocol that builds on the Domain Name System (DNS) and its caching mechanism. With EphPub, sensitive content is published encrypted and the key material is distributed, in a steganographic manner, to randomly selected and independent resolvers. The availability of content is then limited by the evanescence of DNS cache entries. The EphPub protocol is transparent to existing applications, and does not rely on trusted hardware, centralized servers, or user proactive actions. We analyze its robustness and show that it incurs a negligible overhead on the DNS infrastructure. We also perform a large-scale study of the caching behavior of 900K open DNS resolvers. Finally, we propose an Android application, Firefox and Thunderbird extensions that provide ephemeral publishing capabilities, as well as a command-line tool to create ephemeral files. This work has been published in ICNP 2011 [36].

**Differentially private smart metering**
Several countries throughout the world are planning to deploy smart meters in households in the very near future. The main motivation, for governments and electricity suppliers, is to be able to match consumption with generation. Traditional electrical meters only measure total consumption on a given period of time (i.e., one month or one year). As such, they do not provide accurate information of when the energy was consumed. Smart meters, instead, monitor and report consumption in intervals of few minutes. They allow the utility provider to monitor, almost in realtime, consumption and possibly adjust generation and prices according to the demand. Although smart metering might help improving energy management, it creates many new privacy problems. Smart meters provide very accurate consumption data to electricity providers. As the interval of data collected by smart meters decreases, the ability to disaggregate low-resolution data increases.

We developed a new privacy-preserving smart metering system. Our scheme is private under the differential privacy model and therefore provides strong and provable guarantees. With our scheme, an (electricity) supplier can periodically collect data from smart meters and derive aggregated statistics while learning only limited information about the activities of individual households. For example, a supplier cannot tell from a user’s trace when he watched TV or turned on heating. Our scheme is simple, efficient and practical. Processing cost is very limited: smart meters only have to add noise to their data and encrypt the results with an efficient stream cipher.

This work was presented at IH’11 (the Information Hiding Conference, 2011) [34].

- **Protecting against Physical Resource Monitoring**

This work considers the problem of resource monitoring. We consider the scenario where an adversary is physically monitoring on the resource access, such as the electricity line or gas pipeline, of a user in order to learn private information about his victim. Recent works, in the context of smart metering, have shown that a motivated adversary can basically profile a user or a family solely from his electricity traces. However, these works only consider the case of a semi-honest-but-non-intrusive adversary that is only trying to learn information from the consumption reports sent by the user. This work, instead, considers the much more challenging case of an intrusive semi-honest adversary, i.e. a semi-honest adversary that is in addition physically monitoring the resource by modifying the distribution network. We aim at answering to the following question: is it possible to design a resource distribution scheme that prevents resource monitoring and provides strong protection? We propose and analyze several possible solutions. The proposed solutions provide different privacy bounds and performance results. This work was presented at WPES’11 (ACM Workshop on Privacy in the Electronic Society) [35].

- **The Failure of Noise-Based Non-Continuous Audio Captchas**

CAPTCHAs, which are automated tests intended to distinguish humans from programs, are used on many web sites to prevent bot-based account creation and spam. To avoid imposing undue user friction, CAPTCHAs must be easy for humans and difficult for machines. However, the scientific basis for successful CAPTCHA design is still emerging. This project examines the widely used class of audio CAPTCHAs based on distorting non-continuous speech with certain classes of noise and demonstrates that virtually all current schemes, including ones from Microsoft, Yahoo, and eBay, are easily broken. More generally, we describe a set of fundamental techniques, packaged together in our Decaptcha system, that effectively defeat a wide class of audio CAPTCHAs based on non-continuous speech. Decaptcha’s performance on actual observed and synthetic CAPTCHAs indicates that such speech CAPTCHAs are inherently weak and, because of the importance of audio for various classes of users, alternative audio CAPTCHAs must be developed.

This work was presented at IEEE Security and Privacy 2011 [33].
• **BlueBear: Privacy in P2P systems**

We have started a new project called bluebear on privacy threats in the Internet. Indeed, the Internet has never been designed with privacy in mind. For instance, the Internet is based on the IP protocol that exposes the IP address of a user to any other users it is communicating with. However, we believe that current users of the Internet do not realize how much they compromise their privacy by using the Internet. Indeed, the common wisdom is that there are so many users in the Internet that it is not feasible for an attacker, apart may be for national agencies, to globally compromise the privacy of a large fraction of users. Therefore, finding a specific user is like looking for a needle in a haystack. The goal of the bluebear project is to raise attention on privacy issues when using the Internet. In particular, we want to show that without any dedicated infrastructure, it is possible to globally compromise the privacy of Internet users. BitTorrent is arguably the most efficient peer-to-peer protocol for content replication. However, BitTorrent has not been designed with privacy in mind and its popularity could threaten the privacy of millions of users.

In a first study we showed that it is possible to continuously monitor from a single machine most BitTorrent users and to identify the content providers (also called initial seeds). We performed a very large monitoring operation continuously “spying” on most BitTorrent users of the Internet from a single machine and for a long period of time. During a period of 103 days, we collected 148 million IP addresses downloading 2 billion copies of contents. We then identified the IP address of the content providers for 70% of the BitTorrent contents we spied on. We showed that a few content providers inject most contents into BitTorrent and that those content providers are located in foreign data centres. We also showed that an adversary could compromise the privacy of any peer in BitTorrent and identify the big downloaders that we define as the peers who subscribe to a large number of contents. This is a major privacy threat as it is possible for anybody in the Internet to reconstruct all the download and upload history of most BitTorrent users. This work was published in LEET 2010.

To circumvent this kind of monitoring, BitTorrent users are increasingly using anonymizing networks such as TOR to hide their IP address from the tracker and, possibly, from other peers. We explored in a second study whose goal was to Exploit P2P Applications to Trace and Profile Tor Users, to which extent a P2P protocol such as BitTorrent, when not designed to protect users information, leak information that may compromise the identity of users. We quantified such an issue with BitTorrent on top of anonymizing networks. We also designed an attack that reveals the identity of Tor users (We showed that it is possible to retrieve the IP address for more than 70% of BitTorrent users on top of TOR). Moreover, once the IP address of a peer is retrieved, it is possible to link to the IP address other applications used by this peer on top of TOR [45].

The fact that it is hard for a person to map an IP address to an identity mitigates the impact of the privacy attacks we described. However, we show that we can exploit a peer-to-peer VoIP system to associate a social identity (name, email address, etc.) to an IP address [46]. This means that anybody can now find this mapping that was only known by ISPs or big companies (like Google and Facebook), but never communicated unless in case of a legal action. The privacy threat is thus very high because this mapping enables blackmail, social attacks, targeted phishing attacks, etc.

As a proof of concept, we show that it is possible to track VoIP users mobility and BitTorrent downloads [46] using Skype, one of the most popular VoIP system with more that 500 millions registered users.

All these works received a very large media coverage (see [http://www-sop.inria.fr/members/Arnaud.Legout/Projects/bluebear.html](http://www-sop.inria.fr/members/Arnaud.Legout/Projects/bluebear.html)).
6.3. Network measurement, modeling and understanding


The main objective of our work in this domain is a better monitoring of the Internet and a better control of its resources. We work on new measurement techniques that scale with the fast increase in Internet traffic and growth of its size. We propose solutions for a fast and accurate identification of Internet traffic based on packet size statistics and host profiles. Within the ECODE FP7 project, we work on a network-wide monitoring architecture that, given a measurement task to perform, tune the monitors inside the network optimally so as to maximize the accuracy of the measurement results while limiting the overhead resulting from collected traffic. Within the ANR CMON project, we work on monitoring the quality of the Internet access by end-to-end probes, and on the detection and troubleshooting of network problems by collaboration among end users.

Next, is a sketch of our main contributions in this area.

- **Internet traffic classification by means of packet level statistics**
  One of the most important challenges for network administrators is the identification of applications behind the Internet traffic. This identification serves for many purposes as in network security, traffic engineering and monitoring. The classical methods based on standard port numbers or deep packet inspection are unfortunately becoming less and less efficient because of encryption and the utilization of non standard ports. In this activity, we come up with an online iterative probabilistic method that identifies applications quickly and accurately by only using the size of packets. Our method associates a configurable confidence level to the port number carried in the transport header and is able to consider a variable number of packets at the beginning of a flow. By verification on real traces we observe that even in the case of no confidence in the port number, a very high accuracy can be obtained for well known applications after few packets were examined. In another work [39], we make a complete study about the inter-packet time to prove that it is also a valuable information for the classification of Internet traffic. We discuss how to isolate the noise due to the network conditions and extract the time generated by the application. We present a model to preprocess the inter-packet time and use the result as input to the learning process. We discuss an iterative approach for the on line identification of the applications and we evaluate our method on two different real traces. The results show that the inter-packet time is an important parameter to classify Internet traffic.

  We pursued this activity further by accounting for the communication profiles of hosts for the purpose of a better traffic classification [39], [38], [40]. We use the packet size and the inter-packet time as the main features for the classification and we benefit from the traffic profile of the host (i.e. which application and how much) to refine the classification and decide in favor of this or that application. The host profile is then updated online based on the result of the classification of previous flows originated by or addressed to the same host. We evaluate our method on real traces using several applications. The results show that leveraging the traffic pattern of the host ameliorates the performance of statistical methods. They also prove the capacity of our solution to derive profiles for the traffic of Internet hosts and to identify the services they provide.

  For a more thorough study of the traffic classification problem by means of packet statistics and host profiles, we refer to the PhD dissertation of Mohamad Jaber who was the main contributor to this activity inside the EPI Planete.

- **Adaptive network-wide traffic monitoring**
  The remarkable growth of the Internet infrastructure and the increasing heterogeneity of applications and users’ behavior make more complex the manageability and monitoring of ISP networks and raises the cost of any new deployment. The main consequence of this trend is an inherent disagreement between existing monitoring solutions and the increasing needs of management applications.
In this context, we work on the design of an adaptive centralized architecture that provides visibility over the entire network through a network-wide cognitive monitoring system. Given a measurement task, the proposed system drives its own configuration, typically the packet and flow sampling rates in routers, in order to address the tradeoff between monitoring constraints (processing and memory cost, collected data) and measurement task requirements (accuracy, flexibility, scalability). We motivate our architecture with an accounting application: estimating the number of packets per flow, where the flow can be defined in different ways to satisfy different objectives (e.g., Domain-to-Domain traffic, all traffic originated from a domain, destined to a domain). The architecture and the algorithms behind it are explained in paper published in 2010 for the case of a proactive control and in [43] for the case of a reactive control. In [44] the architecture and its algorithms are specified to a flow counting application. In all these works, the performances of our architecture are being validated in typical scenarios over an experimental platform we developed for the purpose of the study. Our platform is called MonLab (Monitoring Lab) and is described with more details in the Section on produced softwares. For now, MonLab presents a new approach for the emulation of Internet traffic and for its monitoring across the different routers. It puts at the disposal of users a real traffic emulation service coupled to a set of libraries and tools capable of Cisco NetFlow data export and collection, the overall destined to run advanced applications for network-wide traffic monitoring and optimization.

The activities in this direction are funded by the ECODE FP7 STREP project (Sep. 2008 - Dec. 2011). The dissertation of Imed Lassoued [21] provides an introduction to the field in addition to details on our contributions and the MonLab emulation platform.

- **Spectral analysis of packet sampled traffic**

In network measurement systems, packet sampling techniques are usually adopted to reduce the overall amount of data to collect and process. Being based on a subset of packets, they hence introduce estimation errors that have to be properly counteracted by a fine tuning of the sampling strategy and sophisticated inversion methods. This problem has been deeply investigated in the literature with particular attention to the statistical properties of packet sampling and the recovery of the original network measurements. Herein, we propose a novel approach to predict the energy of the sampling error on the real time traffic volume estimation, based on a spectral analysis in the frequency domain. We start by demonstrating that errors due to packet sampling can be modeled as an aliasing effect in the frequency domain. Then, we exploit this theoretical finding to derive closed-form expressions for the Signal-to-Noise Ratio (SNR), able to predict the distortion of traffic volume estimates over time. The accuracy of the proposed SNR metric is validated by means of real packet traces. The analysis and the expressions of the SNR that stemmed from are described in [26]. In [52], we adopt such a model to design a real-time algorithm, that sets the IPFIX counter export timers in order to grant, to each flow, a target estimation accuracy. The work within this direction has been partially supported by the FP7 ECODE project.

- **Monitoring the quality of the Internet access by end-to-end probes**

The detection of anomalous links and traffic is important to manage the state of the network. Existing techniques focus on detecting the anomalies but little attention has been devoted to quantify to which extent network anomaly affects the end user access link experience. We refer to this aspect as the local seriousness of the anomaly. In order to quantify the local seriousness of an anomaly, we consider the percentage of affected destinations, that we call the impact factor. In order to measure it, a host should monitor all possible routes to detect any variation in performance, but this is not practical in reality. In this activity, funded by the ANR CMON project, we work on finding estimates for the impact factor and the local seriousness of network anomalies through a limited set of measurements to random nodes we call landmarks.
We initially study the user access network to understand the typical features of its connectivity tree. Then, we define an unbiased estimator for the local seriousness of the anomaly and a framework to achieve three main results: (i) the computation of the minimum number of paths to monitor, so that the estimator can achieve a given significance level, (ii) the localization of the anomaly in terms of hop distance from the local user, and (iii) the optimal selection of landmarks. We are using real data to evaluate in practice the local seriousness of the anomaly and to determine the sufficient number of landmarks to select randomly without knowing anything on the Internet topology. The localization mechanism leverages the study on the connectivity tree and the relationship between the impact factor and the minimum hop distance of an anomaly. Our first results show that the impact factor is indeed a meaningful metric to evaluate the quality of Internet access. The current work focuses on extending this solution towards a collaborative setting where different end users collaborate together by exchanging the results of their observations. The objective will be a better estimation of the impact factor by each of them and a finer localization of the origin of any network problem.

On the experimental side, we have implemented the solution in a tool called ACQUA, which stands for Application for Collaborative Estimation of QUality of Internet Access \(^2\). We design an anomaly detection mechanism based on the histogram of delay measurements and the likelihood of observations. Then, we give to ACQUA a pipeline based software architecture, and we go deeply into experimentation inside and outside Planetlab. We show what the properties and usage of the algorithm are, focusing also on how this tool can help us to get information about the network anomalies detected. Later we extend the idea of Impact Factor Estimation (IFE) by using what we call Inverse IFE from Planetlab, where the computer of the user whose connectivity is tested has a completely passive role in the measurements procedure. We study its strong and weak points, and we show conditions under which Inverse IFE from Planetlab gives similar results to traditional IFE.

- **Applied Internet Measurements**

  The performance of several Internet applications often relies on the measurability of path similarity between different participants. In particular, the performance of content distribution networks mainly relies on the awareness of content sources topology information. It is commonly admitted nowadays that, in order to ensure either path redundancy or efficient content replication, topological similarities between sources is evaluated by exchanging raw traceroute data, and by a hop by hop comparison of the IP topology observed from the sources to the several hundred or thousands of destinations. In this project, based on real data we collected, we advocate that path similarity comparisons between different Internet entities can be much simplified using lossy coding techniques, such as Bloom filters, to exchange compressed topology information. The technique we introduce to evaluate path similarity enforces both scalability and data confidentiality while maintaining a high level of accuracy. In addition, we demonstrate that our technique is scalable as it requires a small amount of active probing and is not targets dependent. This work has been published in \(^{25}\).

- **Reliability of Geolocation Databases**

  In this project, we question the reliability of geolocation databases, the most widely used technique for IP geolocation. It consists in building a database to keep the mapping between IP blocks and a geographic location. Several databases are available and are frequently used by many services and web sites in the Internet. Contrary to widespread belief, geolocation databases are far from being as reliable as they claim. We conduct a comparison of several current geolocation databases -both commercial and free- to have an insight of the limitations in their usability. First, the vast majority of entries in the databases refer only to a few popular countries (e.g., U.S.). This creates an imbalance

\(^2\) [http://planete.inria.fr/acqua/](http://planete.inria.fr/acqua/)
in the representation of countries across the IP blocks of the databases. Second, these entries do not reflect the original allocation of IP blocks, nor BGP announcements. In addition, we quantify the accuracy of geolocation databases on a large European ISP based on ground truth information. This is the first study using a ground truth showing that the overly fine granularity of database entries makes their accuracy worse, not better. Geolocation databases can claim country-level accuracy, but certainly not city-level. This study has been published in CCR [28].

- **Impact of Live Streaming Traffic**

Video streaming is the most popular traffic in the Internet and a strong case for content centric networks. Therefore, it is fundamental to understand the network traffic characteristics of video streaming. In this work [49], we extensively studied the network traffic characteristics of YouTube and Netflix (the most popular video streaming traffic in the USA). We have shown that the traffic characteristics vastly depends on the type of browser, mobile application, and container (Flash, Silverlight, HTLM5) used.

6.4. Experimental Environment for Future Internet Architecture

**Participants:** Walid Dabbous, Thierry Parmentelat, Baris Metin, Frédéric Urbani, Daniel Camara, Alina Quereilhac, Shafqat Ur-Rehman, Thierry Turletti, Julien Tribino.

- **SFA Federation of experimental testbeds**

The OneLab2 project has come to its end in spring 2010. We are now involved in the NOVI (E.U. STREP) project, the F-Lab (French A.N.R.) project, and have the lead of the “Federation” WorkPackage of OpenLab (E.U. IP) project. Within these frameworks, we are codevelopping with Princeton University a reference implementation for the Testbed-Federation architecture known as SFA for Slice-based Federation Architecture. As a sequel of former activities we also keep a low-noise maintenance activity of the PlanetLab software, which has been running in particular on the PlanetLab global testbed since 2004, with an ad-hoc federated model in place between PlanetLab Central (hosted by Princeton University) and PlanetLab Europe (hosted at INRIA) since 2007.

During 2011 we have focused on the maturation of the SFA codebase, with several objectives in mind. Firstly we have contributed to a major overhaul of the specification as defined essentially within the GENI (N.S.F.) Project, with participations from all over the world. These changes, that affected both the core API and the schema used to expose and manage resource specifications, aimed at reaching a mature level of interoperability between the PlanetLab world and the EmuLab a.k.a. ProtoGeni world that has its own implementation, and are now available in SFA-2.0 issued late 2011.

Secondly, the SFA codebase has been redesigned to provide a more generic shelter that other testbeds can easily leverage in order to come up with their own SFA-compliant wrapper. This is perceived as a powerful means to foster further adoption of the architecture, and the Planète team has been instrumenal in bringing two entirely different testbeds to the federation, namely Senslab - developped in other INRIA Project-teams - and FEDERICA, the outcome of another E.U.-funded Project. Along the same lines we are working, although more remotely, with NICTA in Australia that publishes the O.M.F. testbed for running wireless testbeds, and who are interested in adopting this federation paradigm.

Finally, as part of the pure PlanetLab development, we have added a feature for running nodes in a ‘reservable’ mode, which breaks the usual best-effort PlanetLab model, but turned out very helpful both for making experiments possible, that needed a more reproducible behaviour of experiments, and also in a federation perspective, for closing the usage gap with, notably wireless testbeds, that typically have a reservable-only provisioning mechanisms.
• Content Centric Networks Simulation

We worked this year on the extension of the DCE framework for ns-3 in order to run CCN implementation under the ns-3 simulator. DCE stands for Direct Code Execution, its goal is to execute unmodified C/C++ binaries under ns-3 network simulator. With this tool researchers and developers can use the same code to do simulation and real experiments. DCE operation principle is to catch the standard systems calls done by the real application in the experiment and to emulate them within the ns-3 virtual network topology. Concerning CCN we use the PARC implementation named CCNx which is a well working open source software reference implementation of Content Centric Network protocol. As promised by DCE this integration of CCNx requires no modification of its code, it requires ‘only’ working on adding the system calls used by CCN that are not already supported by DCE. The advantage of this approach is that the integration work of CCN advanced DCE and will be useful in others completely different experiments. Another great advantage is that every evolution of the CCNx implementation is very easy to integrate, all what is needed is to compile the new source code. The next steps will be naturally to use DCE/ns-3 to evaluation CCN protocols in specific scenarios, to improve the coverage of systems calls supported by DCE, and to improve the DCE scheduler to be more realistic and to take into account CPU time spent in router queues. This work is done in the context of the ANR CONNECT project.

• ns-3 Module store

Bake is an integration tool which is used by software developers to automate the reproducible build of a number of projects which depend on each other and which might be developed, and hosted by unrelated parties. This software is being developed with the participation of the Planète group and is intended to be the automatic building tool adopted by the ns-3 project.

The client version of Bake is already working and the Planète group had a significant participation in its development. The contributions were in the context the addition of new functionalities, bug fixing and in the development of the regression tests. We are now starting the development of the ns-3 modules repository, which is a web portal to store the meta-information of the available modules. In the present state we have already designed and implemented the portal data basis and the main interface. It is already possible to register new modules and browse among the already registered ones.

The web portal has to be finished, notably the part that will create the xml file that will be used to feed the bake’s client. We also need to add new functionalities to the client part, to enable incremental build over partially deployed environments. As it is today, bake does not enable the user to add just one new module to an already deployed version of the ns-3 simulator. This work is done in the context of the ADT MobSim in collaboration with Hipercom and Swing Inria project-teams.

• The ns-3 consortium

We have founded last year a consortium between INRIA and University of Washington. The goals of this consortium are to (1) provide a point of contact between industrial members and the ns-3 project, to enable them to provide suggestions and feedback about technical aspects, (2) guarantee maintenance of ns-3’s core, organize public events in relation to ns-3, such as users’ day and workshops and (3) provide a public face that is not directly a part of INRIA or NSF by managing the http://www.nsnam.org web site. This web site is now finalized. However, activities related to developing the consortium have slowed down during 2011 due to the leave of Mathieu Lacage. We plan to put more resources on this aspect in 2012.
• Using Independent Simulators, Emulators, and Testbeds for Easy Experimentation

Evaluating new network protocols, applications, and architectures uses many kinds of experimentation environments: simulators, emulators, testbeds, and sometimes, combinations of these. As the functionality and complexity of these tools increases, mastering and efficiently using each of them is becoming increasingly difficult.

We designed the preliminary prototype of the Network Experiment Programming Interface (NEPI) whose goal is to make easier the use of different experimentation environments, and switch among them easily. NEPI intends to make it possible to write a single script to control every aspect of a potentially mixed experiment, including a hierarchical network topology description, application-level setup, deployment, monitoring, trace setup, and trace collection. We showed how a single object model which encompasses every aspect of a typical experimentation workflow can be used to completely describe experiments to be run within very different experimentation environments.

The development of NEPI started in 2009 with the implementation of the core API, an address allocator, a routing table configurator, but also a prototype ns-3 backend driven by a simple graphical user interface based on QT. Last year, we validated and evolved the core API with the addition of a new backend based on linux network namespace containers and stabilized the existing ns-3 backend. This year we have enhanced the design of NEPI and provided experiment validation, distributed experiment control, and failure recovery functionalities. In particular, we enforced separation between experiment design and execution stages, with off-line experiment validation. We also introduced a hierarchical distributed monitoring scheme to control experiment execution. We implemented a stateless message-based communication scheme, and added failure recovery mechanisms to improve robustness. The NEPI approach has been validated by implementing support for three complementary environments: a physical testbed, a network emulator, and a network simulator. Furthermore, we showed with a concrete experiment use case, available online for reproduction, how easy it is with NEPI to integrate these environments for hybrid-experimentation [48].

• Guidelines for the accurate design of empirical studies in wireless networks

Traditionally, wireless protocol proposals have been often tested and validated using only analytical and simulation models [73]. However, as the wireless environment is very complex to model accurately, and since the cost of wireless cards has decreased in an exponential way, today more and more research papers include evaluation of new proposals using experimentation on real devices. Indeed, experimentation is a mandatory step before possible deployment of new network protocols with real users. However, wireless experimentation is much more complex to set up and run than simulation, and it is important to avoid many pitfalls that can occur during experimentation. The objectives of this work are twofold. First, we described typical problems currently encountered in wireless-based experimentation, and we present simple guidelines to avoid them [50]. Second, we proposed an experimental methodology where the detection of anomalies, calibration of the measurement setup, and clear definition of the scenario (among others) make easier the repeatability of results [55]. This work has been done in collaboration with Cristian Tala, Luciano Ahumada and Diego Dujovne from the Universidad Diego Portales of Chili, in the context of the WELCOME STIC AMSud 2011.

• Multicast Video Streaming over WiFi Networks: Impact of Multipath Fading and Interference
We conducted an experimental study in order to analyse the impact of interference, multipath fading and path loss on multicast video streaming (i.e., goodput, packet loss and ultimately on the video quality) using off-the-shelf fixed WiFi equipment in a wireless (802.11 b/g) local area network (WLAN) environment. We used the ricean K-factor as a measure of multipath fading, spectrum analyzer to estimate channel interference and received signal strength indicator (RSSI) as indication of signal power and attenuation. In order to realistically measure aforementioned metrics, we conducted extensive wireless experiments against six test cases representing common real-world situations using off-the-shelf wireless equipment.

We showed that interference has more impact on performance than multipath fading. Multipath fading can result in considerable performance degradation in environments where moving objects cause perturbance. On the contrary, channel interference is more frequent and more prominent cause of performance degradation in wireless networks because ISM 2.4 GHz band is increasingly being utilized in homes and work places. Being able to quantify the impact of multipath fading and interference is crucial in planning, troubleshooting, managing as well as benchmarking and optimizing wireless networks. This study has been published in MediaWin 2011 [51].

- **Making easier Experimentation**

Wireless experimentations are challenging to evaluate due to the high variability of the channel characteristics and its sensitivity to interferences.

Merging traces represents a complex problem especially in wireless experimentations, due to packet redundancy in multiple probes. Merging traces solutions need to be efficient in order to process the large amount of generated traces. These solutions should provide an output data structure that allows easy and fast analysis and must be scalable in order to be used in large and various experimental settings. We have designed an algorithm that performs trace synchronization and merging in a scalable way. The algorithm output is stored in a configured MYSQL database allowing for smart packet trace storage. This solution reduces processing time by 400% and storage space by 200% with regard to raw trace file solutions. It has been implemented in an open source software called CrunchXML, available under the GNU General Public License v2 at [http://twiki-sop.inria.fr/twiki/bin/view/Projets/Planete/CrunchXML](http://twiki-sop.inria.fr/twiki/bin/view/Projets/Planete/CrunchXML).

- **An Integration Framework for Network Experimentation**

Many different experimentation environments address complementary aspects of network protocol evaluation, but because of their disparities and complexities it is often hard to use them to reproduce the same experiment scenario.

Simulation is often used for the evaluation of new network protocols and architectures. In order to perform more realistic simulations, modern simulators such as ns-3 integrate more detailed models and even support direct execution of real protocol code. However, such complex models require more computational and memory requirements. We have studied the feasibility of a hybrid approach based on distributing a complex simulation scenario on several nodes in a grid network. We showed that by exploiting the real time operation of the ns-3 simulator, it is possible to map such complex scenarios on grid nodes. We also proposed a basic mapping algorithm to distribute a simulation scenario in several nodes [42].
ARIANA Project-Team

6. New Results

6.1. Optical imagery for remote sensing

6.1.1. Phase Field-Higher Order Active Contours for Object Modelling and Image Segmentation

Participants: Ikhlef Bechar, Josiane Zerubia [contact].

This work is done in collaboration with Dr Ian Jermyn of Durham University (United Kingdom) and was funded by a contract with the EADS foundation [http://www-sop.inria.fr/ariana/Ikhlef.Bechar/hoacs/index.html].

The problem of object segmentation from imagery is an essential preliminary task for many applications (target recognition, automated navigation, organ segmentation in medical imaging, etc). The problem of adding prior knowledge about objects to the image segmentation process has received a lot of interest since recently, and active contours [38] provide us with such a tool.

We consider a new class of active contours called Higher Order Active Contours (HOACs) introduced initially in [41], and which consider an optimal contour as the one which minimizes an energy involving three additive terms; namely the length of a contour, its area and a term of interaction between all possible pairs of its points via an interaction function $\Psi(t)$. The three terms being weighed by means of three scalar parameters. The main advantage of HOACs over traditional segmentation methods is that they offer an unprecedented means for including shape prior about an object via the interaction function $\Psi(t)$.

The HOAC set up has been applied successfully to various object extraction problems such as the extraction of networks [40], circular shapes [37], etc, using a specific family of the $\Psi(t)$'s. Our main task in the framework of this project is to extend their work to more general shapes.

Our contributions so far have been in the numerical computation of the optimal HOAC parameters for a given shape. We have shown indeed that the HOAC energy can be fully made linear with respect to $\Psi(t)$, which then makes it easy to solve for $\Psi(t)$ numerically (cf. Fig. 6). This is achieved by first choosing a linear basis to represent $\Psi(t)$ and using K-K-T (Karush-Kuhn-Tucker) optimality criteria to express the fact that a target contour is a local minimum of the energy. Consequently, looking for the optimal values of the coefficients for a given shape amounts first to solve an eigen value problem, and second, to find the linear combinations of the found eigen vectors that satisfy both K-K-T minimality criteria. The computation being carried out in the Fourier domain for sake of computational efficiency. We are currently testing the proposed model on simple shapes such as the butterfly one shown in figure 6.

6.1.2. Optimization of the compression-restoration chain for satellite images

Participants: Mikael Carlavan, Laure Blanc-Féraud [contact].

This project involves the French Space Agency (CNES) and the CESBIO, on collaboration with TAS and I3S (Marc Antonini).

This work concerns the study of the optimal imaging chain in the context of satellite imaging. The main goal of this study is to propose a new method to address the problem of decoding-deconvolution-denoising and consists in a characterization and optimization of the compression/restoration processes considering the instrumental characteristics (FTM, noise, sampling). A theoretical study first showed that current processes of compression and restoration are better fitted if the restoration is performed on-board before the compression. Indeed, current restoration algorithm is designed to remove the blur and the instrumental noise but does not take into account the coding noise, and it is well-known that compression algorithms do not perform
Figure 6. An example of estimation of the $\Psi$ function for a given shape. (a) An example of a butterfly-like shape; (b) The numerical estimation of its optimal $\Psi$ function; (c) The result of a gradient descent algorithm on the HOAC energy with $\Psi$ of figure (b) until convergence (in black the original shape, and in red the shape found by the gradient descent algorithm). One can see that the estimated $\Psi$ makes the HOAC energy achieve a local minimum at a shape which is very close to the target shape of figure (a).
properly on noisy data. More generally, we concluded that the image should be the closest possible to the real image before the step of coding, encouraging, thus, to move the restoration step on-board before the compression. Figure 7 shows the global distortion w.r.t. the coding rate if the restoration step is done either before (on-board) or after (on-ground) the compression. We see that using an on-board restoration leads to a quality improvement of the final image regardless the coding rate. For example, at the usual coding rate of 2.5 bits/pixel, using an on-board restoration improves the quality of the final image about 0.5 dB.

![Figure 7. Global distortion w.r.t coding rate. The solid line is the global distortion with the restoration step done on-ground after the compression and the dashed line is the global distortion if the restoration is performed on-board before the compression.](image)

### 6.1.3. Aerial Image Restoration

**Participants:** Daniele Graziani, Laure Blanc-Féraud [contact].

*This project involves as partners: ATE, Coreti, and Gilles Aubert from the J.A.D. Laboratory at the University of Nice Sophia Antipolis.*

The goal of the project is to built an airborne camera system, and our part is to process aerial images provided by ATE: restoration, microscanning, video, color images. We investigate a convex variational framework to compute high resolution images from a low resolution video. We analyze the image formation process to provide a well designed model for warping, blurring, downsampling and restoration. The microscanning is modeled as a convex minimization problem, which is solved with a domain decomposition technique based on the recent work of M. Fornasier, A. Langer and C. Schonlieb. ("A convergent overlapping decomposition method for total variation Minimization", Numerische Math. to appear), which allows parallel computing and a realization of a real time algorithm.

### 6.1.4. Contribution of object recognition on forest canopy images to the building of an allometric theory for trees and natural, heterogeneous forests

**Participants:** Jia Zhou, Xavier Descombes, Josiane Zerubia [contact].

*This work is done in collaboration with Dr. Pierre Couteron and Christophe Proisy at IRD, UMR AMAP, Montpellier.*
Individual tree detection methods are more and more present, and improve, in forestry and silviculture domains with the increasing availability of satellite metric imagery. Automatic detection on these very high spatial resolution images aims to determine the tree positions and crown sizes. The mathematical model based on marked point processes has showed advantages w.r.t. several individual tree detection algorithms for plantations. We used this detection method to analyze natural mangrove forests in French Guiana, eucalyptus plantations in Brazil, and other types of tropical forests. The simulated optical images were also used to improve the method and calibrate the detection parameters. To analyze a eucalyptus plantation in Brazil [23], we used 2 optical images acquired by the WorldView-2 satellite. A tentative detection simultaneously with 2 images of different dates (multi-date) was tested for the first time, which estimates individual tree crown variation during these dates. In this work, we tried to find the trees localizations and crown sizes in order to provide a plantation map, and estimate the tree crown growth during the period between 2 images, and compared these results with the field measurements and expected dynamics of corresponding populations. An example of multi-date detection result is showed in figure 9.

The detection method was also applied on simulated optical DART (Discrete Anisotropic Radiative Transfer) images, where exact field inventory could be provided on large surfaces. We assessed the detection results with these “ground-truth” maps.

6.2. SAR imagery for remote sensing

6.2.1. Stochastic modeling for very high resolution SAR image processing

Participants: Aurélie Voisin, Vladimir Krylov, Josiane Zerubia [contact].

This work is done in collaboration with DIBE, University of Genoa, with Dr Gabriele Moser and Prof. Sebastiano B. Serpico [http://spt.dibe.unige.it/] with partial financial support of the French Defense Agency, DGA [http://www.defense.gouv.fr/dga/] and the Italian Space Agency, ISA [http://www.asi.it/en/]. The data are provided by the Italian Space Agency, ISA [http://www.asi.it/en/].
We deal with the environmental risk assessment by addressing the problem of classifying SAR images of urban areas. Several difficulties need to be considered to address the SAR classification problem. The first one is related to the inherent multiplicative noise known as speckle, which degrades appreciably the registered imagery. Another difficulty is the heterogeneity of urban areas on very high resolution (VHR) images that leads to heterogeneous statistical modeling, reflecting the different ground materials such as asphalt, concrete, metal, etc. We propose a hierarchical statistical Bayesian supervised classification approach that consists of two steps. The first step deals with the SAR amplitude statistical modeling for each target class (e.g. vegetation, urban, etc.) by using a finite mixture model, estimated by resorting to a dictionary-based stochastic expectation maximization (DSEM) algorithm. More specifically, the SAR amplitude probability density functions (PDFs) are assumed to be mixtures of $K$ PDFs automatically chosen inside a predefined dictionary of SAR-specific distribution families. Such mixtures are intended to take into account the above mentioned VHR SAR statistics heterogeneity. We further consider an additional source of information obtained by extracting a textural feature map from the original SAR image in order to optimize the detection of urban areas. Typically, the textural feature is generated by using a Grey Level Co-occurrence Matrix (GLCM)-based method. The marginal PDFs of the original SAR image and the textural feature are combined via copulas, leading to a joint PDF for each class. On the second step the classification map is generated, using the joint copula-based statistics. To improve the robustness with respect to speckle noise, we consider a contextual model based on Markov random fields (MRFs), and, more specifically, a hierarchical MRF, which offers the possibility to take into account the multi-scale information and to deal with multi-resolution imagery [28]. A variety of algorithms were proposed to estimate the labels on hierarchical graphs. The consideration of a specific graph, here a quad-tree, allows to benefit from its good properties (e.g. causality) and to apply non iterative algorithms. Among the different algorithms employed in the literature, we chose to take into account an exact estimator of the marginal posterior mode (MPM). The cost function associated to this estimator offers the possibility to penalize the errors according to their number and the scale at which they occur: an error at the coarsest scale is stronger penalized than an error at the finest scale. Moreover, we introduce a prior estimation update that experimentally leads to improved results and is less affected by speckle noise when compared to a predefined prior [35]. The challenge of the problem considered here is that our given input is a single-polarized SAR image at a single resolution. To improve the classification, we extract an extra information in the form of a multi-scale wavelet decomposition from the initial image. Then, at each level, the textural feature map is obtained from each image in the decomposition stack. Finally, at each level, the wavelet image is combined with the textural image by using copulas, as described previously in Ariana activity reports. The hierarchical method was tested on real COSMO-SkyMed images. We illustrate the obtained results with an example of a SAR acquisition of the Port-au-Prince quay (Haiti). Spatially disjoint training and test areas were manually annotated. The classification is done following 3 classes: urban areas, natural landscape and wet areas. The
results are shown qualitatively in figure 10. The computation of numerical results gives an average accuracy of 95.65 percent for the considered test areas.

Figure 10. Left: Initial SAR image of Port-au-Prince (Haiti) (©ISA, 2009). Right: Classification map obtained with the hierarchical method for the 3 classes (Blue: water; Green: vegetation; Red: urban area).

Our previous work was based on single-scale MRF, thus the hierarchical approach is a direct extension. Part of our work was dedicated to the comparison of these two methods [35], [28], and we also compared the MRF-based model to a novel products of experts approach [18].

6.2.2. Parameter estimation procedures for HR SAR image classification

Participants: Vladimir Krylov, Josiane Zerubia [contact].

This work is conducted in collaboration with DIBE, University of Genoa with Dr. Gabriele Moser and Prof. Sebastiano Serpico [http://spt.dibe.unige.it/] with the support of the Italian Space Agency, ASI [http://www.asi.it/en].

Parameter estimation of probability density functions is one of the major steps in the mainframe of statistical image and signal processing. We have explored the properties and limitations of the recently proposed method of logarithmic cumulants (MoLC) parameter estimation approach which is an alternative to the classical maximum likelihood (ML) and method of moments (MoM) approaches. We have derived the general sufficient condition of strong consistency of MoLC estimates which represents an important asymptotic property of any statistical estimator [33]. We have demonstrated the strong consistency of MoLC estimates for a selection of widely used distribution families originating (but not restricted to) synthetic aperture radar (SAR) image processing. We have then derived the analytical conditions of applicability of MoLC to samples generated from several distribution families in our selection. We have conducted various synthetic and real data experiments to assess the comparative properties, applicability and small sample performance of MoLC notably for the generalized gamma and $K$ family of distributions. The synthetic-data experiments have demonstrated a competitive accuracy of MoLC estimates and a reliable behavior of this estimator for small samples which is a critical issue in applications. We have performed real-data image processing experiments to the problem of supervised classification applied to high resolution satellite SAR imagery. These experiments confirmed the stability of MoLC estimator with respect to sample size and at the same time illuminated the critical side of MoLC given by applicability restrictions. The experiments suggested the efficiency of use of the MoLC estimator for finite-mixture estimation problems [3], [19], and ML-based classification approaches [4], [28], [35].
6.2.3. Unsupervised amplitude and texture based classification of SAR images with multinomial latent model

**Participants:** Koray Kayabol, Aurélie Voisin, Vladimir Krylov, Josiane Zerubia [contact].

The participants would like to thank the Italian Space Agency (ASI) for providing the COSMO-SkyMed images. The TerraSAR-X images are provided from http://www.infoterra.de/.

We combine both amplitude and texture statistics of the Synthetic Aperture Radar (SAR) images using Products of Experts (PoE) approach for classification purpose. We use Nakagami density to model the class amplitudes and a non-Gaussian Markov Random Field (MRF) texture model with t-distributed regression error to model the textures of the classes. A non-stationary Multinomial Logistic (MnL) latent class label model is used as a mixture density to obtain spatially smooth class segments. The Classification Expectation-Maximization (CEM) algorithm is performed to estimate the class parameters and to classify the pixels [18]. Determining the necessary number of classes to represent the data and initialization are some drawbacks of the EM type algorithms. In [17] and [32], we combine hierarchical agglomeration, CEM and Integrated Classification Likelihood (ICL) criterion to get rid of the drawbacks of EM. We obtained some classification results of water, land and urban areas in both supervised and unsupervised cases on TerraSAR-X, as well as COSMO-SkyMed data [18], [17], [32]. The proposed unsupervised ATML-CEM (Amplitude and Texture density mixtures of MnL with CEM) method provides significantly better results, see Fig. 12, compared to the corresponding results obtained with K-MnL and its performance is close to supervised ATML-CEM.

6.3. 3D-modelling of urban scenes

6.3.1. Building reconstruction from aerial LiDAR data

**Participants:** Yannick Verdie, Florent Lafarge [contact], Josiane Zerubia.

The generation of 3D representations of urban environments from aerial and satellite data is a topic of growing interest in image processing and computer vision. Such environments are helpful in many fields including urban planning, wireless communications, disaster recovery, navigation aids, and computer games. Laser scans have become more popular than multiview aerial/satellite images thanks to the accuracy of their measurements and the decrease in the cost of their acquisition. In particular, full-waveform topographic LiDAR constitutes a new kind of laser technology providing interesting information for urban scene analysis. We study new
Figure 12. (a) The original SAR image (COSMO-SkyMed, ©ASI), (b), (c) and (d) classification maps obtained by K-MnL, supervised and unsupervised ATML-CEM methods. Blue, red and green colors represent water, urban and land areas, respectively.
stochastic models for analysing urban areas from LIDAR data. We aim to construct concrete solutions to both urban object classification (i.e. detecting buildings, vegetation, etc.) and the 3D reconstruction of these objects. Probabilistic tools are well adapted to handling such urban objects, which may differ significantly in terms of complexity, diversity, and density within the same scene. In particular, jump-diffusion based samplers offer interesting perspectives for modelling complex interactions between the various urban objects. We investigated a first approach aiming at producing accurate, watertight and compact meshes from planar patches under planar constraint especially designed for urban scenes. The LiDAR point cloud is classified through a non-convex energy minimization problem. The planar structures are extracted and connected to generate a compact, and watertight mesh of the building. Experiments highlight the potential of our method in term of performance, compactness, and accuracy. This work has been published in [22]. We illustrated our results in figure 13.

6.3.2. Modeling large urban environments from unstructured point clouds

Participant: Florent Lafarge [contact].

We present a robust method for modeling cities from unstructured point data. Our algorithm provides a more complete description than existing approaches by reconstructing simultaneously buildings, trees and topologically complex grounds. Buildings are modeled by an original approach which guarantees a high generalization level while having semantized and compact representations. Geometric 3D-primitives such as planes, cylinders, spheres or cones describe regular roof sections, and are combined with mesh-patches that represent irregular roof components. The various urban components interact through a non-convex energy minimization problem in which they are propagated under arrangement constraints over a planimetric map. We experimentally validate the approach on complex urban structures and large urban scenes of millions of points as illustrated on Figure 14.

6.3.3. Parallel Monte Carlo sampler for point processes

Participants: Yannick Verdie, Florent Lafarge [contact], Ioan Dragan.

We designed a new parallel scheme for Markov point processes. These probabilistic models exploit random variables whose realizations are configurations of parametric objects, each object being assigned to a point positioned in the scene. The number of objects is itself a random variable. Another strength of Markov point processes is their ability to take into account complex spatial interactions between the objects and to impose global regularization constraints. Moreover, we proposed to use space-partitioning tree such as quadtree (for 2D data) or octree (for 3D data) for non-homogeneous measure adapted to the problem. We illustrate the results in figure 15.
Figure 14. Reconstruction of the cities of Marseille and Amiens, France, from Lidar point cloud.

Figure 15. Results of the new Marked Point process on (left) 2D data, and (right) LiDAR data.
6.4. Biological imagery

6.4.1. Regularizing parameter estimation with Poisson noise

Participants: Mikael Carlavan, Laure Blanc-Féraud [contact].

The problem is to automatically estimate the regularizing parameter in Poisson noisy image deconvolution using the $L_1$-norm regularization as a total variation or frame coefficients. This problem is addressed using the discrepancy principle. The standard weighted criterion composed of a data term and a regularization term is rewritten as a constrained minimization problem. The constraint is designed on the data term using the discrepancy principle and a new estimation of the bound is proposed as well as an efficient algorithm to solve this constrained minimization problem. This work is published in [10], [24], [9].

6.4.2. Brain vascular network segmentation

Participant: Xavier Descombes [contact].

This work was conducted in collaboration with Franck Plouraboué and Abdelhakim El Boustani from IMFT Toulouse and Caroline Fonta from CerCo Toulouse. It has been partially supported by a PEP II project from CNRS.

Micro-tomography produces high resolution images of biological structures such as vascular networks. We have proposed a new approach for segmenting vascular network into pathological and normal regions from considering their micro-vessel 3D structure only. We consider a partition of the volume obtained by a watershed algorithm based on the distance from the nearest vessel. Each region is characterized by its volume and the local vascular density. The volume and density maps are first regularized by minimizing the total variation. Then, a new approach has been proposed to segment the volume from the two previous restored images based on hypothesis testing. Results are presented on 3D micro-tomographic images of the brain micro-vascular network (see Fig. 16).

![Figure 16. Brain micro-vascular network segmentation](image)

6.4.3. Blind restoration of 3D biological image

Participants: Saima Ben Hadj, Laure Blanc-Féraud [contact].
Fluorescence microscopy is a powerful imaging technique providing three-dimensional images of biological living specimen. However, these images are degraded by a depth variant blur due to light diffraction phenomenon as well as refractive index mismatch between the different mediums composing the system and the biological sample. They are also distorted with noise from non-ideal imaging conditions. In order to provide biologist with more suitable images for quantitative studies, many restoration methods were developed. In most of them, the blur function, called Point Spread Function (PSF) is assumed to be piecewise constant in order to avoid the intensive computing time when using a pointwise varying PSF. However, this usually leads to blocking effect in the restored image. In our work, we extend the Space varying (SV) blur model previously proposed in [39] for 2D astronomical images to 3D microscopy images. In that model, the degraded image is a convex combination of convolutions with a space-invariant (SI) PSF. Furthermore, we fit to that model two restoration procedures which are basically developed for a SI PSF. On the one hand, we use the Richardson-Lucy method with Total Variation regularization which is carried out under Poisson noise assumption in order to restore confocal microscopy images. We employ another method with total variation regularization adapted to images with an additive Gaussian noise in order to restore Wide Field Microscopy images. For that, we rely on a fast optimization method based on a domain-decomposition technique [36]. In particular, we study its convergence properties when using the SV blur model [31],[27]. To illustrate the interest of the proposed method, we show in Fig. 17 some results obtained on a simulated bead image of Wide Field Microscopy.

![Image](a) (b) (c) (d)

Figure 17. (X,Z) slices of the (a) original image, (b) degraded image (c) restored image using a SV blur model, and (d) restored image with a space-invariant PSF.

### 6.4.4. Axon imaging

**Participant:** Florence Besse [contact].

During brain development, neurons extend cellular processes (dendrites and axons) to connect to specific targets and establish functional networks. Understanding how axonal processes migrate to reach their targets and how they form new branches to build up a complex axonal tree is thus key. To characterize the properties of axonal trees, we have generated a collection of confocal 3D pictures of normal and mutant single axons labeled using a fluorescent protein. To analyze the formation of axonal trees in real-time, we have developed a protocol to dynamically image growing axons within intact Drosophila brains. In this protocol, entire brains are cultured in conditions where they can undergo cell differentiation and maturation. Axons are labeled by the fluorescent molecule GFP and are imaged over 12h, with very low-photobleaching and no associated phototoxicity, using an ultra-sensitive 2-photon microscope. Several image sequences corresponding to the growth of axons in normal conditions have been acquired (see Fig. 18).

### 6.4.5. Detection of Axons in Neuronal Images

**Participants:** Alejandro Mottini, Florence Besse, Xavier Descombes [contact].
Imaging techniques such as confocal and two-photon microscopy provide an efficient way of analyzing supra-cellular structures. It is known that the precise shape of these objects provides information on their functioning and allows the characterization of pathological states. Therefore, the analysis of the morphological differences between normal and pathological structures is of paramount importance. In particular, the analysis of neuronal axon topologies allows biologists to study the causes of neurological diseases such as Fragile X Syndrome and Alzheimer’s disease. For this purpose biologists have acquired static 3D images of mature neuron axons using fluorescent confocal microscopy.

Due to the high volume of generated image data and the tortuous nature of the axons, manual processing is infeasible. Therefore, it is necessary to develop techniques for the automatic extraction and analysis of the neuronal structures. However, since both types of images present different characteristics, two different methodologies need to be developed. The main objective of the static case is the study of the length and number of bifurcation points of the two populations of neurons. On the other hand, the focus in the dynamic
case is put on the tracking of the axonal tips. The automatic extraction of axons from confocal microscope images is a key problem in the field of neuron axon analysis. In this work we propose a solution which combines algorithms for the denoising, binarization, skeletonization, gap filling, point detection and statistical analysis in a pipeline capable of extracting the axons. Furthermore, these algorithms were incorporated in a software, developed in Matlab, which includes an easy to use GUI along with functionalities to facilitate high data throughput analysis. The software was evaluated on several 3D confocal microscope images of normal and mutated axons. Our results support the potential use of the software in helping biologist perform automatic morphological analysis of axons in fluorescent confocal microscopy images.

Figure 20. Original (left) and extracted (right) normal axon image (2D maximum intensity projection).

6.4.6. Axons tracking

Participants: Alejandro Mottini, Huei-Fang Yang, Florence Besse, Xavier Descombes [contact].

This work was funded by project ARC-DADA (INRIA/CNRS/UNSA) [ http://www-sop.inria.fr/members/Xavier.Descombes/DADA/home.html ]. It is done in collaboration with Serpico team at INRIA Bretagne (C. Kervrann, P. Houllier)

To study axon growth process biologists have acquired dynamic 3D+t image sequences of developing neurons using fluorescent two-photon microscopy.

Live cell two-photon microscopy is an effective tool for the analysis of dynamical processes occurring in living samples that, when combined with fluorescence, allows the detection of objects of interest in 3D space and time. These labeled objects appear as bright spots which need to be detected. The low resolution and signal-to-noise-ratio (SNR) make this analysis difficult even for experienced biologists. As a consequence, automatic detection techniques need to be robust and flexible. To this end, the Marked Point Process (MPP) detection framework was selected. Since axonal extremities appear either as circular objects or as ending points of thin filaments, we proposed two different MPP models. These models were formulated using a Gibbs energy function and optimized with Multiple Births and Deaths, a newly proposed algorithm which guarantees a fast convergence to the global minimum. The first model is designed to detect spheres or disks and the second filaments (both in 3D). Both models were tested on several 3D static images. To evaluate the performance of the detection, results were compared against images labeled by an experienced biologist.

Figure 22 shows the results obtained on one frame of a video sequence (the shown image is a slice of the 3D frame) for the disk/sphere model. One can appreciate that the axonal tip marked by the biologist (in white, left) was detected. However, a false positive is also present.
Figure 21. Dynamic images (2D maximum intensity projection).

Figure 22. Original (left) and resulting (right) images (detected disks in pink). The true axonal tip is circled in white (original image) and red (result). In the original image, axons are labeled in white.
Figure 23 shows the result obtained on one of the slices (for the same video frame) for the filament model. Although the good filament was detected, many false positives are also present in the image. Similar results were obtained for other images.

![Figure 23. Original image (left) (targeted filament circled in red) and result (right) (detected filaments in red, circled in red).](image)

Once the detection of the extremities in each frame of the video has been solved, the solution should be embedded into a tracking algorithm to obtain an estimation of the axon’s trajectory during the growing stage. To this end, the particle filtering technique was considered. This technique consists in estimating the posterior distribution of the current state $x_t$ of the target of interest at time $t$ based on the measurements $z_{1..t}$:

$$p(x_t | z_{1..t}) \propto p(z_t | x_t) \int p(x_t | x_{t-1}) p(x_{t-1} | z_{1..t-1}) \, dx_{t-1},$$

where $p(x_t | x_{t-1})$ is the transition distribution, and $p(z_t | x_t)$ is the likelihood. In the current implementation, a simple kinematic model is used for the transition distribution, and a color histogram is applied to the computation of the likelihood. Figure 24 shows the preliminary results obtained by applying the particle filtering technique on the coronal KESM (Knife-Edge Scanning Microscopy) sections of the mouse cerebellum, where the green rectangles indicate the tracked objects.

![Figure 24. Tracking results on the coronal KESM sections of the mouse cerebellum. The green rectangles are the tracked objects. The user gives an initial starting region on the first image, and the algorithm tracks the region of interest in the subsequent images. Note that the images are cropped for a better view.](image)
The next step will be to design a more sophisticated transition distribution and likelihood model that are suitable for tracking the trajectories of axons during their growth.

6.5. Dermatology

6.5.1. Statistical analysis of skin pigmentation under treatment

Participants: Sylvain Prigent, Xavier Descombes, Josiane Zerubia [contact].

This work was partially funded by a contract with Galderma R&D [$http://www.galderma.com/R&D.aspx$].

One of the steps to evaluate the efficiency of a therapeutic solution is to perform measurements on a series of patients who received the studied treatment. In parallel another treatment is tested on another group of people or on the same group of patients on another skin area. This second treatment is the reference one for the studied pathology or a placebo. We will call it ‘vehicle’.

For facial hyper-pigmentation, for each studied treatment, a group of $N_e$ patients receive the treatment on one cheek and the vehicle on the other. To this end, patients are selected to have the same hyper-pigmentation severity on the two cheeks. Then multi-spectral images are taken at different times $t$ along the treatment period.

We propose a methodology to estimate the efficiency of a treatment by calculating a spectral criteria that maximizes the visibility of the disease comparatively to a healthy reference area. To design such a criterion, we compare three approaches. The first one gives weights to spectral bands in order to get an equivalent of the luminance from the $CIEL*a*b$ decomposition. This is the standard measure in dermatology. We compare this spectral measurement to a spectral signature obtained by ICA (Independent Components Analysis) in a whole study, and a criterion that searches for the highest contrasted band.

Once a criterion is designed, we compute the hyper-pigmentation severity of a patient by normalizing the pathological area with the healthy area and the active treatment measurement by the vehicle. That gives severity measurement distributions from patients at different times. A Student paired test allows to determine if an active treatment has an effect between two measurement times. The experiments done on 3 treatments and their associated vehicles brought to the following conclusion: The severity measure based on the selection of an optimal band allows to detect a treatment effect sooner than the two other approaches. The severity measurements obtained by the optimal band method is illustrated in figure 25.

Figure 25. Evolution of skin pigmentation in the time and its quantification.
6. New Results

6.1. Perception and autonomous navigation

Participants: Patrick Rives, Pascal Morin, Andrew Comport, Alexandre Chapoulie, Gabriela Gallegos, Cyril Joly, Maxime Meilland, Glauco Scandaroli.

6.1.1. Indoor SLAM: Self-calibration of the camera frame with respect to the odometry frame

Fusing visual data and odometry information is a standard technique to improve the robustness of the SLAM solution. Odometry data is considered as an input of the motion prediction equation while the visual data constitutes the filter observation. However, such method requires that the system is well calibrated: the pose of the camera frame with respect to the odometry frame has to be known. Usually, this pose is directly obtained by an hand made measurement yielding to incorrect values. We propose a new self calibration method to get these calibration parameters automatically. In practice, the state in the SLAM formulation is augmented with the unknown camera parameters (with respect to the odometry frame). This method requires to adapt a few Jacobians with respect to the original SLAM algorithm which assumes that these parameters are known. The accuracy and the stability of the estimation scheme clearly depends on the observability and the conditioning properties of the new system.

In 2010, we presented results in the case where the camera frame location has only 3 degrees of freedom (two translations and a rotation with respect to the vertical z axis). This year, these results were extended to the full calibration problem. As in the previous case, we assume that the robot is moving on a horizontal ground and observes 3D landmarks from the images delivered by the on board camera. The five parameters introduced by the calibration problem - 2 translations and 3 rotations (only 2 translations since the z component is not observable due to the planar motion of the robot) - are estimated simultaneously in addition to the "classical" SLAM parameters. The implementation of the algorithm is based on a Smoothing And Mapping (SAM) approach which computes a solution by considering the whole trajectory (instead of only the current pose as with the EKF approach).

As a theoretical result, we prove that this augmented system remains observable if and only if the curvature of the robot trajectory changes. This analysis was validated on real data with our indoor robot. Fig. 2 shows the mobile platform and the camera. It can be seen that an important rotational offset was added on the camera to test the capability to deal with large rotational values (the parameters are initialized with identity). Results are provided on Fig. 3 - 4 and table 1. They show that the trajectory and the map seem consistent; moreover, the algorithm was able to correct the odometry drift (green trajectory on Fig. 3). Then, the observability analysis was validated since the estimation of the camera frame parameters begins when there is a significant change in the radius of curvature of the trajectory (see the confidence bounds on Fig. 4). Finally, the estimation of these parameters was consistent with the ground truth (table 1). These results were presented at IROS’11 conference [30].

6.1.2. Outdoor Visual SLAM

Safe and autonomous navigation in complex outdoor urban-like environment requires a precise and real time localization of the robot. Standard methods, like odometry, typically performed by wheel encoders or inertial sensors, are prone to drift and not reliable for large displacements. Low cost GPS stations are inaccurate and satellite masking effect happens too frequently due to corridor-like configurations. We develop a real time and accurate localization method based on vision only without requiring any additional sensor.
Figure 2. Left: Robot used for the experiment – Right: The omnidirectional camera mounted with a rotational offset

Figure 3. Red: trajectory and map provided by the algorithm — Green: odometry integration — Black cross: end of the trajectory — Blue: 99% confidence region for the last robot position
Figure 4. Evolution of the curvature and the estimation of the camera parameters

Table 1. Numerical results concerning the camera parameters

<table>
<thead>
<tr>
<th>Camera param.</th>
<th>Reference</th>
<th>Estimation</th>
<th>3σ bounds</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_x$ (m)</td>
<td>-0.16</td>
<td>-0.149</td>
<td>[-0.165, -0.132]</td>
</tr>
<tr>
<td>$t_y$ (m)</td>
<td>-0.35</td>
<td>-0.38</td>
<td>[-0.064, -0.013]</td>
</tr>
<tr>
<td>$\gamma_x$ (deg)</td>
<td>6.05</td>
<td>6.609</td>
<td>[6.276, 6.942]</td>
</tr>
<tr>
<td>$\gamma_y$ (deg)</td>
<td>-7.34</td>
<td>-7.714</td>
<td>[-7.889, -7.539]</td>
</tr>
<tr>
<td>$\gamma_z$ (deg)</td>
<td>0</td>
<td>0.668</td>
<td>[0.330, 1.007]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Final Pose</th>
<th>Reference</th>
<th>Estimation</th>
<th>3σ bounds</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_{end}$ (m)</td>
<td>-0.06</td>
<td>-0.11</td>
<td>[-1.32, 1.09]</td>
</tr>
<tr>
<td>$y_{end}$ (m)</td>
<td>0.15</td>
<td>0.56</td>
<td>[-1.76, 2.89]</td>
</tr>
<tr>
<td>$\theta_{end}$ (deg)</td>
<td>0</td>
<td>-1.18</td>
<td>[-12.96, 10.58]</td>
</tr>
</tbody>
</table>
Our approach relies on a monocular camera on board the vehicle and the use of a database of spherical images of the urban scene acquired during an offline phase. This geo-referenced database allows us to obtain a robust “drift free” localization. Basically, the database is constituted of spherical images augmented by depth which are positioned in a GIS (Geographic information system). This spherical robot centered representation accurately represents all necessary information for vision based navigation and mapping (\[26\]). During the online navigation, the current vehicle position is computed by aligning the current vehicle camera view with the closest reference sphere extracted from the database.

A spherical augmented acquisition system has been developed and tested on our Cycab vehicle. This system is composed of six wide angle stereo cameras in overlap, which permits to extract depth information by dense correspondence. Since the depth is available, we are able to construct 360 degrees spherical images with a unique center of projection. Those 3D spheres are then used in an image-based spherical odometry algorithm to obtain the trajectory of the vehicle (\[31\]), fuse the spheres and construct the database.

During the online navigation, we consider a vehicle equipped with a simple camera (perspective, omnidirectional...). Here the aim is to register the current view on the closest sphere stored in the database. To achieve this we have developed a spherical image-based registration which allows efficient and accurate localization. But since the database of augmented visual spheres can be acquired under different illumination conditions than the online camera is experiencing, a robust algorithm combining model based localization and online visual odometry has been developed \[32\]. This method performs in real-time (45 Hz), and allows to handle large illumination changes and outliers rejection (see figure 5).

As a part of the ANR CityVIP project, the localization and mapping system has been successfully tested in Clermont Ferrand (France). A database of augmented images has been built along a learning trajectory. The aim was to automatically “replay” the learning trajectory using the database and a monocular camera. To avoid collisions and pedestrians, a laser was mounted on the front of the Cycab. The system was able to autonomously follow large scale trajectories (over 400 meters), in crowded urban environments (see figure 6).

Figure 5. Top left, robust outliers weights. Top center, augmented reference image. Top right, reference depth-map. Bottom left, intensity error after alignment. Bottom center, registered image. Bottom right, original current image.
6.1.3. Loop closure detection for spherical view based SLAM

Although more precise than the odometry computed from the wheels encoders, the visual odometry also suffers from the problem of drift when large displacements are performed. It is possible to correct this drift if the robot is capable to determine if the the place it is visiting has already been visited. re-observes a scene previously observed. This is often referred as the loop closure detection problem and several methods exist in the literature using perspective cameras. We develop new methods more reliable by exploiting the peculiar properties of spherical cameras.

Standard perspective cameras have a limited field of view leading to an incapability to encompass all the surrounding environment. This limitation of the field of view drastically limits the performances of visual loop closure algorithms. We propose to use spheres of vision computed by mosaicing images from 6 wide angle cameras mounted on a ring. Such a representation offers a full 360° field of view and keeps the spherical image invariant to the changes in orientations (Figure 7).

Loop closure detection can be exploited in a SLAM context at two levels: firstly, in the metric representation to retro-propagate along the robot’s trajectory the cumulative errors due to the drift, secondly, in the topological representation, to fusion in the graph representation the nodes corresponding to a same place.

Existing algorithms are not point of view independent: loop closures are detected uniquely when a place is revisited by the robot coming from the same direction but if the robot comes back in a different direction, the algorithms fail. Our solution relies on the presented spherical view and an efficient way of information extraction from it. We extract local information describing the points of interest of the scene. We enhance this local information with a global descriptor characteristic of the distribution of the points of interest over the sphere thereby describing the environment structure. These informations are used to retrieve the already visited places. Our algorithm performs well and is robust to the point of view variation [27]. This has led to an accepted paper at OMNIVIS 2011.
The figure below presents obtained results. The trajectory is corrected (drift reduction) using the loop closure constraint. Red and green dots represent the loop closing places, they are linked by a red line.

### 6.1.4. Context-based segmentation of scenes

In a topological SLAM framework based on vision, the places are often represented by images gathered at regular time/distance intervals. It is nevertheless a meaningless representation in the context of topology. We
would prefer a definition like "in front of a building" or "entrance of the campus" instead of "image number i". Places are thus a set of images we need to group. This is what we call context based segmentation. In order to achieve this segmentation a criterion for "changing place" is needed, we propose to evaluate the environment structure using a global spatial descriptor (computed on the spherical view) called GIST. The algorithm relies on a statistical process control monitoring for an out-of-control signal involving a changing place event. The algorithm still needs to be improved for better robustness on the localization of the changing place events when we come back on previous visited paths.

The figure 9 presents the preliminary results. On the bottom left is the similarity matrix of the images GIST while on the bottom right is the segmented trajectory followed by our robot.

![Figure 9. Context-based segmentation of scenes](image)

### 6.1.5. Nonlinear observers for visual-inertial fusion with IMU-bias and camera-to-IMU rotation estimation

This work concerns the fusion of visual and inertial measurements in order to obtain high-frequency and accurate pose estimates of a visual-inertial sensor. While cameras can provide fairly accurate pose (position and orientation) estimates, the data acquisition frequency and signal processing complexity limit the capacities of such sensors in the case of highly dynamic motions. An IMU (Inertial Measurement Unit) can efficiently complement the visual sensor due to its high frequency acquisition, large bandwidth, and easy-to-process signals. IMU biases and calibration errors of the displacement between the camera frame and the IMU frame, however, can severely impair the fusion of visual and inertial data. Identification of these biases and calibration of this displacement can be achieved with dedicated measurement tools, but this requires expensive equipment and it is time consuming. We propose instead to address these issues via the design of observers. Last year, we had proposed a nonlinear observer to fuse pose and IMU measurements while identifying additive IMU biases on both gyrometers and accelerometers. We have extended this work to the self-calibration of the
rotation between the pose sensor frame (camera) and the IMU frame. Simulation and experimental results have confirmed that this calibration significantly improves the final pose estimation and allows to process motions with faster dynamics. This work has been presented at the IROS conference in october [35]. It is a joint work with G. Silveira from CTI in Brazil. We are currently extending this result to include the self-calibration of the translation displacement between both sensors.

6.2. Control of mobile robots

**Participants:** Claude Samson, Pascal Morin, Minh-Duc Hua [Post Doc, I3S, CNRS-Univ. of Nice-Sophia Antipolis], Daniele Pucci, Glauco Scandaroli, Luca Marchetti, Tarek Hamel [Univ. of Nice-Sophia Antipolis].

6.2.1. New developments of the Transverse Function control approach

We are pursuing the development of the Transverse Function control approach for highly nonlinear systems via the application of the approach to challenging mechanical systems with various structural control properties.

6.2.1.1. Control of a redundant wheeled snake mechanism using transverse functions on $SO(4)$

The Transverse Function approach is applied to the control of a nonholonomic three-segments/snake-like wheeled mechanism, similar to the planar low-dimensional version of Hirose’s Active Cord Mechanism (ACM) previously studied [65], but with two additional internal degrees of freedom (d.o.f.) whose actuation yields more flexible and efficient control solutions (see figure 10). From a theoretical point of view, these complementary d.o.f. modify the Control Lie Algebra of the system so that only first-order Lie brackets of the control vector fields are needed to satisfy the Lie Algebra Rank Condition (LARC). The fact that four independent (angular velocity) control inputs are used also implies for this system the existence of Transverse Functions (TF) defined on the six-dimensional special orthogonal group $SO(4)$. Several examples of mechanisms whose control involve TF defined on $SO(3)$ have been pointed out in the past [54], [64], [65]. Beyond the specific control problem addressed here, a motivation for the present study is to illustrate for the first time how functions defined on the larger set $SO(4)$ can be determined and used for the control of a physical system. This study is complemented with recalls concerning the parametrization of $SO(4)$ by pairs of isoclinic quaternions and with the derivation of complementary differential calculus relations associated with this parametrization. The results will soon be submitted for presentation at an international conference.

6.2.1.2. Control of three hooked vehicles with off-axle hitches

An extension of the study [65] performed last year on Hirose’s Active Cord Mechanism (ACM) concerns the case when one of the wheeled-trains (the middle one, for instance) possesses actuators giving it tracing and rotating capacities (alike a unicycle-like vehicle), while the other two vehicles are passively hooked to this tracing vehicle. This type of actuation departs from the one of Hirose’s Active Cord Mechanism for which the tracing capacity of the mechanism relies exclusively on the deformation of the system of vehicles via the control of the inter-connecting angles, and it makes an important difference at the control level. This system may also be seen as a unicycle-type vehicle with two trailers and off-axle hitches. Unlike the simpler hitch-on-axle case commonly addressed in the literature, this system is not differentially flat and “complete” feedback solutions ensuring practical stabilization of any, feasible or non-feasible, trajectory remained an open issue. This actuation allows for the complete alignment of the three vehicles without going through actuation singularities, and for the asymptotic tracking of a reference frame moving along a straight line or a circle. On the other hand, in order to fully take advantage of the extra possibilities offered by it, one has to consider higher-order Lie bracket manoeuvring motions that significantly complicate the feedback control design. The Transverse Function approach is applied using the fact that a dynamic extension of this two-control-inputs system is left-invariant on a 6-dimensional Lie group. Transverse functions calculated as the group product of “elementary” functions defined either on toruses or on $SO(3)$, and yielding feedback controls ensuring asymptotic stabilization of “feasible” reference trajectories under common "persistent excitation" properties (as in the case of classical feedback control solutions based on a linear approximation of the associated tracking error system) are proposed. As usual, the superiority of the transverse function solution over more classical solutions comes from that it also applies to the case of non-feasible reference trajectories for which (practical) stabilization involves complex maneuvers. The results of this study will be submitted next year for presentation at an international conference.
Figure 10. Three segments snake robot with two steering wheels. View from above
6.2.1.3. Control of an extended trident-snake vehicle

This study is part of a thesis work on the control of non-standard nonholonomic mobile robots by W. Magiera under the dual supervision of Prof. K. Tchon (Wroclaw University of Technology) and C. Samson. This collaboration involves several long term visits of the PhD student at INRIA, starting this year (2 months), and for the next two years. This year’s objective is to address a particularly challenging control problem and evaluate the possibilities offered by the Transverse Function approach to solve it. The system under consideration is based on the "common" trident snake mechanism [54] complemented with one, two, or three additional "passive" wheeled extensions, each of them subjected to the rolling-without-slipping constraint (see figure 6.2.1.3). Transverse Functions solutions tested so far involve a mixt (product) of functions defined either on the torus, or on special orthogonal groups, and future improvements may involve the search for new transverse functions.

Figure 11. Trident-snake mechanism with passive extensions. View from above

6.2.2. Control of aerial vehicles

6.2.2.1. Vehicles subjected to lift forces

The development of a general theory for the control of underactuated (ground, marine, and aerial) vehicles whose main propulsion relies on a thrust force exerted in a single (vehicle’s related) direction was continued
this year. Part of this program, more specifically devoted to aerial vehicles, is the subject of D. Pucci’s thesis research project. This year’s focus was the prolongation of the work initiated last year on the modelling of lift forces and on their effects on the flight and control of aerial vehicles. Among the new results obtained on the subject, an extension and generalization of a previous feedback control strategy developed for spherically-shaped vehicles only subjected to drag forces [52], based on an “ideal” generic model of lift and drag forces associated with bi-symmetric wings, has been accepted for presentation at an international conference ([34]). The proposed solution involves a change of thrust control input in order to render the dynamics of the transformed system independent of the angle of attack associated with the vehicle’s main wing. A weakness of the aforementioned model is that it does not account for the so-called stall phenomenon, which is an abrupt loss of lift when the angle of attack increases beyond a certain value called stall angle. Taking it into account adds considerable complexity to the vehicle’s dynamics, especially in the case of a vehicle moving within a fluid endowed with a large Reynolds number for which the stall phenomenon can no longer be neglected. We showed that, although this phenomenon never forbids the existence of an attitude equilibrium given a desired reference velocity, the uniqueness of this equilibrium is not always granted. As a consequence, modifications of the desired velocity may result in the abrupt disappearance of an equilibrium so that the asymptotic stabilization of a desired velocity profile may become an ill-conditioned problem. To avoid this complication a possibility consists in characterizing “good” velocity profiles –associated, for instance, to transition maneuvers between hovering and high-velocity cruising– for which the existence of continuously changing equilibria is ensured. First results on this topic and research direction have been submitted for presentation to an international conference.

### 6.2.2.2. Nonlinear control of VTOL UAVs with uncertain position measurements

This work concerns the feedback control of VTOL UAVs (Vertical Take-Off and Landing Unmanned Aerial Vehicles). The objective is to asymptotically stabilize a reference equilibrium configuration with a “semi-global” convergence domain, i.e global convergence domain in position and semi-global in orientation, knowing that a global convergence domain in orientation cannot be obtained with continuous feedback laws due to the topology of the rotation space $SO(3)$. Several solutions to this problem have been proposed in the past years, under the assumption that the pose (i.e. position and orientation) is completely known. This work concerns the case when the relation between the "position measurements" and the true position vector is uncertain. In practice, such uncertainties are related, e.g., to ill-calibrated sensors or to incomplete knowledge of the environment in the case of proximetry sensors. It is assumed that position measurements are given by $\overline{p} = Mp$ where $p$ is the true position error with respect to the reference position, and $M$ is an unknown invertible matrix. As a first contribution, we propose a class of feedback laws that achieve semi-global stability of the equilibrium $p = 0$ for any matrix $M$ that satisfies the stability criteria $\|M - I_3\| < \delta(k)$ where $I_3$ is the $3 \times 3$ identity matrix, $\delta$ is a strictly positive function, and $k$ is the vector of control parameters. An explicit expression of the function $\delta$ is provided, thus relating the control parameters to the stability margin. The second contribution of this work is the application of this control approach to the visual servoing of VTOL UAVs with respect to a planar vertical structure (wall, etc). From the homography matrix that relates the current camera image to a reference image (taken at the reference pose), we derive a signal output of the form $\overline{p} = Mp$. The matrix $M$ typically depends on unknown parameters but we show that a very rough knowledge of them is sufficient to design a stable controller based on the above-mentioned stability criterion. These results have been submitted for publication at an international conference. This is a joint work with H. de Plinval and P. Mouyon from ONERA Toulouse.

### 6.2.3. Development of an autonomous shopping cart

This work, which consists in developing a shopping cart with autonomy capabilities (automatic user following, obstacle avoidance, etc), is part of the national INRIA PAL project (Personally Assisted Living) which aims at developing robotic tools for disabled persons or elderslies.

The architecture of “Autonomous Shopping Cart” has been developed in three layers. The first one is responsible for connecting the services layer to physical (or virtual) devices. During this year, all necessary components to access the devices have been implemented:
- the Phidgets library wraps the API of Phidgets devices and abstracts the access to the peripherals on the wheelchair robot;
- the Hannibal library wraps the interface to access the Hannibal robot (through Carmen library);
- the Simulator library wraps the simulator interface.

All of them expose a common interface to the software modules. Thus, the higher components do not have to be changed if the test platform changes.

The second level is the core of this year’s work. It is composed by the Control module and Modeling module. The Control module aims at stabilizing the trajectory of the robot w.r.t a given reference motion. In practice, this reference motion corresponds to the cart user that needs to be followed, but it could be any virtual reference motion. A first implementation of the Control module has been made using Matlab software. The result is a Control library that contains two different methods for controlling the trajectory:

- position control only;
- full-state control (position+orientation).

These methods have been implemented in Matlab language and then converted in C++. The resulting library has been utilized within the Control module deployed on the robotic platform.

The main objective of the Modeling module is to detect the cart user within the sight of the sensors. This task is generally non-trivial, due to noise in the sensor signals and variations of the environment. For this reason a Multiple Hypothesis Tracker has been used to allow for the presence of several persons in the environment. The method uses the laser scans to extract potential persons and then a Selector algorithm extracts the best hypothesis for the cart user. This hypothesis is then converted into a virtual reference point given to the Control module for trajectory tracking.

The third layer is represented by the Behavior module. This component manages the other modules, starts and stops services on request, enables the initialization procedure and so on. As for now, it starts all modules and initiates the starting procedure. In particular, it selects the first person to be tracked, among possible candidates. Experiments have been successfully conducted both on the mobile robot Hannibal and on the wheeled walking aid ANG (Assistive Navigation Guide) developed by the EPI Coprin.
5. New Results

5.1. Introduction

This year we obtained new results in our three sub-objectives and also related to Focuslab platform and software valorization:

1. **Sub-objective 1 - Mining for Knowledge Discovery in Information Systems**: this we get five results (with one achieved PhD thesis).

   Let us note that six past works on this sub-objective described in previous AxIS annual reports have been published this year as articles in international journals ([22], [11]) or conferences, one in a national journal [50], two in a french-speaking conference [35], [46], one book [20] and one book chapter [52] at international level. Indeed
   - The work in 2009 on mining data streams by Marascu in her thesis [96] has been published in [11] with more details in the algorithms and in the experiments.
   - The work published in 2008 on discovering frequent behaviors [107] has been published in [22] with more details in the algorithms and in the experiments.
   - Our previous work on satellite image mining in 2010 [99] has been published in French at EGC [35].
   - Our past work on Functional data analysis involving data described by regular functions rather than by a finite number of real valued variables has been published as a scientific book chapter [52]. In this paper we propose to use a clustering approach that targets variables rather than individual to design a piecewise constant representation of a set of functions. The contiguity constraint induced by the functional nature of the variables allows a polynomial complexity algorithm to give the optimal solution.
   - In the context of the WRUM project (Morocco) and Zemmouri’s PhD thesis, we have a long paper accepted at JFO 2011 related to past works (2010) on how to integrate domain knowledge in a multi-view KDD process [46].

2. **Sub-objective 2 - Information and Social Networks Mining for Supporting Information Retrieval**: Three results (with one achieved Ph-D thesis). Let us remind the best paper [15] obtained By E. Smirnova at ECIR 2011 for her research on expert finding.

3. **Sub-objective 3 - Multidisciplinary Research For Supporting User Oriented Innovation**: this interdisciplinary research is dedicated to the design, tailoring and refinement of methodologies and tools for a better users’ involvement in innovation processes. We have seven results this year.

Concerning our activity in terms of FocusLab Experimental Platform and Software, a) we first applied ATWUEDA on another context of evolutive data (on system monitoring data at EDF) which is different of Web usage data) to show the genericity of the approach [9], and b) we develop a Web-based version of the FocusLab experimental platform for analysis usage data (hardware and software parts).

5.2. Mining for Knowledge Discovery in Information Systems

5.2.1. Mining Data Streams: Clustering and Pattern extraction

**Participant**: Chongsheng Zhang.
In Zhang’s thesis [19] (supervised by F. Masseglia), which was partially founded by ANR MIDAS (cf. 6.2.1), we present our study of the management and mining issues on data streams with evolving tuples, caused by model updates or tuple revisions. For instance, in an online auction system where bids on auction items are streaming, it is possible that some users may bid for more than one item within the user-specified time interval. As a result, the profiles of the users can be updated or revised in such applications. Data streams having evolving tuples bring new challenges as well as research opportunity. In this work, he develops novel and efficient models and methods for managing and mining data streams with evolving tuples. (I) To model data streams with evolving tuples, we propose the Anti-Bouncing Streaming model (ABS) for usage streams. ABS fits data streams with evolving tuples and it enables methods for processing of data streams to handle tuple updates or revisions. (II) To find frequent itemsets from data streams with evolving tuples over pane-based sliding windows, we conduct theoretical analysis and propose theorems which can avoid scanning the past slides to check for possible itemsets that may become frequent. We also design novel data structures which can manage the data streams with evolving tuples efficiently and facilitate the mining of frequent itemsets. Moreover, we devise an efficient counting algorithm to verify the frequentness of the candidate frequent itemsets. We also propose two running frameworks for this problem. (III) To extract important feature set from data streams (including the ones with evolving tuples), based upon ABS, we devise the streaming feature set selection algorithm for data streams which is the first in the literature. This method is based on information theory to extract the informative feature sets. To further accelerate the extraction of the most informative feature set from high-dimensional data, we propose a framework that reduces the huge search space to a rather small subset while still guarantee the quality of the discovered feature sets.

In 2011, Chongsheng Zhang has mainly worked on a data stream mining method, intending to extract frequent itemsets. This method has not been published yet and is described in Chapter 5 (page 79) of his thesis document [19].

### 5.2.2. Clustering on Multiple Dissimilarity Matrices

**Participants:** Yves Lechevallier, Francisco de A.T. de Carvalho, Thierry Despeyroux, Alessandra Silva Anyzewski.

In [23] we introduce hard clustering algorithms that are able to partitioning objects taking into account simultaneously their relational descriptions given by multiple dissimilarity matrices [49]. The aim is to obtain a collaborative role of the different dissimilarity matrices in order to obtain a final consensus partition. These matrices could have been generated using different sets of variables and a fixed dissimilarity function or using a fixed set of variables and different dissimilarity functions, or using different sets of variables and dissimilarity functions.

These methods, which are based on the dynamic hard clustering algorithm for relational data as well as on the dynamic clustering algorithm based on adaptive distances, are designed to furnish a partition and a prototype for each cluster as well as to learn a relevance weight for each dissimilarity matrix by optimizing an adequacy criterion that measures the fitting between clusters and their representatives. These relevance weights change at each algorithm iteration and can either be the same for all clusters or different from one cluster to another. The usefulness of these partitioning hard clustering algorithms are shown on two time trajectory real world datasets.

### 5.2.3. Clustering of Constrained Symbolic Data

**Participants:**Marc Csernel, Francisco de A.T. de Carvalho.

In the context of our FACEPE collaboration with Brazil (cf. section 6.4.3.1), we have presented a method which allows clustering of symbolic descriptions constrained by presence rules in a polynomial time instead of a combinatorial one. This method allows to deal with "false missing values". Such a method can be applied on various classification problems [26].

### 5.2.4. Web Page Clustering based on a Community Detection Algorithm

**Participants:** Yves Lechevallier, Yacine Slimani.
Extracting knowledge from Web user’s access data in Web Usage Mining (WUM) process is a challenging task that is continuing to gain importance as the size of the web and its user-base increase. That is why meaningful methods have been proposed in the literature in order to understand the behaviour of the user in the web and improve the access modes to information. In this work [42], we are interested in the analysis of the user browsing behavior. The objective is to understand the navigational practices of users (teachers, students and administrative staff). First we clean the data by removing irrelevant information and noise. During the second step, remaining data are arranged in a coherent way in order to identify user sessions. After we defined a new approach [42] of knowledge extraction. This approach treats the data resulting from the preprocessing phase (first and second steps) as being a set of communities. Our approach extends the Modularity measure, proposed by Newman and Girvan [97], in the Web Mining context in order to benefit from their classifying capacity in the communities discovery.

This work is done in collaboration with the LRIA laboratory – Université Ferhat Abbas, Sétif, Algérie

5.2.5. Critical Edition of Sanskrit Texts

**Participants:** Marc Csernel, Nicolas Béchet, Ehab Hassan, Yves Lechevallier.

New progresses concerning the computer assisted elaboration of Sanskrit texts have been made. First Nicolas Béchet and Marc Csernel have worked on the problem of moved texts. After an alignment between two versions of the texts, we discover that some parts of the text appears to have been moved according to the technics developed in [48]. Until now, we were not able to discover when a text has been moved in a manuscript.

Now using a words-grams technique proposed in [48], we were able to obtain quite good results on the moved texts problem and we were able to optimize the different possible parameters. A paper on the subject has been submitted to the Cicling 2012 conference (http://www.cicling.org/2012/).

After the new treatment related to the moved text problem, we need to provide an interactive display of the critical edition. During his internship, Ehab Hassan has been working on the subject and obtained good results. These results need to be deeply examined by Sanskritists to see if they always fulfill their needs.

5.3. Information and Social Networks Mining for Supporting Information Retrieval

5.3.1. Clustering of Relational Data and Social Network Data

**Participants:** Yves Lechevallier, Amine Louati.

The automatic detection of communities in a social network can provide this kind of graph aggregation. The objective of graph aggregations is to produce small and understandable summaries and can highlight communities in the network, which greatly facilitates the interpretation.

Social networks allow having a global view of the different actors and different interactions between them, thus facilitating the analysis and information retrieval.

In the enterprise context, a considerable amount of information is stored in relational databases. Therefore, relational database can be a rich source to extract social network. The extracted network has in general a huge size which makes its analyses and visualization difficult tasks. In [45], we propose a social network extraction approach from relational database.

Often, the network has a large size which makes its analysis and visualization difficult.

The aggregation step is a necessary task, so we offer [33] and [32] an aggregation step based on the k-SNAP algorithm [109] that produces a summary graph by grouping nodes based on attributes and relationships selected by the user.

This work is done in collaboration with Marie-Aude Aufaure, head of the Business Intelligence Team, Ecole Centrale Paris, MAS Laboratory.
5.3.2. Networks Solutions for Expert Finding and People Name Disambiguation

Participants: Elena Smirnova, Yi-Ling Kuo, Brigitte Trousse.

The task of finding people who are experts on a given topic has cently attracted close attention. State-of-the-art expert finding algorithms uncover knowledge areas of candidate experts based on textual content of associated documents. While powerful, these models ignore social structure that might be available. Therefore, we develop a Bayesian hierarchical model for expert finding that accounts for both content and social relationships. The model assumes that social links are determined by expertise similarity between candidates. The results of EGC experiments on UvT expert collection have demonstrated the effectiveness of our algorithm [43].

E. Smirnova visited Intellius, people search technology company (Aug 8 - Oct 5, 2011): the goal of this visit was to validate the research on expert finding in social networks on real dataset and further advance it. As a real dataset, we have taken a sample of United States LinkedIn public profiles. We built an organizational network by connecting a LinkedIn user and his colleagues at different workplaces. We also constructed a geographical network from user’s current location in the United States. We used Amazon’s Mechanical Turk framework (http://aws.amazon.com/code/923) to collect user-oriented judgements for model evaluation. We found that the user-oriented model is statistically significantly prefered to the baseline model on 72.5% of queries.

Her work on name disambiguation done in 2010 has been integrated in an article related to the problem of quick detection of top-k Personnalized PageRank (PPR) in [24]. The effectiveness of the chosen approach based on Monte Carlo methods for quick detection of top-k PPR lists has been demonstrated on the Web and Wikipedia graphs.

Yi-Ling Kuo during her internship has worked on Person Name Disambiguation and started by managing the analysis of the very huge Yahoo! Web graph.

This topic has been done in the context of Smirnova’s thesis [18] which has been defended on december 15 (thesis supervised by B. Trousse (AxIS) and K.Avrachenkov (Maestro)).

5.3.3. Towards an On-Line Analysis of Tweets Processing

Participant: Nicolas Béchet.

Tweets exchanged over the Internet represent an important source of information, even if their characteristics make them difficult to analyze (a maximum of 140 characters, etc.). In [25], we define a data warehouse model to analyze large volumes of tweets by proposing measures relevant in the context of knowledge discovery. The use of data warehouses as a tool for the storage and analysis of textual documents is not new but current measures are not well-suited to the specificities of the manipulated data. We also propose a new way for extracting the context of a concept in a hierarchy. Experiments carried out on real data underline the relevance of our proposal.

This work is done inside a collaboration with LIRMM and CEMAGREF.

5.4. Multidisciplinary Research For Supporting User Oriented Innovation

5.4.1. Usability Design and Evaluation Methods

Participants: Dominique Scapin, Yves Lechevallier, Pascal Marie-Dessoude, Claudia Detraux.

We pursued our work on articulation of usage mining approach and human factors expertise for the design and evaluation of information systems. Namely, collaborative clustering techniques were used to analyze data issued from users via a card sorting technique, with respect to an a priori (“expert”) clustering. Considering the difficulties that people have in managing large information sets in their everyday life, for either professional or non-professional purposes (administration, social relationships, leisure, etc.), our recent research focusses on personal information space for which information bits are currently scattered many places.
In this PIMs field, there is little research with a user-centric approach, with the view that users-based knowledge might help specifying computer-based tools and a state-of-the-art [37] showed little work specifically on usability. Studies address a variety of questions from user needs to accessibility (including studies on older people) or user acceptance, among others. In the context of user-centered and long-term studies to understand the evolution of user information practice, we looked in a study at the intuitive way people organize their personal information, with or without computer systems, in order to help the design of future systems. Also, we recently surveyed 15 tools that claim to support personal information management.

5.4.2. Living Lab Landscape

Participants: Marc Pallot, Brigitte Trousse, Bernard Senach, Dominique Scapin.

In order to provide to the research community a comprehensive landscape of research streams in the Living Lab domain, we launched a study on the state-of-the-art about the ubiquitous notion of User Experience. During this continuous study, a landscape [34], [69] has progressively emerged that we organized through 4 main axes: focus granularity (individual/group), user’s role in the design process (observed subject/value creator), collaboration style (structured/unstructured), and evaluation purpose (reliability/adoptability). Our landscape of research streams has been used by Finnish colleagues who conducted an empirical study on the use of the Living Lab research domain landscape as a tool for assessing the maturity level of 16 Finnish Living Labs [36]. The Living Lab research domain Landscape has allowed the study team to identify four categories of Living Labs.

5.4.3. Future Internet Domain Landscape

Participants: Marc Pallot, Brigitte Trousse, Bernard Senach.

There are many different Internet research areas and corresponding technologies that were already investigated, experimented and progressively deployed such as peer-to-peer, autonomous, cognitive and ad hoc networking, that have already demonstrated how to improve network performance and user experience. Peer-to-peer networking for large-scale distributed systems and widely used applications has proved both the feasibility and economic potential for delivering services to millions of users. Others emerged more recently in the context of the future Internet (FI), such as Cloud Computing for transparently sharing among users scalable elastic resources over a limitless network. As it remains difficult to visualise the conceptual evolution and articulate the various Internet research areas, we conducted a study for identifying the appropriate concepts that could populate the FI domain landscape [69], [71] over three different periods of time (1990-1999, 2000-2005 and 2006-2011). Several INRIA research teams are involved in FIRE (Future Internet Research Experimentation) Testbed projects, namely: PlanetLab, OneLab, TEFIS, SensLAB, and BonFIRE whose scientific leaders were interviewed during the development of the FI domain landscape. Four dimensions were used for landscaping the Future Internet research domain: evolution approaches (from incremental evolution design to Clean Slate re-design or radical evolution), Internet routing (from the basic data packet delivery towards more sophisticated content distribution and retrieval capacities such as content Centric Networking), network type (from wired communication to wireless communication networks), evolution trend (from computer network towards network computing). The resulting tentative landscape of FI research areas shown in Figure 1 is intended to provide a faster and broader understanding of the different Internet research streams and related topics.

5.4.4. Future of Internet and User-open innovation for Smart Cities

Participants: Caroline Tiffon, Marc Pallot, Brigitte Trousse, Bernard Senach.

The goal of the Fireball project is to bring together three different constituencies: user driven open innovation, Future Internet, and Smart Cities [34], [39]. It aims at defining a roadmap [29], based on analysis of needs, opportunities and gaps, to benefit a wide scale implementation of the methodologies and concepts elaborated. A first objective in the project was to get a clear picture of the state-of-the-art in each domain. During the review, progressively emerges a landscape [69] that we organized along 4 main axes: wiring (wired/wireless), user’s role (subject/actor), Internet evolution approach (structured/unstructured), evaluation purpose (reliability/adoptability). A large variety of FI research have been engaged. If initial efforts in Future
Internet research have been directed towards the goal of providing the technical infrastructure supporting the next network generation, a rising trend in this research field is to consider now a higher level layer, the layer of services.

5.4.5. Method and Tool for Selection of Open Innovation Software Tools

Participants: Mylène Leitzelman, Brigitte Trousse.

In spite of an important number of tools supporting open innovation, there is few comparative evaluation and no grid or evaluation criteria helping to choose a product. A 2011 review in the Computer Aided Innovation field \(^6\) provides a large overview of available tools in relation with a wide range of innovation cycle features. This useful top-down categorization approach is of little help to choose a specific tool. To find the best OI tool supporting idea exchanges among a community of participants, we built an exploratory method on the Web and we elaborate an assessment grid of OI tools based on the QSOS method which is a method designed to qualify, select and compare free and open source software in an objective, traceable and argued way. It publicly available under the terms of the GNU Free Documentation License. In our QSOS method, evaluation criteria are organized in a tree-hierarchy grid with, a scoring method procedure of each tree-leaf criteria (from 0 - not covered to 2 - completely covered). To achieve the construction of the OI tools criteria assessment grid (first step), we used different mining tools for Web crawling, network analysis, criteria classification and from the 29 top rated OI tools, we finally selected 6 of them from which we extracted the tree-map categorization used to build the reference software criteria sheet. In the following steps, after appropriate weighting, we used the provided OS3 Web application and we were able to compare 4 top selected OI tools [31].

To support our method, we developed a QSOS-based OI Grid for supporting the OI tool selection. The QSOS Grid to compare OI tools, which is in an XML format, has been translated into a MindMap. As a first mock-up, the QSOS Web Interface O3S will be installed at the beginning of 2012 on our server. For the future, we

---

\(^6\) Hüsig, University Regensburg & Khon, Otto Beisheim School of Management
will extend our own QSOS comparative method to other softwares and to other categories interesting for the Living Lab community.

Figure 2. Radar Positioning of Selected Open Innovation Tools

5.4.6. Extension of Methods for Idea Generation Process

Participants: Anne-Laure Negri, Caroline Tiffon, Bernard Senach, Brigitte Trousse.

Internet of Things (IoT) is supposed to be a physical world where everyday objects, rooms and machines are connected to one another and to the larger digital world. In this web of people and objects, individuals as well as things will have their own unique URL and, according to interactions between all these entities, social networks will be articulated with a ring of connected objects. This mash up of "hyper groups" and "hyper objects" will be the next step towards a deeper level of automation in which the user interface has vanished, the explicit control over the world is no longer the rule and where the users will have to understand the dynamic changes of their environment in relation to their social interactions and to their physical behaviour. The design of IoT based services therefore raises many challenges related to the user experience and requires a deep understanding of users’ needs in their real life environment as well as many field experiments; this is why the Living Lab approach appears as the qualification device which has been lacking so far. As described in a paper accepted for ServDes2012 conference (February 2012) among 80 papers submitted, we developed and tested two methods designed for Internet of Things service idea generation: GenIOT (Generative Technique for the Internet of Things) and Aloha! (Animation Lens: Object/Human Actoring!) methods.

- The GenIOT method aims at providing citizen a tool for identifying and sharing examples of situations where they may benefit from an Internet of Things based service. Indeed the objective is to help the introspective effort of analysing one’s own daily behavior and unveil situated data in order to develop grounded internet of things service ideation. Fake sensors are placed by the participants in their daily environment and pictures of the sensors in situ are shared and tagged on a collaborative platform. Other participants are invited to discuss online the ideas behind the pictures. Gamification rules are implemented in the platform in order to enhance collaboration and participation.

- The Aloha! method tackles another creative dimension as it asks participants to role or play characters or objects and bodystorm a collaborative scenario meeting the characters need. Participants report it to be a fun and effective method, alternative to traditional brainstorming and integrating serendipity as a creative asset.

For the exploration step in the context of ELLIOT, we developed NiceAir, an android mock-up for visualizing air quality data on Google map in Nice as well as some interest points (free bikes, free park places, bus stop, etc.). Such a mock-up will be available in the co-creation step of green services in the ELLIOT project and during the European Mobility Week.
5.4.7. Mock-ups for two innovation processes (exploration step)

Participants: Lucile Gramusset, Guillaume Pilot, Mohamed Gaieb, Bernard Senach, Brigitte Trousse.

In the context of two contracts (TICTAC and ELLIOT) related to user oriented innovation process, we have developed two mock-ups:

- Based on the user feedback from the first experiment in TICTAC (cf. section 6.1.1), we decided to develop a mock-up MobilTIC of a real-time information service related to public transportation (Envibus & SNCF) usable for any smartphone with internet access for Sophia-Antipolis workers. We designed a Web interface the most simple based on PHP/Mysql technologies and accessible from a computer or 3G mobile phone. Usage analysis of MobilTIC has been anticipated by providing rich logs. An exploration task with citizen is planned in January 2012 with a new experiment with an improved version.

- A first mock-up called Nice Air has been developed in the context of ELLIOT (cf. section 6.3.1.1) by L. Gramusset and M. Gaieb on android smartphone for providing information related to air quality and noise on a map of the area of Nice Cote d’Azur with some interest points such as bus stops and Vélib parkings.

5.5. FocusLab Experimental Platform and Software

5.5.1. ATWUEDA based Clustering Approach for System Monitoring

Participant: Yves Lechevallier.

Progressive advances in hardware and software technologies have enabled the production and storage of system monitoring data streams in a wide range of fields (e.g. telecommunications, sensor networks, etc.). Traditional clustering methods are unable to deal with data of such a voluminous and dynamic nature. In this work [51], we propose an efficient clustering approach (ATWUEDA) for monitoring massive time-changing data streams. This work considers a real case study on condition monitoring data streams of an electric power plant provided by EDF.

This work is done in collaboration with Alzennyr Da Silva of BILab laboratory (Telecom ParisTech and EDF R&D Common Laboratory).

5.5.2. FocusLab Experimental Platform (CPER Telius 2008-2012)

Participants: Xavier Augros, Mohamed Gaieb, Brigitte Trousse, Yves Lechevallier.

The FocusLab platform aims to be a major delivery mechanism of previous and current work in AxIS. It is a way to make methodological contributions (including software) available for the scientific community, but also a way for stimulating further research. This work has slowly started on the software part due to the absence of human resource funding and due to the absence of engineers in the team until 2010. Mid 2011, we started the specification and the development of a first version of a platform with three parts (hardware, software and methods) with the arrival of two engineers on the ELLIOT and TIC TAC contracts. A first version is available since the end of September as a Web portal and a second one is planned for the end of 2012 with advanced features for the software part (cf. service oriented platform, SOA architecture and interoperability).

Related to the software part we are in the process of developing several AxIS methods as Web services: we started with SCDS (cf. 4.2.3) which was demonstrated in the context of the MIDAS project on two applications (Orange labs mobile portal and vehicle trajectories) and in the ELLIOT platform (linked to the San rafaelle Hospital media use case) at the first review meeting (cf. 6.3.1.1) applied on data issued from San Rafaello Hospital use case. ATWUEDA (cf. 4.3.2), GEAR (Marascu’s thesis) are under development as web services.

Our work on mining evolutive data (ATWUEDA) and data streams (such as SCDS) have been used in real applications in the context of Internet of things and sensors: ATWUEDA (system monitoring for EDF (cf. section 5.5.1) and SDMS (cf. section 4.2.3).
6. New Results

6.1. Robotics

6.1.1. Calibration and identification

6.1.1.1. Calibration of a cable-driven robot

Participants: David Daney, Julien Alexandre dit Sandretto, Jean-Pierre Merlet, Gilles Trombettoni.

To improve the accuracy of a cable manipulator, it is necessary to identify the uncertainties of its model. The robots, studied in Cogiro, an ANR National initiative, are redundantly actuated: the number of powered wires is larger than the number of degrees of freedom of the manipulator. Under some cable properties hypothesis, this over-constraint mechanism allows to perform a self-calibration - i.e. the identification of the parameters does not need additional external measurement. A first experimentation, done in Montpellier, validated a novel approach which consists in a simultaneous identification of parameters and robot position (unknown in self-calibration process).

6.1.1.2. Cable properties

Participants: Julien Alexandre dit Sandretto, Gilles Trombettoni, David Daney.

The majority of researches done on cable-driven robot modeling need to take into account that the mass and the elasticity of wires are neglectable. However, they can not prove that these hypotheses are acceptable regarding these objectives. We have proposed an algorithm based on interval analysis to judge the validity of these assumptions for a cable-robot in a specific workspace. This method have been tested on the Tecnalia/LIRMM’s prototype and used for the construction of the Cogiro robot.

6.1.1.3. Optimal calibration poses of a 3-RPR planar parallel robot

Participant: David Daney.

The choice of the measurement configuration is crucial to improve the robustness of the calibration for measurement uncertainties. This year, a geometrical approach has been used to determine formally the set of the optimal poses for the identification of the kinematic parameter of a 3-RPR planar parallel robot. This result is important because it explains the influence of the location of some particular robot poses in their workspaces during the model identification process. A generalization is explored to construct automatically an optimal set for robot calibration and moreover, to improve experimental design algorithms. The aim is now to obtain similar results for cable driven robots calibration.

6.1.1.4. Geometric calibration of a space telescope

Participants: Thibault Gayral, David Daney, Jean-Pierre Merlet.

In October 2010 begun a collaborative work with Thales Alenia Space on the calibration of the mechanical structure of a space telescope. Its architecture is based on a parallel manipulator (type active wrist 6-PUS) used to correct the relative position of two mirrors. The aim is to identify the parameters of this robot, to improve its accuracy and then increase the quality of the images provided by the telescope. Thus, a geometric calibration procedure was considered and a campaign of photogrammetry was performed on the telescope. Using a kinematic description, a final accuracy of at worst 10 µm was reached on the position of the platform of the telescope. The aim of these measures were also to valid or not the model of comportment of the flexible parts of the device. This campaign brought to light the necessity to consider forces and torques acting on the structure in the deformation of the flexible parts in order to reach a submicrometric accuracy.

6.1.1.5. Modelization of flexible articulations of the telescope

Participants: Thibault Gayral, David Daney, Jean-Pierre Merlet.
In order to improve the final accuracy of the above-mentioned space telescope, a novel model including the statics equations in order to calculate deformations of the flexible articulations is currently under study. The main difficulty is to identify parameters (stiffness matrix and geometric parameters) that have different units and are not of the same order of magnitude. To solve this issue, we are focusing our effort to write the problem in a better robust form.

6.1.1.6. Interval Identification

**Participants:** David Daney, Julien Alexandre dit Sandretto, Gilles Trombettoni.

There are many approaches to identify the parameters of a model. In most cases, it consists in providing a particular solution of an over-constraint set of equations which must be robust to measurement to errors: in least square sense, with some statistical properties... However, the interpretation and the validity of the result can be difficult and prone error. We propose to investigate some interval approaches in order to associate to the result some information and a certification of solutions.

6.1.2. Rehabilitation and biomechanics

**Participants:** Sami Bennour, David Daney, Mandar Harshe, Jean-Pierre Merlet [correspondant].

The focus of the work is on analyzing knee joint motion during a walking activity. The measurement system is based on the wire actuated parallel robot architecture. To increase the reliability of our analysis, and decrease the influence of Skin Tissue Artifacts (STA), we also incorporate a passive wire measurement system, IR camera based motion capture system, accelerometers, and force sensors to measure human motions.

The main principle of the system is to observe relative motions of the collars attached to tibia and femur. These are connected to the base by wires and also hold the other sensors. Measurements in the global frame and collar specific local frames give precise data to reconstruct collar (and thus, knee joint) motion.

Over the past year we have finalized the experimental setup, by calibrating the collars and the sensor systems, and adapting the existing wire robot system (MARIONET-REHAB) to work along with the other sensors. The software developed uses a single unified input file to specify all sensor configurations, streamlining experiments. We performed our preliminary experimental trials for walking motion on three subjects using the wire sensors, accelerometers and optical motion capture system.

We began work on processing the data obtained from these trials. Post-processing functions have also been developed to calculate additional collar properties, perform sensor data processing (filtering, noise removal and estimation) and access files in the C3D file format, which is used a binary file format used by the motion capture system.

The main challenge we are working on is to perform sensor data fusion and increase reliability of results. For this we must identify parameters that correlate the different sensor measurements and perform error analysis. Possible solutions include using interval analysis methods to address the uncertainties.

6.1.3. Kinematics of wire parallel robots

**Participant:** Jean-Pierre Merlet.

The kinematics of wire robot is a complex problem because a solution is possible only if the tension in the wire is positive. Hence the static equilibrium has to be taken into account. This problem is not well addressed in the literature. Curiously the forward kinematics of robot (i.e. finding the possible poses of the platform for given wire lengths) with at least 6 wires is straightforward: the distance equations allows to determine all poses and then we use the static equations to calculate the wire tensions and discard the one having at least one negative tension. For robot having less than 6 wires we have to consider simultaneously the distance equations and the static equations in order to get a square system (of \( n + 6 \) equations for a \( n \)-wires robot). We have investigated the case of a 3-wires robot with all wires attached at different points on the platform and have shown that all solutions can be computed provided the solving of an univariate polynomial of degree 158 [17]. Although we are not able to guarantee that the degree of this polynomial cannot be decreased, we believe that nevertheless the order will be too high for robust determination of the solutions and can only be used to determine an upper bound for the maximal number of solutions. We have also investigated theoretically and experimentally the...
kinematics of a \( n \)-wires \( (n \geq 4) \) robot with all wires attached at the same point (i.e. only the position of the center of the platform can be controlled). Although this robot is apparently redundant, we have shown that in any pose at most 3 wires will be simultaneously under tension and therefore that the redundant wires cannot be used to control the wire tensions.

As the wire length measurements are not sufficient to determine the current pose of the platform (which is necessary for control purposes) we are investigating the use of additional sensors. Our prototypes \textsc{Marionet-Assist} and \textsc{Marionet-VR} are instrumented to measure wire directions, but with a large uncertainties. We have started a theoretical investigation to determine under which conditions these uncertainties may lead to a non-unique solution and we will validate the results on the two prototypes.

6.1.4. Rehabilitation robots for the immersive space

\textbf{Participants:} Michael Burman, Jean-Pierre Merlet.

The on-site immersive room provides 3D visualization but is lacking of haptic feedback and motion capabilities. We plan to implement in this room a movable system, constituted of:

- a 6 degrees-of freedom motion base: the motion system 710-6-500-220 by Servos Simulation Inc. has been selected and is now operational and fully calibrated. If necessary the user may stand on this motion base
- the \textsc{Marionet-VR} wire-driven parallel robot: this robot uses the same actuation principle than the \textsc{Marionet-Rehab} robot (linear actuator with a pulley system for coiling and uncoiling of the wires), but is able to lift a person. The prototype is basically functionnal but its installation in the immersive room has been delayed because of lack of appropriate fixing elements

The full system will be installed in the immersive room at the beginning of 2012.

6.1.5. Assistance robotics

\textbf{Participants:} Michael Burman, David Daney, Jean-Pierre Merlet.

As mentioned earlier in the report we have started in 2008 a long term strategic move toward assistance robotics, with the objectives of providing low-cost, simple to control, robotized communicant devices that may help disabled, elderly and handicapped people in their personal life, with the credo that they have to be adapted to the end-user and to its everyday environment (by contrast with the existing trend of focusing on a "universal" robot, to which the end-user and its environment have to adapt) \cite{18}, \cite{14}, \cite{21}. We have started last year the development of a simulated flat in order to explore various full scale scenarii that cover a part of the daily life of an elderly, to develop specific assistance devices and to test them. We describe in the following sections several devices that have been developed/improved during this year\(^1\). Note that our demonstration in assistance are highlighted during the visit of Sophia (275 visitors have attended our demonstration during 14 visits) and have received serious press coverage (5 papers, 2 TV interviews).

6.1.5.1. Walking aids

Wheeled walking aids are usually the first tools that are used when motricity problems occur. We are developing the family of robotized \textit{Assistive Navigation Guide} (ANG), which are based on commercially available Rollators, with several objectives (we mention only a few of them):

- fall prevention/detection: fall is a major problem for elderly (it is estimated that fall is the main cause of 10 000 elderly deaths per year in France).
- mobility help: provide an on-demand mobility help
- gait pattern monitoring: we believe that being able to monitor the trajectory of the walking aid will provide useful information on the gait pattern of the user

\(^{1}\) pictures of this assistive flat are available at \url{http://www-sop.inria.fr/coprin/developpements/main.html}
For reaching these objectives we have developed two walking aids:

- **ANG-light**: a walking aid with encoders in the wheels, 3D accelerometer, gyrometer. These sensors allow to measure the trajectory of the walking aid and several features of the user’s gait (step pattern, gait asymmetry,...). ANG-light has been tested by the CHU of Nice-Cimiez that was willing to perform an in-depth investigation of its use. For that purpose we have asked in September 2009 for the necessary formal authorization to the local CPP, which has been granted only in December 2011. To prepare this study we have organized a large scale experiment at INRIA, where 24 users were asked to perform the trajectories of the protocol twice, with and without the aid. When not using the aid the users were equipped with 3D accelerometer on the wrists and knees and were using specific shoes with force sensors in the sole. Initial analysis of the records shows that indeed we are able to obtain significant information on the gait pattern, that are not available using the existing tools, and detect differences in the gait pattern for user having even a light pathology in the lower limb. The experiment with elderly patients at CHU will take place in January 2012.

- **ANG-II**: this aid is an evolution of the motorized walker ANG, with a lower weight and better integration.

6.1.5.2. MARIONET-ASSIST

This wire-driven parallel robot is installed in the ceiling of the flat. It has been used this year in the 4-1 configuration (4 wires attached at the same point), which allows for controlling the position of the platform, but not its orientation. Several platforms have been developed, all of them incorporating a webcam and allowing for a free rotation around the vertical axis, while an accelerometer measure the tilt angle of the platform (which is used to determine in which direction the end-user is willing to move). One of the platform incorporate a 4 d.o.f. robot that may grasp light object (one of our objective is to use also the robot as a manipulator for bringing object back to the user in a more or less autonomous way, which is the subject of the PhD thesis of R. Ramadour).

We have shown that the the robot can be used for sit-to-stand transfer and for lifting handicapped people. A specific attention has been devoted to propose very simple control interface: joystick, remote TV set, control box whose tilt determine the motion axis.

6.1.5.3. Other flat equipments

Our scenario includes the management of emergency situations such as the fall. Fall detection can be performed by the ANG walkers but we have also started investigating the inclusion of fall detection system in the clothes of the end-user either through a GEO-300 devices or by incorporating an Arduino Lilypad processor. When a fall is detected indoor an alert is transmitted to a coordinator (a Nabaztag) which will order the walker and the MARIONET-ASSIST robot to move close to the user to provide a support. At the same time two mobile robots will converge to the same location: a remote-controlled, webcam equipped ROVIO (which can provide images of the end-user to a rescue center) and a Pekee II, that we will equip to provide first aid.

An important point in assistance is to be able to have at all time a rough idea of the localization of the patient. Although we plan to use a Kinect for that purpose, we will also investigate the use of non-vision sensors (which are much less intrusive and therefore can be more easily accepted) such as RFID tag (ANG-II has a RFID tag reader), directive distance sensors and light barriers.

Another axis for assistance is to reduce the risks of fall by using the principle that the objects has to come to the hand of the user (or of the robot), not the opposite. This implies instrumenting the environment with drawer openers and doors manipulation and we have started implementing them on drawers and on the fridge of the flat.

### 6.2. Interval analysis

#### 6.2.1. Inner Regions and Interval Linearizations for Global Optimization

**Participants:** Gilles Trombettoni [correspondant], Bertrand Neveu.
Researchers from interval analysis and constraint (logic) programming communities have studied intervals for their ability to manage infinite solution sets of numerical constraint systems. In particular, inner regions represent subsets of the search space in which all points are solutions. Our main contribution is the use of recent and new inner region extraction algorithms in the upper bounding phase of constrained global optimization. Convexification is a major key for efficiently lower bounding the objective function. We have adapted the convex interval taylorization proposed by Lin & Stadtherr for producing a reliable outer and inner polyhedral approximation of the solution set and a linearization of the objective function. Other original ingredients are part of our optimizer, including an efficient interval constraint propagation algorithm exploiting monotonicity of functions.

We end up with a new framework for reliable continuous constrained global optimization. This interval Branch & Bound significantly outperforms the best reliable global optimizers [22], [25], [28].

6.2.2. An Interval Extension Based on Occurrence Grouping

Participants: Bertrand Neveu [correspondent], Gilles Trombettoni.

We proposed last year a new “occurrence grouping” interval extension $[f]_{og}$ of a function $f$. When $f$ is not monotonic w.r.t. a variable $x$ in a given domain, we try to transform $f$ into a new function $f^{og}$ which is monotonic w.r.t. two subsets $x_a$ and $x_b$ of the occurrences of $x$: $f^{og}$ is increasing w.r.t. $x_a$ and decreasing w.r.t. $x_b$. $[f]_{og}$ is the interval extension by monotonicity of $f^{og}$ and produces a sharper interval image than the natural extension does.

This year we have improved the linear program and algorithm that minimize a Taylor-based over-estimate of the image diameter of $[f]_{og}$. We have detailed the proofs of correctness and reliability of this occurrence grouping algorithm [8], [29].

6.3. Miscellaneous results

6.3.1. Equilibrium strategies for linked Electricity and CO2 markets

Participant: Odile Pourtallier.

In collaboration with M. Bossy (INRIA -TOSCA Team) and N. Maïzi (CMA - Mines ParisTech) O. Pourtallier the study of equilibrium model for coupled electricity and CO2 allowance exchange markets has been pursued. (see also Section 7.1). A static equilibrium model has been studied under various assumptions on the CO2 market design. All the CO2 market designs do not lead equilibrium, which interferes on the (short term day ahead) electricity market, which in turn interferes on the electricity mix and consequently on the total emission. Together with El-Hadj Dia (INRIA -TOSCA Team) we have also pursued an indifference pricing methodology which is presented in more details in INRIA -TOSCA Team section.

6.3.2. Symbolic tools for modeling and simulation

Participant: Yves Papegay.

This activity is the main part of a long-term ongoing collaboration with Airbus whose goal is to directly translate the conceptual work of aeronautics engineers into digital simulators to accelerate aircraft design. An extensive modeling and simulation platform has been designed which includes a dedicated modeling language for the description of aircraft dynamics models in term of formulae and algorithms, and a symbolic compiler producing as target an efficient numerical simulation code ready to be plugged into a flight simulator, as well as a formatted documentation compliant with industrial requirements of corporate memory.
Implementation of this platform is a modeling and simulation environment based on symbolic computation tools. It contains several components:

- a model editor, that makes it possible and easy to enter the whole set of equations describing large and complex industrial models,
- an highly interactive and modular evaluation workbench allowing to simulate the models and to visualize the results inside the modeling environment with the benefits for the designer of being able to directly use all its computational functionalities.
- a C code generator which, using these models, automatically generates the numerical real-time simulation engines
- a technical documentation generator

During the year 2011 the technology demonstrated by our prototype has been transferred to our industrial partner. A lot of work has been done on our modeling and simulation environment to improve its robustness and its development level of quality toward industrial standards. Final version of our prototype is to be delivered to Airbus at the end of the year.

6.3.3. Multi-agent aircraft design

Participant: Yves Papegay.

The modeling environment described in the previous section is used, in collaboration with other teams at Airbus, in the framework of the ID4CS project founded by ANR and dedicated to multi-agent optimization of large scale system. Several models of aircraft engines and of aircrafts have been developed as user cases for the project. Automatic generation of extended models namely computing first order derivatives of the original models has been implemented.
6. New Results

6.1. Graph Based Knowledge Representation

6.1.1. Knowledge Graph Abstract Machine

Participants: Olivier Corby, Catherine Faron-Zucker, Fabien Gandon.

KGRAM (Knowledge Graph Abstract Machine) is a generic interpreter for W3C SPARQL Query Language that operates not only on RDF graphs but on labelled graphs. The interpreter interacts with the target graph through proxies that implement an interface: *Producer* enumerates edges from the target graph, *Evaluator* evaluates filters and *Matcher* takes entailments into account.

This year, work have been done to leverage KGRAM up to SPARQL 1.1 Query Language & Update. It implements most of current version of the recommendation, except the *service* statement. It passes almost all W3C SPARQL 1.1 test cases.

In addition, the Corese Semantic Web Factory has been redesigned and modularized into release 3.0 entirely based on KGRAM interfaces and proxies. Corese 3.0 is a new lightweight RDF/S implementation with SPARQL 1.1. We ported the former Inference Rule engines (forward and backward engines) onto Corese 3.0. We also ported former SPARQL extensions: approximate search based on ontological distance, SQL and XPath in SPARQL 1.1, edge enumeration and length of Property Path, pragmas.

This new version is already used in several applications among which: cartography at IGN [28], design constraint modeling at CSTB [35], technological watch in ISICIL ANR project. It is also used in several PhD Theses in the team. A list of applications can be found on Corese Web site6.

6.1.2. Semantic Web Graph Visualization

Participants: Olivier Corby, Nicolas Delaforge, Erwan Demairy, Fabien Gandon [contact].

Thanks to an INRIA grant (ADT), we design and develop a Semantic Web Gephi Plugin. This plugin is coupling Corese and the Gephi Open Graph Visualization Platform to provide a framework to query and visualize RDF data taking into account their schemas. See the web pages78.

6.1.3. Semantic Social Network Analysis

Participants: Guillaume Erétéo, Fabien Gandon.

The PhD thesis of Guillaume Erétéo [14] in the context of the ANR project ISICIL allowed us to analyze the characteristics of the heterogeneous social networks that emerge from the use of web-based social applications, with an original contribution that leverages Social Network Analysis with Semantic Web frameworks. Social Network Analysis (SNA) proposes graph algorithms to characterize the structure of a social network and its strategic positions.

Semantic Web frameworks allow representing and exchanging knowledge across web applications with a rich typed graph model (RDF), a query language (SPARQL) and schema definition frameworks (RDFS and OWL). In this thesis, we merged both models in order to go beyond the mining of the link structure of social graphs by integrating two approaches: (1) semantic processing of the network typing and (2) emerging knowledge of online activities.

---

6 http://www-sop.inria.fr/teams/edelweiss/software/corese/
8 https://gforge.inria.fr/projects/segviz-public/
In particular we investigated how (1) to bring online social data to ontology-based representations, (2) to conduct a social network analysis that takes advantage of the rich semantics of such representations, and (3) to semantically detect and label communities of online social networks and social tagging activities.

This work was published at [15], [14].

6.1.4. Index Summarizing the Content of RDF Triple Stores

Participants: Adrien Basse, Fabien Gandon, Isabelle Mirbel.

We are interested in designing an architecture to support the distribution of a SPARQL query on a small and fixed number of RDF repositories. To do so, the key stage is to characterize the content of the base of each server in order to be able to predict if a server could contribute or not to the answer of a query. In the context of the PhD Thesis of Adrien Basse we propose an algorithm to extract a compact representation of the content of an RDF store. We improved the canonical representation of RDF graphs based on DFS code proposed in the literature by providing a join operator to reduce the number of generated redundant patterns.

6.1.5. Rules for the Web of Data

Participants: Oumy Seye, Olivier Corby.

In the context of this PhD thesis, the focus is on Rules for the Web of data. We are interested in integrating Rule Interchange Format (RIF) - W3C recommendation for exchanging rules on Web - to others W3C technologies. The aim of this year is to study the integration possibilities of RIF-BLD into semantic Web technologies. RIF-BLD is the dialect of RIF for logic-based systems. Firstly, we have studied the state of the art. Secondly we improved the RIF-BLD parser for presentation syntax and XML syntax. As RIF-BLD can be used with RDF data and OWL ontologies, it is interesting to consider RIF inferences in queries on RDF graph structure. That is why we finally study the integration of RIF-BLD into the Corese Semantic Web engine. In this last step, we have implemented the mapping of abstract syntax tree of RIF-BLD to abstract syntax tree of SPARQL. Thus, we can now execute logic inferences of RIF-BLD in the backward engine of Corese.

We have a paper accepted at EGC 2012 presenting RIF2SPARQL [44], a translation of RIF-BLD statements in SPARQL to perform the logical inferences of RIF-BLD on the Corese Semantic Web Factory. These inferences are implemented in backward chaining approach. We have designed and implemented the mapping of RIF-BLD to SPARQL.

6.1.6. Collaborative Management of Interlingual Knowledge

Participants: Maxime Lefrançois, Fabien Gandon.

We are interested in bridging the gap between the world of natural language and the world of the Semantic Web, in particular to support multilingual access to the Web of Data and management of interlingual knowledge bases. We introduce the ULiS approach, that aims at designing a pivot-based NLP technique called Universal Linguistic System, using Semantic Web formalisms, and being compliant with the Meaning-Text theory. Through ULiS, a user could interact with an Interlingual Knowledge base (IKB) in controlled natural language. Linguistic resources themselves (e.g., dictionary, grammar) are part of a specific IKB, thus, actors may enhance them (i.e. the model of the controlled natural language), through requests in controlled natural language (e.g., add a new lexical units, add grammar rules).

In [30] we proposed a novel approach to define Interlingual Lexical Units classes in the Interlingual Lexical Ontology so that they support the projection of their lexicographic definition on themselves using the OWL formalism. This approach is compliant with the Meaning-Text Theory.

In [31], [40] we introduced three basic interaction scenario for ULiS and we proposed and overviewed the layered architecture of ULiS: meta-ontology, ontology, facts; and ontology, interlingual knowledge, situational knowledge.

We have started a collaboration with the RELIEF project that deals with the construction of a French Lexical Network (Alain Polguère, CNRS-ATILF).
6.1.7. Reuse of Data Analytics Contents and Processes

Participants: Corentin Follenfant, Fabien Gandon, Olivier Corby.

Industrial Business Intelligence (BI) proposes tools and methods to perform data analysis over heterogeneous enterprise sources. They allow one to harvest, federate, cleanse, annotate, query, organize and visualize data in order to support decision making with human-readable documents such as reports, dashboards, mobile visualizations. Such processes currently require expertise in technical domains like relational modeling in order to produce relevant content.

Users willing to do so without following the learning curve have to reuse existing content to create new one, and need to be guided throughout the workflow. Recommender systems can contribute to easing their progression, but most of them will operate inside walled garden for specific tasks instead of assisting the user throughout his workflow.

Semantic Web tools allow us to provide a common ground for modeling the different operations that compose BI workflows with RDFS vocabularies, capturing usage of the underlying transformations operators within document repositories with RDF graphs, and enabling further composition and reuse of BI operations to achieve new analysis. We introduced with [38] an extension of the RDF Data Cube vocabulary to describe these operations as flexible services that are composed by matching multidimensional data structures interfaces, and validated this model on a production repository containing 900 BI documents decomposed into 8000 documents snippets.

The underlying sequence of operations specific to each snippet was then extracted into a unique RDF graph. Aggregate SPARQL queries allow us to compute basic usage statistics for BI operations that can feed recommender systems such as BI workflows wizards. Besides refining the proposed model, next steps include evaluating the technical usability of SPARQL property paths patterns for data lineage and to identify frequent patterns in sequences of BI operations.

This PhD Thesis is done with a CIFRE industrial grant from SAP Research.

6.2. Interaction Design

6.2.1. Question Answering over Linked Data

Participant: Elena Cabrio.

While an increasing amount of semantic data is being published on the Web, the crucial issue of how typical Web users can access this body of knowledge comes to light. This PostDoc project focuses on the development of methods for a flexible mapping between questions in natural language, and data objects. The main purpose is to allow an end user to submit a query to an RDF triple store in English and get results in the same language, masking the complexity of SPARQL expressions and RDFS/OWL inferences involved in the resolution, but at the same time profiting from the expressive power of these standards. In particular, we address the problem of automatic identification of the relevant relations in Question Answering (QA), to capture the context in which the requests should be interpreted, to be able to determine the constraints on the database query.

We aim at investigating the applicability of the Textual Entailment (TE) approach, recently proposed as a general framework for applied semantics, where linguistic objects are mapped by means of semantic inferences at a textual level [55]. According to such framework, entailment relations can be detected between an input question and a set of relational patterns that represent possible lexicalizations of the relations of interest. Such relations, collected in a pattern repository, can be associated to a SPARQL query to the database. A TE system should therefore first try to establish an entailment relation between an input question and each of the relations in the pattern repository. Then, the SPARQL queries associated to the relations for which the entailed patterns have been found are composed in a single query to the database.

---

Since this PostDoc research work has just started, our early efforts were directed toward the study of the state of the art on QA over Linked Data. We are currently carrying out a feasibility study on the extraction of the relational patterns from Wikipedia (as the source of free text) and the use of DBpedia\(^\text{11}\) as a linked data resource. For the experimental part, we are considering energy and environment as the reference scenario.

### 6.2.2. Mobile Access to the Web of Data

**Participants:** Luca Costabello, Fabien Gandon.

This thesis, directed by F. Gandon and I. Herman (CWI and Semantic Web Activity Lead at W3C) deals with accessing the Web of Data from mobile environments. The first year addressed the multi-faceted relationship between ubiquitous consumption of Linked Data and mobile context. More specifically, focus has been put on RDF adaptive representation and on context-aware SPARQL endpoints access control.

When accessed from devices immersed into ubiquitous environments, RDF resources must be properly adapted to the mobile context in which the consumption is performed. A domain-independent, lightweight vocabulary for displaying Web of Data resources in mobile environments has been designed (PRISSMA, Presentation of Resources for Interoperable Semantic and Shareable Mobile Adaptability [36]). The vocabulary is the first step towards an adaptive rendering engine for RDF data coupled with a declarative framework to share and re-use presentation information for context-adaptable user interfaces for Linked Data.

Another line of research regards the role of mobile context in restricting access to the Web of Data. Ubiquitous connectivity enables new scenarios in consuming Linked Data and access control in such pervasive environments must not ignore the mobile context in which RDF consumption takes place, as uncontrolled access in given situations may be undesired by data providers. The work led to enhance the access control framework for SPARQL endpoints proposed by teammate Serena Villata (see 6.2.3) with the notion of mobile context provided by PRISSMA.

### 6.2.3. Access Control for the Web of Data

**Participant:** Serena Villata.

This research activity is mainly focussed on the field of Knowledge Representation. First, we have proposed a new access control model for the Web of Data and the Social Web. In particular, we have introduced the S4AC ontology\(^\text{12}\) where the meaning of the access policies and their components is defined. This access control model proposes, also, a contextual evaluation of the accessors’ information. This model has been applied both to the world of Linked Data and to the world of social networks. This research activity has been addressed in the context of the DataLift ANR project [21], [20].

Second, we have continued a research activity in the area of argumentation theory. In this context, we are exploring the use of argumentation theory for modeling trust in those systems which deal with incomplete knowledge, and for providing explanations about the agents’ choices [22], [19], [23], [25].

### 6.2.4. ISICIL

**Participants:** Nicolas Delaforge, Fabien Gandon.

As the leading team of the ISICIL project, we have developed many software components (client-side and server-side) to enrich the ISICIL platform. First, the whole server mechanism was redesigned, in order to improve the server performance, to strengthen and modularize the framework as well. Many semantic REST services were added (activity stream, syndication, subscription/notification, graphs and charts visualizations).

In collaboration with Erwan Demairy, in charge of the SegViz ADT, a Gephi-ISICIL connector was implemented, allowing ISICIL users to visualize the results of their SPARQL queries directly into a dynamic graph. A demo of this work was presented during the ISICIL public seminarium in September. Furthermore, projects such as Datalift and ISICIL had brought out the need of an access control model for the Web of Data. For this purpose, we designed the S4AC model and ontology and we realized a prototype to evaluate it based on the ISICIL dataset Figure 4.

---

\(^{11}\) [http://dbpedia.org](http://dbpedia.org)

\(^{12}\) [http://ns.inria.fr/s4ac/](http://ns.inria.fr/s4ac/)
Since the Philoweb conference in 2010, a workshop dedicated to philosophical engineering was attached to the French IC conference in Chambery. We presented there the advancement of a brand new bookmark model called Webmarks which semantically models the user interest on a web resource (Figure 5 & 6). A long paper on this work was accepted in the EGC 2012 conference and will be published in the RNTI journal (Hermann editions) [27].

We also collaborate with the I3S team in a task of semantization of a commercial wiki, called Mindtouch. This wiki is enhanced with semantic description of its content (Figure 7), its users are part of the ISICIL social network and their activities on the wiki are reported on the metadata server. This tool represents the editorial layer of the software bundle developed to improve the business intelligence tasks. This work was also accepted as a long paper in the EGC 2012 conference [26].

Figure 4. S4AC Access Policy Editor

6.2.5. Models and Methods for Representing and Identifying Groups of Individuals and Their Activities

6.2.5.1. Models and Methods for Representing and Identifying "Collective Personas"

Participant: Alain Giboin.

Context of the work: ISICIL project.

As opposed to Individual Personas (which are user models represented as specific, realistic humans), Collective Personas are models representing specific, realistic groups of people as such (e.g., teams, communities). Collective Personas are aimed to design groupware more closely adapted to groups. In 2010, we updated our review of the existing methods for elaborating collective Personas. This year, we published the updated review [17].
Figure 5. Webmark GUI
Figure 6. Webmark Model
Figure 7. SweetDeki Resource Model
6.2.5.2. Models and methods for Representing and Identifying Relationships between Individuals

**Participants:** Alain Giboin, Neji Bouchiba.

Context of the work: AVISICIL project, in collaboration with researchers from the Kewi team (I3S, UNS) and from the Laboratoire de Psychologie Cognitive et Sociale (UNS) who are involved in affective computing design projects (designing systems intended to help elderly people maintain their relationships, or autistic children to build relationships with others).

Digital technologies have been claimed to contribute to prevent elderly people from social isolation or loss of social ties. For example, ubiquitous computing, online social networking and affective computing have been reported to facilitate social interaction [64] or to enhance social connectedness [61] among the elderly. Participating to a project aimed to design a system for recognizing, through various sensors, the affective states (emotions) that indicate a loss or maintenance of social ties, we conducted a social ergonomic study to provide elements of design and evaluation of such a system. Noting that depressive states are among the most significant signs of an actual or potential loss of social ties (see, e.g., [65]), we focused the study in particular on: (a) the models describing the depressive states and the process of their recognition, and the links between these states and the state of social ties; (b) the sensors that can contribute to this recognition. In order to evaluate our solution (so-called GeREmo) with the elderly, we also identified, from an analysis of existing studies on the acceptability of digital technologies, criteria for assessing the acceptability of the GeREmo solution [50].

6.2.6. Comparing and Bridging Models of Shared Representations and Representation Sharing Processes

**Participant:** Alain Giboin.

Context of the work: GDR CNRS Psycho Ergo, Groupe thématique Coopération homme-homme et Coopération homme-machine. Action de recherche RefCom (Référentiel commun), co-led with Pascal Salembier (UTT).

Sharing representations or shared representations are often claimed to be a key factor for a collaboration to succeed. The notions of shared representations and representation sharing processes are examined in the research literature from several points of view; this variety of viewpoints gave rise to different conceptualizations, which are referred to using such terms as Common Frame of Reference, Mutual Intelligibility, Shared Context, Team/Situation Awareness, etc. In 2010, in order to achieve mutual intelligibility between researchers working on such conceptualizations, we elaborated and asked participants to the RefCom joint research action to test and to apply a grid for collaboratively comparing and bridging the conceptualizations (see Edelweiss activity report 2010 [13]). This year, we analyzed and reported the results of the test and application of the grid [39]. This resulted in a revision of the grid.

6.2.7. Frameworks for taking pragmatic dimensions of ontologies into account

**Participant:** Alain Giboin.

Context of the work: Follow-up to the Palette European project. This work was done in collaboration with the Centre de Recherche sur l’Instrumentation, la Formation et l’Apprentissage, ULg (Belgium).

When designing ontologies, ontologists (i.e., knowledge engineers specialized in ontology engineering) most often focus on the semantic dimensions of ontologies (such as expressiveness, level of granularity, etc.). Pragmatic dimensions, i.e. dimensions related to the context of use (including the purpose) of the ontologies, are often neglected whereas they are critical to users: ontologies indeed are used in context. In brief, pragmatic dimensions are not taken seriously into account when engineering ontologies but they have to.

---

13 http://raweb.inria.fr/rapportsactivite/RA2010/edelweiss/
We developed a framework to analyze the way we attempted, in the context of the Palette EU project, to contextualize the ontologies underlying a set of semantic knowledge services dedicated to communities of practice. The framework was derived from the Ontology Framework elaborated by members of the Ontology Engineering community during the Ontology Summit 2007 [62]. Both frameworks define a series of "pragmatic dimensions" of ontologies. Because our derived framework did not cover all possible dimensions, we complemented it, by relying on existing work from the Ontology Engineering community in general, and from the Pragmatic Web community in particular [16].

6.2.8. Explanation of Semantic Web Query Results

Participants: Rakebul Hasan, Fabien Gandon, Olivier Corby.

This PhD thesis, directed by Fabien Gandon and Olivier Corby, aims at opening the query-solving mechanism to the users, and handling and explaining the distribution of a query over several sources on the Semantic Web. This work is part of the Kollow ANR project.

The current Semantic Web search engines are not able to explain how a given query result is obtained or why it has failed to obtain a result. The goal of opening the query-solving mechanism is to enable the Semantic Web query engines to explain the query solving process taking into account the inferences used to obtain the results for a given query. In addition, explanation of the performance indicators of the query-solving process contributes to the understanding of the resolution process. These performance indicators can be effectively used to help in formulating queries by suggesting alternative queries based on the history of the performance of the query-solving process. Another focus of this thesis is on how the distribution of the queries can be performed over the distributed sources and how explanation can be used to better understand the queries and their results over the distributed sources.

In the early stage of this thesis, our current focus is on explaining the Semantic Web query results taking the inferences into account. We are working on justification of results for SPARQL query with RDFS entailment. Our next focus will be on the different abstractions of these justifications with different degree of details and different types of presentations depending on different level of user expertise.

6.2.9. Pervasive Sociality through Social Objects

Participants: Nicolas Marie, Fabien Gandon.

The work is related to semantic spreading activation algorithm, from idea to first results and visualization. Spreading activation is a method for searching semantic networks by labeling a set of initial nodes with weights (called activation), propagating (spreading) that activation out to other nodes linked to the source nodes and iterating propagation. Previously, at the end of 2010, we designed an ontology called OCSO [14]. This ontology aims at describing in a structured format social objects (content augmented by social functionalities independently of its nature: video, place, text, etc.) and corresponding social activity. Then, the need of powerful and semantic sensitive algorithm to process such data led us to follow the track of semantic spreading activation.

Two posters were published at IC [42] and Web Science [43] presenting OCSO model and research axis about semantic spreading activation. A state of the art about exploitation of semantics in spreading activation and its position in the general context of this algorithm family was written. Then a formal proposition was made and algorithm development started leading to first experimental results. The state of the art, the formal proposition and early results were published at the Social Objects workshop [41]. The end of the year was mainly focused about results visualization through Gephi and knowledge acquisition on algorithm and its behavior through multiple tests.

[14] http://ns.inria.fr/ocso/V0.2/
6. New Results

6.1. Note

Note that we do not include here the results from Souhila Kaci and Tjitze Rienstra since they joined the team in September 2011.

6.2. Ontological Query Answering with Rules

Participants: Jean-François Baget, Marie-Laure Mugnier, Michaël Thomazo, Michel Leclère, Eric Salvat, Mélanie König.

In collaboration with: Sebastian Rudolph (Karlsruhe Institute of Technology)

We have developed a framework based on rules that have the ability of generating unknown individuals, an ability sometimes called value invention in databases. These rules are of the form $body \rightarrow head$, where the body and the head are conjunctions of atoms (without function symbols except constants) and variables that occur only in the head are existentially quantified, hence their name existential rules hereafter. E.g., $\forall x (\text{Human}(x) \rightarrow \exists y (\text{isParent}(y,x) \land \text{Human}(y)))$. These rules can be seen as the logical translation of conceptual graph rules, historically a main focus of the team [70] [55]. Existential rules have the same logical form as the well-known Tuple-Generating Dependencies (TGDs) in databases [45]. TGDs have been extensively used as a high-level generalization of different kinds of constraints, e.g., for data exchange [57]. Recently, there has been renewed interest for TGDs seen as rules in the context of ontological query answering. Indeed, the value invention feature has been recognized as crucial in an open-world perspective, where it cannot be assumed that all individuals are known in advance. The deductive database language Datalog allows to express some ontological knowledge but it does not allow for value invention. This motivated the recent extension of Datalog to TGDs (i.e., existential rules), which gave rise to the Datalog +/- family [52], [53], [54]. In KRR and in the Semantic Web, ontological knowledge is often represented with formalisms based on description logics (DLs). However, DLs traditionally focused on reasoning tasks about the ontology itself (the so-called TBox), for instance classifying concepts; querying tasks were restricted to ground atom entailment. Conjunctive query answering with classical DLs has appeared to be extremely complex, hence less expressive DLs more adapted to conjunctive query answering on large amounts of data have been designed recently, namely DL-Lite [51], $\mathcal{EL}$ [41], [63], and more generally Horn DLs (see e.g., [60]), cf. also the tractable profiles of the Semantic Web language OWL2. Existential rules cover the core of lightweight DLs dedicated to query answering, while being more powerful and flexible [53], [44],[21]. In particular, they have unrestricted predicate arity (while DLs consider unary and binary predicates only), which allows for a natural coupling with database schemas, in which relations may have any arity; moreover, adding pieces of information, for instance to take contextual knowledge into account, is made easy by the unrestricted predicate arity, since these pieces can be added as new predicate arguments.

Building on our previous work on conceptual graphs, while meeting this new trend, we have developed a knowledge representation framework centered on existential rules, which can be seen both as logic-based and graph-based.

Entailment, hence query answering, with existential rules is not decidable, thus finding decidable classes of rules as expressive as possible is a crucial issue. We have pursued our previous work on better understanding the border between decidability and undecidability. We have also extended rule dependency to k-dependency, which takes into account sequences of rule applications.

- Results published in Artificial Intelligence Journal [13] (extending the work in [3], [44]);
- keynote talk synthesizing this work at RR’2011 [20]; extension to k-dependency at RR’2011 [22]
For newly exhibited decidable classes (namely, “frontier-one”, “frontier-guarded” and “weakly-frontier-
guarded” rules), the problem complexity was unknown, moreover there was no algorithm for computing entailment. First, we have classified these classes with respect to combined complexity (i.e., usual complexity) with both unbounded and bounded predicate arity, and data complexity (i.e., restricting the input of the decision problem to the facts). An interesting result is that some of the new classes (namely frontier-one and frontier-
guarded rules) have a polynomial time data complexity. Secondly, we have provided a generic algorithm for query entailment with a large class of rules including these classes, which is worst-case optimal for combined complexity (with or without bounded predicate arity) as well as for data complexity.

- Results partially published at IJCAI'2011 [21]. Long paper in preparation with extended com-
plexity results and all proofs, for submission to a major artificial intelligence journal.

6.3. Processing Conjunctive Queries with Negation

Participants: Marie-Laure Mugnier, Michel Leclère, Khalil Ben Mohamed, Michaël Thomazo.

Conjunctive queries have long been recognized as the basic queries in database and knowledge-based systems. The fundamental decision problems on these queries, namely query inclusion checking (given two queries $q_1$ and $q_2$, is $q_1$ included in $q_2$, i.e., is the set of answers to $q_1$ included in the set of answers to $q_2$ for all databases) and query entailment (is a given query entailed by the database) are NP-complete. When atomic negation is added to queries and databases, these problems become $\Pi^P_2$-complete (with the open world assumption for the query entailment problem). Note that these problems can be recast as entailment in the FOL fragment of existentially closed conjunction of literals (without function symbols except constants). On the one hand, we have led a theoretical complexity study: we have investigated the role of pairs of literals called “exchangeable” (which generalizes the notion of unifiable literals) in the complexity increase. The main results are that when the number of exchangeable pairs is bounded, say by $k$, then the complexity falls from $\Pi^P_2$-complete to $P^{NP}-$complete for any $k \geq 3$, and is NP-complete for $k \leq 1$ (the case $k = 2$ being open).

In collaboration with: Geneviève Simonet (LIRMM Algeco team)

- Results partially published at Information and Computation.

On the other hand, we have proposed, refined and compared experimentally several algorithms. This study follows first results of us in [61] and is the core of Khalil Ben Mohamed’s PhD thesis defended in December 2010 [64].

- Results published DEXA 2011 [24] (extending our work in RFIA 2010 [48], DEXA 2010 [46], AIMSA 2010 [47]).

Let us point out that both theoretical and practical results still hold when the predicates are preordered, which allows to take very light ontologies into account, i.e., where concepts and relations are organized in a specialization preorder.

6.4. Argumentation Systems for Decision Making

Participants: Rallou Thomopoulos, Madalina Croitoru, Jérôme Fortin, Marie-Laure Mugnier.

In collaboration with: Joël Abecassis (IATE/INRA), Jean-Rémi Bourguet (UM3), Patrice Buche (IATE/INRA), Sébastien Destercke (IATE/CIRAD) Nir Oren (Univ. of Aberdeen, Scotland)

Scientific investigations in this axis are guided by applications of our partners in agronomy (IATE laboratory). Substantial part of the work has consisted of analyzing the proposed applications and the techniques they require in order to select appropriate applications with respect to our team project.
Argumentation is a reasoning model based on the construction and the evaluation of arguments. In his seminal paper, Dung has proposed an abstract argumentation framework [56]. In that framework, arguments are assumed to have the same strength. This assumption is unfortunately strong and often unsatisfied. Consequently, several generalizations of the framework have been proposed in the literature. In [49] and [50], we have led a comparative study of these generalizations. It clearly shows under which conditions two proposals are equivalent. We have also integrated those generalizations into a common more expressive framework.

An instantiation of Dung’s abstract framework with the conceptual graph framework has been proposed. This representation uses default conceptual graph rules, an extension of classical conceptual graph rules (equivalent to existential rules, see Axis 1) with Reiter’s defaults [67] allowing for non-monotonic reasoning, that we developed independently of the argumentation framework [42], [43]. In the conceptual graph representation, arguments are represented as nested graphs, attacks between arguments can be computed from the structure of arguments and default rules allow to compute several kinds of extensions (i.e., maximal sets of arguments jointly acceptable according to a given semantics).

This approach has been applied to agrifood chain analysis, which is a highly complex procedure since it relies on numerous criteria of various types: environmental, economical, functional, sanitary, etc. Quality objectives imply different stakeholders, technicians, managers, professional organizations, end-users, public organizations, etc. Since the goals of the implied stakeholders may be divergent, decision-making raises arbitration issues. Arbitration can be done through a compromise—a solution that satisfies, at least partially, all the actors—or favor some of the actors, depending on the decision-maker’s priorities. We have analyzed a case study concerning risks/benefits within the wheat-to-bread chain. It concerns the controversy about the possible change in the ash content of the flour used for commonly used French bread. Several stakeholders of the chain are concerned, in particular the Ministry of Health through its recommendations in a national nutrition and health program, millers, bakers and consumers.

As already pointed out, the proposed approach is novel both for theoretical and application aspects.

Results presented in [30], [28].

Let us mention additional results related to the applications in agronomy on decision making combining machine learning based on decision trees and ontologies [58], [30], as well as results obtained by our collaborators on semi-automatic data extraction from web data (tables), data reliability, and the representation and flexible querying of imprecise data with fuzzy sets [16], [14], [17], [26], [31], [25], [27], [33], [34]. These investigations are complementary to the above mentioned results on argumentation and generally relate to other aspects in the same applicative projects.

6.5. Semantic Data Integration

Participants: Michel Leclère, Michel Chein, Madalina Croitoru, Rallou Thomopoulos, Léa Guizol.

It often happens that different references (i.e. data descriptions), possibly coming from heterogeneous data sources, concern the same real world entity. In such cases, it is necessary: (i) to detect whether different data descriptions really refer to the same real world entity and (ii) to fuse them into a unique representation. Since the seminal paper [66], this issue has been been studied under various names: “record linking”, “entity resolution”, “reference resolution”, “de-duplication”, “object identification”, “data reconciliation”, etc., mostly in databases (cf. the bibliography by William E. Winckler 1). It has become one of the major challenges in the Web of Data, where the objective is to link data published on the web and to process them as a single distributed database. Most entity resolution methods are based on classification techniques; Fatiha Saiṣ, Nathalie Pernelle and Marie-Christine Rousset proposed the first logical approach [68]. Many experiments on public data are underway, in France (cf. DataLift2 and ISIDORE3 projects) or internationally (e.g., VIAF project4 led by OCLC5, whose aim is to interconnect authority files coming from 18 national organizations).

2 DataLift, http://datalift.org/
3 ISIDORE, http://www.rechercheisidore.fr/
4 The Virtual International Authority File, http://www.oclc.org/research/activities/viaf/
Two years ago, we began a collaboration with ABES (National Bibliographic Agency for Universities, which takes part in the VIAF project). The aim of this collaboration is to enable the publication of ABES metadata bases on the Web of Data and to provide an identification service dedicated to bibliographic notices. ABES bibliographic bases, and more generally document metadata bases, appear to be a privileged application domain for the representation and reasoning formalisms developed by the team. This work has an interdisciplinary dimension, as it also requires experts in the Library and Information Science domain. We think that a logical approach is able to provide a generic solution for entity resolution in document metadata bases, even though it is generally admitted in Library and Information Science that “there is no single paradigmatic author name disambiguation task—each bibliographic database, each digital library, and each collection of publications, has its own unique set of problems and issues” [69].

6.5.1. SUDOC Metadata Formalization

The first step of collaboration with ABES was to formalize the SUDOC catalogue, which contains all French academic libraries bibliographic notices, into a knowledge base using a suitable knowledge representation and reasoning language. This required to first analyze SUDOC content, as well as document description standards (CRM-CIDOC, FRBR, Dublin Core). We then designed an ontology expressed in the Semantic Web languages RDFS + OWL, compatible with document description standards, as well as translations from any SUDOC set of notices into a set of RDF facts according to this ontology. These translations have been implemented, which allows to export SUDOC bases into Semantic Web formats. Moreover, using the RDFS to CG second translation mentioned above, we are now able to import SUDOC bases into our tools CoGUI + CoGITaNT.

▶ Technical report [40].

6.5.2. Implementation of an Entity Identification Service

In order to perform entity resolution (for entities restricted to "authors" for now), we have defined a set of rules allowing to enrich Sudoc descriptions; then, using enriched descriptions, authors can be classified according to a proximity criterion. A prototype providing this service has been implemented on top of Cogui. Experiments are currently led in the context of the SudocAd project jointly conducted by ABES and GraphIK. SudocAd aims at enriching the author field of a bibliographic record describing a document with links to Sudoc authorities referring to the authors of the target document. A general description of the implemented approach, an analysis of this approach on a representative sample of bibliographic records and first results on 13400 bibliographic records extracted from a corpus independent from Sudoc catalog are presented in the final report of SudocAd.

▶ Link to SudocAd Final Report: http://www.abes.fr/Media/Fichiers/Sudoc-Fichiers/SudocAD_rapportFinal

Finally, we have defined an extension of our own logical framework (existential rules, constraints, homomorphism-based mechanisms) based on Hector J. Levesque and Gerhard Lakemeyer’s Standard Names [62], and the notion of knowledge base faithfulness with respect to the entity resolution problem (intuitively, the fact that the knowledge base is non-ambiguous). This is still ongoing work.

▶ Research Report [38].
6. New Results

6.1. Introduction

This year Pulsar has tackled several issues related to its two main research axes: scene understanding for activity recognition and software engineering for activity recognition.

6.1.1. Scene Understanding for Activity Recognition


This year Pulsar has proposed new algorithms in computer vision (people head and face detection and people re-identification), in reasoning (activity recognition and uncertainty handling). More precisely, the new results for this research axis concern:

- People detection in monocular video sequences (6.2)
- Online Parameter Tuning for Object Tracking Algorithms (6.3)
- Fiber Based Video Segmentation (6.4)
- Multiple Birth and Cut Algorithm for Multiple Object Detection (6.5)
- Exhaustive Family of Energies Minimizable Exactly by a Graph Cut (6.6)
- Steepest Descent in Banach Spaces with Application to Piecewise-Rigid Evolution of Curves (6.7)
- Object Tracking Using a Particle Filter based on SIFT Features (6.8)
- Human Re-identification using Riemannian Manifolds (6.9)
- Global Tracking of Multiples Actors (6.10)
- Crowd Data Collection from Video Recordings (6.11)
- Events Recognition and Performance Evaluation (6.12)
- Group interaction and group tracking for video-surveillance in underground railway stations (6.13)
- Action Recognition in Videos (6.14)
- Activity Recognition Applied on Health Care Application (6.15)
- A Cognitive Vision System for Nuclear Fusion Device Monitoring (6.16)
- Scenario Recognition with depth camera (6.17)
- Trajectory Clustering for Activity Learning (6.18)

6.1.2. Software Engineering for Activity Recognition

This year Pulsar has improved the SUP platform. This latter is the backbone of the team experiments to implement the new algorithms proposed by the team in perception, understanding and learning. We improve our meta-modeling approach to support the development of video surveillance applications based on SUP. We continue the development of a scenario recognition module relying on formal methods to support activity recognition in SUP platform. We also continue to study the definition of multiple services for device adaptive platform for scenario recognition. Finally, we are implementing the new theoretical results obtained last year to improve the Clem toolkit.

The new results related to this research axis concern:

- SUP Software Platform (6.19)
- Model-Driven Engineering and Video-surveillance (6.20)
- Scenario Analysis Module (6.22)
- Multiple Services for Device Adaptive Platform for Scenario Recognition (6.23)
- The Clem Toolkit (6.24)

6.2. People detection in monocular video sequences

Participants: Etienne Corvee, François Brémond, Silviu-Tudor Serban, Vasanth Bathrinaryanan.

A video understanding system analyzes human activity by detecting people in video sequences and tracking their displacement and movement throughout the sequences. The better the detection quality, the higher the semantic level of the information is. People activity can differ greatly from one application to another e.g. the presence of a person in one zone can simply be detected from a moving pixel region in a manually specified zone whereas detecting people fighting in a subway requires more complex information. For people activity to be recognized, one needs to detect people accurately in videos and at real time frame rate. Current state of the art algorithms provide generic people detection solutions but with limited accuracy. In the people monitoring domain, although cameras remain mostly fixed, many issues occur in images. For example, outdoor scenes display strong varying lighting conditions (e.g. sunny/cloudy illumination, important shadows), public spaces can be often crowded (e.g. subways, malls) and images can be obtained with a low resolution and can be highly compressed. Hence, detecting and tracking objects in such complex environment remains a delicate task to perform. In addition, detecting people has to face one major difficulty which is caused by occlusion where important information is hidden. When people overlap onto the image plane, their foreground pixels cannot be separated using a standard thresholding operation from a background reference frame. Therefore vision algorithms need to use information held by the underlying pixels and located at specific locations such as body parts.

We have extended our work by implementing and testing a novel people, head and face detection algorithm using Local Binary Pattern based features and Haar like features. The traditional and efficient Adaboost training scheme is adopted to train object features from publicly available databases. This work has been published in the ICVS Workshop [36].

The work has been tested for group tracking in Vanaheim videos (see section 8.2.1.2) and for people tracking in VideoId videos. The VideoId project aims to re-identify people across a network of non overlapping cameras using iris, face and human appearance recognition. An example of tracked people, head and faces in a testing database is shown in figure 7. An example of re-identified face is shown in figure 8 by the VideoId interface in a Paris underground video.

We have evaluated our people detection algorithm on the test human dataset provided by INRIA against state of the art algorithms which we refer as HOG [59] and LBP-HOG [77]. The INRIA human dataset is composed of 1132 human images and 453 images of background scenes containing no human. The results are displayed in figure 9 which shows that we obtain slightly better performances than the HOG-LBP technique in terms of missed detection rate vs. FPPI i.e. False Positive Per Image. In this figure, two extreme functioning modes could be chosen: approximately 2 noisy detections are obtained every 1000 background images for 50% true positive detections or 1 noisy detection every 2 frames for a detection rate of approximately 88%.
Figure 7. Example of tracked people, head and face

Figure 8. Face recognition in Paris underground
The same evaluation scheme of people detection above is used for face detection evaluation. The FPPI rates are obtained on 997 NICTA [66] background images of 720x576 pixels. 180 faces provided by a CMU test face image database are used to evaluate true positive rates. We have compared our results with the 2 versions of Haar feature provided by the OpenCv library i.e. the standard ‘default’ and alternative ‘alt’ training parameters. The results in table 1 show that the Haar ’alt’ technique performs better than the traditional Haar one. And our haar based technique called CCR provides similar face detection rates while giving a less false alarm rate. The proposed approach is approximately 1% less successful in detecting faces than the Haar technique while this latter is 32% more noisier than our CCR technique.

<table>
<thead>
<tr>
<th>technique</th>
<th>TP(%)</th>
<th>FPPI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haar (default)</td>
<td>91.57</td>
<td>4.132</td>
</tr>
<tr>
<td>Haar (alt)</td>
<td>92.13</td>
<td>1.685</td>
</tr>
<tr>
<td>CCR</td>
<td>91.01</td>
<td>1.274</td>
</tr>
</tbody>
</table>

6.3. Online Parameter Tuning for Object Tracking Algorithms

Participants: Duc Phu Chau, Monique Thonnat, François Brémond.

Many approaches have been proposed to track mobile objects in a scene. However the quality of tracking algorithms always depends on scene properties such as: mobile object density, contrast intensity, scene depth and object size. The selection of a tracking algorithm for an unknown scene becomes a hard task. Even when the tracker has appropriately selected, it is difficult to tune online its parameters to get the best performance.

Therefore we propose a new control approach for mobile object tracking. More precisely in order to cope with the tracking context variations, this approach learns how to tune the parameters of object appearance-based tracking algorithms. The tracking context of a video sequence is defined as a set of features: density of mobile objects, their occlusion level, their contrasts with regard to the background and their 2D areas. In an offline supervised learning phase, satisfactory tracking parameters are searched for each training video sequence. Then these video sequences are classified by clustering their contextual features. Each context cluster is associated with the learned tracking parameters. In the online control phase, two approaches are
proposed. In the first one, once a context change is detected, the tracking parameters are tuned for the new context using the learned values. In the second approach, the parameter tuning is performed when the context changes and the tracking quality (computed by an online performance evaluation algorithm [56]) is not good enough. An online learning process enables to update the context/parameter relations.

We have also proposed two new tracking algorithms to experiment the proposed control method. The first tracker relies on a Kalman filter and a global tracking which aims at fusing trajectories belonging to the same mobile object. This work has been published in [35]. The second tracker relies on the similarities of eight object descriptors (2D and 3D positions, area, shape ratio, HOG, color histogram, color covariance and dominant color) to build object trajectories. This work has been published in [34].

![Figure 10. (a) CARETEKER: Illustration of the Caretaker video; (b) CAVIAR: Illustration of the Caviar video](image)

The proposed controller has been experimented on a long, complex video belonging to the Caretaker European project 1 (see figure 10 (a)) and 26 videos of Caviar dataset 2 (see figure 10 (b)). For the Caretaker video, when the controller is used, the tracking quality increases from 52% to 78%. For the Caviar dataset, the experimental results show that the tracking performance increases from 78.3% to 84.4% when using the controller. The tracking results on Caviar videos with the proposed controller are as good as the ones obtained with manual parameter tuning.

### 6.4. Fiber Based Video Segmentation

**Participants:** Ratnesh Kumar, Guillaume Charpiat, Monique Thonnat.

The aim of this work is to segment objects in videos by considering videos as 3D volumetric data (space×time). Figure 11 shows 2D slices of a video volume. Bottom right corner of each figure shows the current temporal depth in the volume, while top right shows the X-time slice and bottom left shows Y-time slice. In this 3D representation of videos, points of static background form straight lines of homogeneous intensity over time, while points of moving objects form curved lines. Analogous to the fibers in MRI images of human brains, we name fibers, these straight and curved lines of homogeneous intensity. So, in our case, to segment the whole video volume data, we are interested in a dense estimation of fibers involving all pixels.

For the detection of these fibers, we use motion flow vectors and intensity correlation of 2D patches over time. As these techniques are not reliable everywhere in the image domain, we sort the fibers based on the reliability of the detections from these techniques. The subsequent goal is then to pick high ranked fibers to propagate motion information and boundary fronts to other regions of the 3D volume.

---

To reliably propagate information from a fiber, we express the reliability of detection of a fiber and the cost of propagation of information from it. The later can be based on a distance measure of a pixel from a fiber, while reliability of a fiber involves motion coherency, color homogeneity, duration along time axis etc.

Our work closely relates to [72]. A video is represented by a set of particles (trajectory of an image point sample). The algorithm then extends and truncates particle trajectories to model motion near occlusion boundaries.

Figure 12 shows some straight fibers found in a video volume. The reliability of these fibers is based on temporal length. Fibers which have temporal span same as that of the video are colored in green, while fibers which have temporal span of less than 10% of the video are colored blue. Red colored fibers have temporal length in between green and blue colored fibers.

6.5. Multiple Birth and Cut Algorithm for Multiple Object Detection

Participant: Guillaume Charpiat.

In collaboration with the Ariana team (Ahmed Gamal-Eldin, Xavier Descombes and Josiane Zerubia), we developed a new optimization method which we call Multiple Birth and Cut (MBC). It combines the recently proposed Multiple Birth and Death (MBD) algorithm and the Graph-Cut algorithm. MBD and MBC optimization methods are applied to energy minimization of an object based model, the marked point process. The most important advantage of the MBC over MBD is the reduction of number of parameters. By proposing good candidates throughout the selection phase in the birth step, the speed of convergence is increased. In
this selection phase, the best candidates are chosen from object sets by a belief propagation algorithm. The algorithm is applied on the flamingo counting problem in a colony [37], [26].

6.6. Exhaustive Family of Energies Minimizable Exactly by a Graph Cut

**Participant:** Guillaume Charpiat.

Graph cuts are widely used in many fields of computer vision in order to minimize in small polynomial time complexity certain classes of energies. These specific classes depend on the way chosen to build the graphs representing the problems to solve. We study here all possible ways of building graphs and the associated energies minimized, leading to the exhaustive family of energies minimizable exactly by a graph cut. To do this, we consider the issue of coding pixel labels as states of the graph, i.e. the choice of state interpretations. The family obtained comprises many new classes, in particular energies that do not satisfy the submodularity condition, including energies that are even not permuted-submodular.

We studied in details a generating subfamily, in particular we proposed a canonical form to represent Markov random fields, which proves useful to recognize energies in this subfamily in linear complexity almost surely, and then to build the associated graph in quasilinear time. We performed a few experiments to illustrate the new possibilities offered [33]. We have also started to use this technique to minimize exactly approximations of Markov random field energies instead of minimizing approximately the exact energies, by projecting energies on the family we know to solve globally efficiently.

6.7. Steepest Descent in Banach Spaces with Application to Piecewise-Rigid Evolution of Curves

**Participant:** Guillaume Charpiat.

This is joint work with Gabriel Peyrée (CNRS, Ceremade, Université Paris-Dauphine). We intend to favor piecewise-rigid motions, i.e. articulated movements, during shape evolutions, especially when computing morphings or image segmentation with shape prior. To do this, we first need a dissimilarity measure between shapes, whose gradient is meaningful. We formulate one using kernels and bistochastization.

The parameters of these kernels are automatically estimated in a fixed-point scheme that guarantees physical relevance, and the notion of bistochastization is extended to continuous distributions. Finally, piecewise rigidity is ensured during gradient descents by a change of the norm from which the gradient is derived. This norm is formulated so as to favor sparse second derivatives, which produces articulated movements without knowing by advance the location of the articulations.

The formula of the norm is actually elegantly simple, involving simple geometric quantities, derivatives, and the $L_1$ norm. Note that this norm does not derive from an inner product but defines a gradient in the sense of [5] as the minimizer of an energy. It turns out that in our case the energy defining the gradient is actually convex, and efficient mimimization follows.

6.8. Object Tracking Using a Particle Filter based on SIFT Features

**Participants:** Malik Souded, François Brémont.

The approach consists in detecting SIFT points of interest on the objects to track, calculating their SIFT descriptors, tracking these points with a particle filter, and finally achieving tracking process by linking them along the time with links which are weighted by measures on SIFT descriptors and reliability.

The main contributions in this work are on three points.

The first point consists in techniques of detection and selection of SIFT points, allowing a better distribution of points of interest on the target and allowing better management of partial occlusions, and secondly an optimized computing time thanks to the parallelization of these SIFT computation on modern processors (see figure 13).
Figure 13. SIFT Points detection and selection

The second point concerns the weighting of the particles during tracking. This is done with a combination of two kinds of information: the similarity measure of the SIFT descriptor and the state of motion of pixels corresponding to the particles. This allows more robust tracking of SIFT points regardless of the quality of the background subtraction providing the detected objects.

The last point concerns the selection of temporal links between tracked objects and detected ones. These links are selected according to their weight. The weight of each link is based on the proportion of common SIFT points to both objects (two successive images) potentially linked, and the reliability of each of these SIFT point. This reliability is calculated for each point by measuring the variation of the SIFT descriptor during the tracking time.

The occlusion management is performed using three types of information: SIFT descriptors used for tracking (matching after reappearance) the dominant colors of the object of interest and finally the width, height, and speed (in real world) of the object, which are learned in Gaussian models during the tracking (tracking being used with video cameras which have been calibrated), see figure 14 .

The approach was tested on 121 sequences of four different datasets: 80 sequences from CAVIAR , 34 sequences from ETISEO, 3 sequences from PETS2001 and 2 sequences from VS_PETS2003. The obtained results are satisfying. The comparison of these results with the state of the art shows improvements for the benchmarking dataset (ETISEO) The following table compares the proposed approach and state of the art results on ETISEO data base:

<table>
<thead>
<tr>
<th>Metrics</th>
<th>ETI-VS1-BE-18-C4</th>
<th>ETI-VS1-BE-16-C4</th>
<th>ETI-VS1-MO-7-C1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed tracker</td>
<td>M1</td>
<td>0.68</td>
<td>0.54</td>
</tr>
<tr>
<td>Chau et al. (VISAPP 2011)</td>
<td>M1</td>
<td>0.64</td>
<td>0.36</td>
</tr>
<tr>
<td>Best team in ETISEO Project (2005)</td>
<td>M1</td>
<td>0.48</td>
<td>0.44</td>
</tr>
</tbody>
</table>

This work was published in [49 ]

6.9. Human Re-identification using Riemannian Manifolds

Participants: Sławomir Bąk, Etienne Corvée, François Brémond, Monique Thonnat.

This work addresses the human re-identification problem, which is defined as a requirement to determine whether a given individual has already appeared over a network of cameras. This problem is particularly hard by significant appearance changes across different camera views. In order to re-identify people, a human signature should handle difference in illumination, pose and camera parameters.
We propose new appearance models based on the mean riemannian covariance (MRC) matrices combining the appearance information from multiple images. These mean covariance matrices not only keep information on feature distribution but also carry out essential cues about temporal changes of an appearance. Using MRC-s, we propose two methods for an appearance representation:

- **Learned Covariance Patches (LCP)** [25] - a distinctive representation is extracted by a boosting scheme. The structure of MRC patches (size, position) is learned using boosting algorithm based on confidence-rated predictions (see Fig. 15 (a)). These confidence-rated coefficients are employed to weight appearance characteristics of a specific individual w.r.t. the reference (training) dataset of humans.

- **Mean Riemannian Covariance Grid (MRCG)** [30] - less computationally demanding technique than LCP. We represent a human appearance by a grid of MRC cells (see Fig. 15 (b)). Relevant cells are identified by an efficient discriminant analysis. This analysis takes into account variance of MRC patch in the class of humans (reference dataset). MRC is assumed to be more significant when its variance is larger in the class of humans: (1) the most common patterns belong to the background (the variance is small); (2) the patterns which are far from the rest are at the same time the most discriminative (the variance is large). All operations, such as mean or variance, are performed on covariance manifold specified as Riemannian.

Both methods are evaluated and compared with the state of the art using publicly available datasets. We demonstrate that the proposed approaches outperform state of the art methods. Further, we extract new sets of individuals from i-LIDS data to investigate more carefully advantages of using many images for human re-identification.

The computation complexity is analyzed in the context of distance operator between two signatures. Comparing two human signatures, it is necessary to compute distance between covariance matrices, which requires solving the generalized eigenvalues problem. This operation is computationally heavy. In [31], we propose an implementation for finding generalized eigenvalues and eigenvectors for distance operator, using NVIDIA GPU architecture. We improve significantly the performance, reaching 66 speedup using Tesla S1070.
6.10. Global Tracking of Multiples Actors

Participants: Julien Badie, François Brémond.

We propose a new approach for long term tracking of individuals. Our main objective is to design a tracking algorithm for people reidentification [30] that can track people even if they come back in the scene after leaving it. This algorithm is based on covariance matrix and we have also added some contextual information of the scene (for instance, zones where people can enter the scene) to improve tracking performance. In addition, a basic noise detection system and a tracking correction system are proposed in order to handle short-term tracking errors such as multiplication of IDs corresponding to only one individual. The noise detection system is designed to find and remove objects that are detected in a very small number of consecutive frames (for instance 4) and disappear afterward. The tracking correction system associates IDs recently lost with IDs that have just started to be tracked based on geometrical features and 3D distance criteria.

As a result, the tracking quality is significantly improved on 5 video sequences tested from the ETISEO dataset 3. The people reidentification algorithm gives encouraging results for future work. The number of IDs associated to one individual is reduced (on average 50% less) and the tracking quality improves due to the IDs stability. This algorithm can detect not only people re-entering the scene but also trajectory interruptions due to occlusions or misdetections.

This approach could enable the detection of new kinds of events on video sequences such as long range people tracking on a camera network.

6.11. Crowd Data Collection from Video Recordings

Participants: Jihed Joobeur, François Brémond.

The aim of this work is to analyze crowd behaviors by detecting specific situations: panic, congestion, fighting etc. We validate our work with subway station videos from VANAHEIM project. We use Mixture of Gaussian based segmentation to extract moving point and then detecting moving objects. Subsequently inside these moving objects we detect FAST feature points and compute HOG descriptors for tracking these points. We compute different features based on these points like speed and orientation. To estimate the crowd density we use features based on Grey-Level Co-occurrence Matrix. As these features depend on the distance of people from the camera, we divide the scene into different zones which have each zone same distance from the camera.

---

3 http://www-sop.inria.fr/orion/ETISEO/
Figure 16. The green lines are the ground-truth and the blue lines are the tracking results using both reidentification and noise removal (left) and with only noise removal (right). The man is tracked throughout the whole video with a single ID (68).
camera. In each area, compiling all the information on speed, direction and learned over a threshold density of the crowd, we can learn and detect different situations. For example, if the density increases and the average speed decreases in a pre-defined zone, that may correspond to a congestion situation.

On figure 17 the FAST feature points are shown in blue points, while the tracking of these points is shown in yellow.

Figure 17. Feature points detection and tracking in different zones.


Participants: Ricardo Cezar Bonfim Rodrigues, François Brémond.

The goal of this work is to evaluate the accuracy and performance of events detection, see workflow in Figure 18. The experiments will be performed using the tools developed in Pulsar team, such as Scene Understand Platform (SUP) 4 a plugin for events detection [50] and ViseVal.

The experiments were performed using video sequences of a subway station (VANAHEIN dataset) where the goal was to detect events such as people waiting, entering, buying tickets and so on. Preliminary results showed a very low accuracy and demonstrated that the scenario configuration parameters are very sensitive in this problem. It means many of the expected events were missed or misclassified, specially composite events (when more than one activity recognition is required) see some issues on Figure 19.

Based on the issues, a second experiment was configured using 3 different video sequences. In this new experiment, the scenario was adjusted to give more tolerances to people detection issues, the camera calibration was refined and some events were remodeled. After these changes the results were significantly improved. This last experiments showed that the engine proposed by Pulsar team is able to detect events accurately however events modeling can be very sensitive to the scenario configuration, see the results in Table 3.

Table 3. Results of events detection based on the people detection performed in experiment 2 using an IDIAP algorithm. The global Precision and sensibility of people detection are respectively 0.91 and 0.95.

<table>
<thead>
<tr>
<th></th>
<th>Sequence 1</th>
<th>Sequence 2</th>
<th>Sequence 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision (global)</td>
<td>0.73</td>
<td>1.00</td>
<td>0.88</td>
</tr>
<tr>
<td>Sensitivity (global)</td>
<td>0.82</td>
<td>0.90</td>
<td>0.85</td>
</tr>
</tbody>
</table>

4 http://www.ebi.ac.uk/pulsar/uid27.html
5 http://www-sop.inria.fr/teams/pulsar/EvaluationTool/ViseVal_Description.html
Figure 18. Events detection workflow

Figure 19. Blue bounding boxes correspond to Annotations (Reference data) and red detected objects. It's possible to observe if there is a detection issue on the left image.
6.13. Group interaction and group tracking for video-surveillance in underground railway stations

Participants: Sofia Zaidenberg, Bernard Boulay, Carolina Garate, Duc-Phu Chau, Etienne Corvée, François Brémond.

One goal in the European project VANAHEIM is the tracking of groups of people. Based on frame to frame mobile object tracking, we try to detect which mobiles form a group and to follow the group through its lifetime. We define a group of people as two or more people being close to each other and having similar trajectories (speed and direction). The dynamics of a group can be more or less erratic: people may join or split from the group, one or more can disappear temporarily (occlusion or disappearance from the field of view) but reappear and still be part of the group. The motion detector which detects and labels mobile objects may also fail (misdetections or wrong labels). Analyzing trajectories over a temporal window allows handling this instability more robustly. We use the event-description language described in [50] to define events, described using basic group properties such as size, type of trajectory or number and density of people and perform the recognition of events and behaviors such as violence or vandalism (alarming events) or a queue at the vending machine (non-alarming events). Two approaches to this problem have been implemented. The first approach takes as input the frame-to-frame tracking results of individual mobiles and tries to gather them into groups based on their trajectories through the temporal window. Each group has a coherence coefficient. This coefficient is a weighted sum of three quantities characterizing a group: the group density (average of distances between mobiles), the similarity of mobile’s speed and the similarity of their motion directions. The update of a group consists in re-calculating the group coherence with new mobiles from the current frame. If adding the mobile does not put the coherence under a defined threshold, the mobiles are added to the group. A pre-selection is made by only considering mobiles that are close enough to the center of gravity of the group. After the update step, all mobiles that have not been assigned to a group are analyzed to form new groups if possible.

A first improvement has been done by integrating the use of the LBP-based people detector described in [36]. This makes the algorithm more robust to false detections such as train doors closing. But on the other hand, it also introduces false negatives as, among other things, people are only detected if fully visible in the image. The group tracking algorithm has been tested both with the original, background subtraction-based mobile object detection (noted S hereafter) and the LBP-based people detection (noted LBP hereafter).

For evaluating the detection, we used 3 annotated sequences: Sequence 1 is a short sequence of 128 frames with just one ground truth object (one group), Sequence 2 has 1373 frames and 9 ground truth objects, and Sequence 3 is 17992 frames long and 25 ground truth objects were annotated. Detection and tracking results are shown in table 4.

The whole algorithm chain has been integrated into the common VANAHEIM platform and sent to partners for pre-integration.

We also used videos from the ViCoMo project, recorded in the Eindhoven airport to test our approach. No formal evaluation has been done yet on these sequences due to the lack of ground truth. Nevertheless, these videos contain several acted scenes which could be successfully recognized: groups merging, splitting and entering a forbidden zone.

This work has been published in [50].

In parallel, a new approach is being developed, making use of a long-term tracker described in [35]. This tracker provides more robust individual trajectories to the group tracker, containing less confusions in cases where people cross each other. We apply the Mean Shift clustering algorithm on trajectories of people through a sliding time window (e.g. 10 frames). If the target is lost in one or several frames, we interpolate its positions. The clustering brings together mobiles having similar trajectories, which is our definition of a group. At each frame, clusters are calculated and then a matching is done to associate clusters to existing groups in the previous frame, and thus track groups. Looking backwards (within a window) on the trajectory of a mobile we might find a mobile on that trajectory that belongs to a group. If such a group is found, it is called the probable
<table>
<thead>
<tr>
<th></th>
<th>Sequence 1</th>
<th></th>
<th>Sequence 2</th>
<th></th>
<th>Sequence 3</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>S</td>
<td>HD</td>
<td>S</td>
<td>HD</td>
<td>S</td>
<td>HD</td>
</tr>
<tr>
<td>True Positives (TP)</td>
<td>72</td>
<td>67</td>
<td>1395</td>
<td>1079</td>
<td>5635</td>
<td>3679</td>
</tr>
<tr>
<td>False Positives (FP)</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td>111</td>
<td>1213</td>
<td>642</td>
</tr>
<tr>
<td>False Negatives (FN)</td>
<td>6</td>
<td>11</td>
<td>269</td>
<td>585</td>
<td>3686</td>
<td>5642</td>
</tr>
<tr>
<td>Precision (global)</td>
<td>1</td>
<td>1</td>
<td>0.99</td>
<td>0.90</td>
<td>0.82</td>
<td>0.85</td>
</tr>
<tr>
<td>Sensitivity (global)</td>
<td>0.92</td>
<td>0.84</td>
<td>0.83</td>
<td>0.65</td>
<td>0.60</td>
<td>0.40</td>
</tr>
<tr>
<td>Tracking confusion</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.99</td>
<td>0.92</td>
<td>0.96</td>
</tr>
</tbody>
</table>

Of the current mobile. Each trajectory cluster is associated to the group that is the probable group of most mobiles in the cluster. Several clusters may be associated to the same group. This cluster association makes the algorithm robust to cases where one or several mobiles temporarily separate from the group. If the separation is longer than the time window, the probable group of these mobiles will be empty and a split will be detected.

Additionally, we work on improving the people detection by combining both methods: background subtraction-based and LBP-based. We compare overlapping mobiles from both methods and choose the best one based on their respective confidence values and their sizes. If a target was detected by only one of the two methods, we keep the target given that the confidence is high enough. If a mobile from the background subtraction method is big enough to cover several LBP-detected people (the LBP-based people detection output targets have the size of a human, whereas the background subtraction can detect a bigger mobile with the size of a GROUP_OF_PEOPLE), we attach the LBP-people as sub-mobiles of the group mobile so no information is lost. This method is a work in progress and no evaluation have been done yet.

Figure 20 shows two examples of group and event detection.


**Participants:** Piotr Bilinski, François Brémond.

The aim of this work is to learn and recognize short human actions in videos. We perform an extensive evaluation of local spatio-temporal descriptors, then we propose a new action recognition approach for RGB camera videos. We also propose a new approach for RGB-D cameras. For all our experiments, we develop an evaluation framework based on the bag-of-words model, SVM and cross-validation technique. We use the bag-of-words model to represent actions in videos and we use non-linear multi-class Support Vector Machines together with leave-one-person-out cross-validation technique to perform action classification.

Local spatio-temporal descriptors have shown to obtain very good performance for action recognition in videos. Over the last years, many different descriptors have been proposed. They are usually evaluated using too specific experimental methods and using different datasets. Moreover, existing evaluations make assumptions that do not allow to fully compare descriptors. In order to explore capabilities of descriptors, we perform an extensive evaluation of local spatio-temporal descriptors for action recognition in videos.

Four widely used state-of-the-art descriptors (HOG, HOF, HOG-HOF and HOG3D) and four video datasets (Weizmann, KTH, ADL and KECK) have been selected. In contrast to other evaluations, we test all the computed descriptors, we perform experiments on several codebook sizes and use several datasets, differing in difficulty. Our results show how the recognition rate depends on the codebook size and the dataset. We
observe that usually the HOG descriptor alone performs the worst but outperforms other descriptors when it is combined with the HOF descriptor. Also, we observe that smaller codebook sizes lead to consistently good performance across different datasets. This work has been published in [32].

We also propose a new action recognition method for RGB camera videos based on feature point tracking and a new head estimation algorithm. We track feature points along a video and compute appearance features (HOG-HOF) for each trajectory. Additionally, we estimate a head position for each visible human in the video, using the following chain: segmentation, person, head and face detectors. Finally, we create an action descriptor based on the combination of all these sources of information. Our approach has been evaluated on several datasets, including two benchmarking datasets: KTH and ADL, and our new action recognition dataset. This new dataset has been created in cooperation with the CHU Nice Hospital. It refers to people performing daily living activities like: standing up, sitting down, walking, reading a magazine etc.

We also study the usefulness of low-cost RGB-D camera for action recognition task. We propose a new action recognition method using both RGB and depth information. We track feature points using RGB videos and represent trajectories in a four-dimensional space using additionally depth information. Experiments have been successfully performed on our new RGB-D action recognition dataset, recorded using Microsoft’s Kinect device.

6.15. Activity Recognition Applied on Health Care Application

Participants: Rim Romdhane, Veronique Joumier, François Brémond.

The aim of this work is to propose a constraint-based approach for video event recognition with probabilistic reasoning for handling uncertainty. This work was validated on health care applications.

6.15.1. Event Recognition

We propose an activity recognition framework which is able to recognize composite events with complex temporal relationships. We consider different aspects of the uncertainty of the recognition during the event modeling and the event recognition process to overcome the noise or missing observations which characterize real world applications.
To reach this goal, we manage the uncertainty in the event modeling and event recognition processes by a combination of logical and probabilistic reasoning for handling uncertainty. We improve the event description language developed in Pulsar team and introduce a new probabilistic description based approach to gain in flexibility for event modeling by adding the notion of utility. Utility expresses the importance of sub-events to the recognition of the whole event. We compute the probability of recognition for both primitive (i.e. elementary) events and composite events based on Bayesian theory.

We compute the probability that the event $e$ is recognized given a sequence of observations $O$ as described in [48]. The observations consist of the set of the physical objects $p_{oe}$ moving in the scene. If the probability of an event is over a predefined threshold, the event is recognized.

6.15.2. Health Care Application

The proposed event recognition approach is validated using the videos from the health care application SWEETHOME (http://cmrr-nice.fr/sweethome/) and CIUSante (https://extranet.chu-nice.fr/ciu-sante). We have worked in close collaboration with clinicians from Nice hospital to evaluate the behaviours of Alzheimer patients. We have first model 69 event models for health care application using our event modeling formalism. With the help of clinicians we have established a scenario protocol. The scenario is composed of three parts: (1) directed activities (10 min), (2) semi directed activities (20 min), (3) free activities (30 min). Experiments have been performed in a room of Nice hospital equipped with 2 video cameras where 45 elderly volunteers have spent between 15 min to 1 hour. Volunteers include Alzheimer patients, MCI (mild cognitive impairment) and healthy elderly.

The study described in [38] and [27] shows the ability of the proposed automatic video activity recognition system to detect activity changes between elderly subjects with and without dementia during a clinical experimentation. A total of 28 volunteers (11 healthy elderly subjects, 17 Alzheimer’s disease patients (AD)) participate to the experimentation. The proposed study shows that we could differentiate the two profiles of participants based on motor activity parameters, such as the duration of the recognized activities, the strike length and the walking speed, computed from the proposed automatic video activity recognition system. These primary results are promising and validating the interest of automatic analysis of video as an objective evaluation tool providing comparative results between participants and over the time.


Participants: Guillaume Charpiat, Vincent Martin, François Brémond, Monique Thonnat.

In collaboration with Victor Moncada, Jean-Marcel Travere and Thierry Loarer (CEA Cadarache), we propose a cognitive vision-based system for the intelligent monitoring of tokamaks during plasma operation, based on multi-sensor data analysis and symbolic reasoning. The practical purpose is to detect and characterize in real time abnormal events such as hot spots measured through infrared images of the in-vessel components in order to take adequate decisions. Our system is made intelligent by the use of a priori knowledge of both contextual and perceptual information for ontology-driven event modeling and task-oriented event recognition. The system is made original by combining both physics-based and perceptual information during the recognition process. Real time reasoning is achieved thanks to task-level software optimizations. The framework is generic and can be easily adapted to different fusion device environments. The developed system and its achievements on real data of the Tore Supra tokamak imaging system can be found in [39].

6.17. Scenario Recognition with depth camera

Participants: Bernard Boulay, Daniel Zullo, Swaminathan Sankaranarayanan, François Brémond.

Thanks to Microsoft and its kinect sensor, RGB-depth camera becomes popular and accessible. The basic idea of depth camera is to combine a visible camera, with an IR camera associated to a laser to determine the depth of each image pixel. This kind of sensor is well adapted for applications which monitor people (e.g. monitoring Alzheimer patient in hospital): because the people are in a predefined area and near the camera.
The depth cameras have two main advantages: first, the output image contains depth information and second, the sensor is independent from the light changes (IR sensor).

In our work, we propose to use the kinect sensor to acquire 3D images, detect the people and recognize interesting activities (see Figure 21).

![Figure 21. People detection and activity recognition in day and night conditions with kinect sensor.](image)

The nestk library is used to manage the kinect sensor. This library is based on OpenNI framework (an open source driver) to acquire the images. Moreover the library is able to compute some treatments (e.g. people detection) and to provide a true 3D map of the scene in the referential of the kinect.

Basic attributes are computed for each detected person: 3D position, 3D height,... These attributes are then used to compute more complex information: speed, global posture to recognize interesting activities thanks to the ScreKs framework (scenario recognition based on expert knowledge) of the SUP platform as walking, stopping, standing, sitting,... (following a protocol delivered by doctors). Then we have a plug and play system able to recognize basic activities associated to a person.

Moreover, if information on the scene, as interesting zones, or equipment are available, complex activities can be recognized as nurse explaining exercise or nurse switching off the light.

The next step, is to use the human skeleton detection to recognize precisely the posture of the patient in order to understand more precise activities and infer a behaviour model.

### 6.18. Trajectory Clustering for Activity Learning

**Participants:** Jose Luis Patino, Guido Pusiol, Hervé Falciani, Nedra Nefzi, François Brémond, Monique Thonnat.

The discovery, in an unsupervised manner, of significant activities observed from a video sequence, and its activity model learning, are of central importance to build up on a reliable activity recognition system. We have deepened our studies on activity extraction employing trajectory information. In previous work we have shown that rich descriptors can be derived from trajectories; they help us to analize the scene occupancy and its topology and also to identify activities [67], [68], [70], [55]. Our new results show how trajectory information can be more precisely employed, alone or in combination with other features for the extraction of activity patterns. Three application domains are currently being explored: 1) Monitoring of elderly people...
at home; 2) Monitoring the ground activities at an airport dock-station (COFRIEND project \(^6\)); 3) Monitoring activities in subway/street surveillance systems.

### 6.18.1. Monitoring of elderly people at home

We propose a novel framework to understand daily activities in home-care applications; the framework is capable of discovering, modeling and recognizing long-term activities (e.g. “Cooking”, “Eating”) occurring in unstructured scenes (i.e. “an apartment”).

The framework links visual information (i.e., tracked objects) to the discovery and recognition of activities by constructing an intermediate layer of primitive events automatically.

The primitive events characterize the global spatial movements of a person in the scene (“in the kitchen”), and also the local movements of the person body parts (“opening the oven”). The primitive events are built from interesting regions, which are learned at multiple semantic resolutions (e.g. the “oven” is inside the “kitchen”). An example of the regions and possible activities for a single resolution is displayed in Fig. 22.

![Figure 22. Example of some learned regions and possible activities.](image)

A probabilistic model is learned to characterize each discovered activity. The modeled activities are automatically recognized in new unseen videos where a pop-up with a semantic description appears when an activity is detected. Examples of semantic labels are illustrated in Fig. 23 (a, b, c).

Recently we introduced 3D (MS. Kinect) information to the system. The preliminary results show an improvement superior to the 30% of the recognition quality. Also, the system can recognize activities in challenging situations as the lack of light. -See Fig.2 (b) and Fig.2 (c) -. The approach can be used to recognize most of the interesting activities in a home-care application and has been published in [43]. Other examples and applications are available online in [http://www-sop.inria.fr/pulsar/personnel/Guido.Pusiol/Home4/index.php](http://www-sop.inria.fr/pulsar/personnel/Guido.Pusiol/Home4/index.php).

### 6.18.2. Monitoring the ground activities at an airport dock-station

The COFRIEND project aims at creating a system for the recognition and interpretation of human activities and behaviours at an airport dock-station. Our contribution is a novel approach for discovering, in an unsupervised manner, the significant activities from observed videos. Spatial and temporal properties from detected mobile objects are modeled employing soft computing relations, that is, spatio-temporal relations graded with

different strengths. Our system works off-line and is composed of three modules: The trajectory speed analysis module, The trajectory clustering module, and the activity analysis module. The first module is aimed at segmenting the trajectory into segments of fairly similar speed (tracklets). The second aims at obtaining behavioural displacement patterns indicating the origin and destination of mobile objects observed in the scene. We achieve this by clustering the mobile tracklets and also by discovering the topology of the scene. The latter module aims at extracting more complex patterns of activity, which include spatial information (coming from the trajectory analysis) and temporal information related to the interactions of mobiles observed in the scene, either between themselves or with contextual elements of the scene. A clustering algorithm based on the transitive closure calculation of the final relation allows finding spatio-temporal patterns of activity. An example of discovery is given in the figure below. This approach has been applied to a database containing near to 25 hours of recording of dock-station monitoring at the Toulouse airport. The discovered activities are: ‘GPU positioning’, ‘Handler deposits chocks’, ‘Frontal unloading operation’, ‘Frontal loading operation’, ‘Rear loading operation’, ‘Push back vehicle positioning’. An example of discovered activity (Frontal loading) is given on the figure 24. When comparing our results with explicit ground-truth given by a domain expert, we were able to identify the events in general with a temporal overlap of at least 50%. The comparison with a supervised method on the same data indicates that our approach is able to extract the interesting activities signalled in the ground-truth with a higher True Positive Rate (74% TPR for the supervised approach against 80% TPR with our unsupervised method). This work has been published in [42]

**Figure 23. Examples of the recognized activities in 3D and under different light conditions.**

![Figure 23](image)

**Figure 24. Example of an activity cluster obtained. The left panel presents the tracklets of the mobiles participating in the Frontal Loading activity. Filled circles indicate the beginning of a tracklet. Empty circles indicate the end of a tracklet. The right panel presents the start frame of the activity.**

### 6.18.3. Monitoring activities in subway/street surveillance systems
In this work we have built a system to extract from video and in an unsupervised manner the main activities that can be observed from the a subway scene. We have setup a processing chain broadly working on three steps: The system starts in a first step by the unsupervised learning of the main activity areas of the scene. In a second step, mobile objects are then characterized in relation to the learned activity areas: either as ‘staying in a given activity zone’ or ‘transferring from an activity zone to another’ or a sequence of the previous two behaviours if the tracking persists long enough. In a third step we employ a high-level relational clustering algorithm to group mobiles according to their behaviours and discover other characteristics from mobile objects which are strongly correlated. We have applied this algorithm to two domains. First, monitoring two hours of activities in the hall entrance of an underground station and showed what are the most active areas of the scene and how rare/abnormal (going to low occupied activity zones) and frequent activities (e.g. buying tickets) are characterized. In the second application, monitoring one hour of a bus street lane, we were able again to learn the topology of the scene and separate normal from abnormal activities. When comparing with the available ground-truth for this application, we obtained a high recall measure (0.93) with an acceptable precision (0.65). This precision value is mostly due to the different levels of abstraction between the discovered activities and the ground-truth. The incremental learning procedure employed in this work is published in [52] while the full activity extraction approach was published in [41].

6.19. SUP Software Platform

Participants: Julien Gueytat, Leonardo Rocha, Daniel Zullo, François Brémond.

SUP is a software platform developed by PULSAR team, written in C and C++ for generating activity recognition systems. These systems should be able to perceive, analyze, interpret and understand a 3D dynamic scene observed through a network of sensors.

These activity recognition systems are a combination of algorithms developed by members of Pulsar or state of the art computer vision libraries. The SUP dissemination is targeted for use in real-world applications requiring high-throughput.

SUP is made as a framework allowing several computer vision workflows to be implemented. Currently, the workflow is static for a given application but our goal is to make it dynamic. A given workflow is the composition of several plugins, each of them implementing an algorithmic step in the video processing chain (i.e. the segmentation of images, the classification of objects, etc.). The design of SUP allows to execute at run-time the selected plugins.

During 2011 several tasks have been accomplished:

- A stable packaged release is available
- 3D simulation from a scenario description
- Existing algorithms have been improved in performance and accuracy
- Kinect sensor has been added to the hardware supported

Several plugins are available:

- 2 plugins are wrappers on industrial implementations of video processing algorithms (made available by Keeneo). They allow a quick deployment of a video processing chain encompassing image acquisition, segmentation, blob construction, classification and short-term tracking. These algorithms are robust and efficient algorithms, but with the drawback that some algorithms can lack accuracy.
- Several implementations by the Pulsar team members which cover the following fields:
  1. Image acquisition from different types of image and camera video streaming.
  2. Segmentation removing the shadows.
  3. Two classifiers, one being based on postures and one on people detection.
  4. Four frame-to-frame trackers, using as algorithm:
1. a simple tracking by overlapping,
2. neural networks,
3. tracking of feature points,
4. tracking specialized for the tracking of people in a crowd.

5. Three scenario recognizers, one generic algorithm allowing expression of probabilities on the recognized events, the second one focusing on the recognition of events based on postures and the third one (see section Extendable Event Recognition algorithm: SED in this document) uses the complete ontology of the domain as a parameter (e.g. the definition of objects of interest, scenario models, etc.).

6. 3D animation generation, it generates a virtual 3D animation from information provided by different plugins of the processing chain together with 3D contextual environment.

7. 3D simulation from description, it generates a virtual 3D animation from information provided from a text file with the description of the scenario.

From a software engineering point-of-view, the goal is to obtain a flexible platform being dynamically reconfigurable for the generated scene understanding systems to be autonomous and adaptable for handling changing environment.

SUP relies on DTK, a generic platform developed by the DREAM service at INRIA Research Center Sophia-Antipolis Méditerranée.

The purpose of DTK is to provide a software infrastructure allowing the generation of a new system by the composition of plugins, each plugin being an algorithmic step of the whole processing chain. SUP is oriented to help developers building activity recognition systems and describing their own scenarios dedicated to specific applications. By relying on the DTK software infrastructure, the possibilities are:

- To simplify the exchanges of algorithms between different INRIA teams using the DTK.
- To use the facilities already provided by the DTK allowing to compose quickly existing plugins. Currently a python interface is operational, and we plan to take advantage of the graphical composer to prototype quickly new workflows, or reconfigure existing ones, for the experimentation conducted by the team.

In order to be confident on the results obtained with the SUP platform, an important effort is done to check:

- The correct behavior of the platform from a software engineering point of view, i.e. that the functionality of the SUP software is correctly provided, or is not broken by modifications.
- A qualitative evaluation tool (see ViSEvAl in this document) for the algorithms, which compares and assesses the results obtained with the algorithms to ground truth for several reference videos.

Both kinds of test are performed on a daily basis and on several hardware/software architectures.

6.20. Model-Driven Engineering and Video-surveillance

Participants: Sabine Moisan, Jean Paul Rigault, Sagar Sen, François Brémond.

In the framework of our research on model engineering techniques for video-surveillance systems, we have focused this year on the runtime adaptation of such systems.

Video-surveillance systems are complex and exhibit high degrees of variability along several dimensions. At the specification level, the number of possible applications and type of scenarios is large. On the software architecture side, the number of components, their variations due to possible choices among different algorithms, the number of tunable parameters... make the processing chain configuration rather challenging. Moreover, the context of an application may change in real time, requiring dynamic reconfiguration of the chain. This huge variability raises problems at design time (finding the configurations needed by the chain, foreseeing the different possible contexts), at deployment time (selecting the initial configuration), and at run time (switching configurations to react to context changes).
The first step was to formalize in an unified way all the necessary concerns — at the specification as well as at the component level — and their relations. To this end, we rely on Feature Models and (semi) automatic model transformations. Feature Models are widely used to represent systems with many possible variation points. Moreover they are liable to formal analysis (using propositional logic and satisfiability techniques) and thus lead to valid configurations, by construction. We have developed two feature models, one for the specification of the application (type of application, context of execution, expected quality of service, etc.) and one for the implementation representation (components and their assembly). Each model has its own internal constraints. Moreover, the two models are not independent: they are connected by cross model transformation rules that formalize the bridge between application requirements and component assemblies that realize them.

Second, we propose a framework to derive valid possible system configurations and to adapt running configurations to context changes. Users can select features describing their application in the specification model, through a simple graphic interface. The outcome is a sub-model of the specification model. Based on the cross model transformation rules, our framework automatically transforms this sub-model into a sub-model of the component model. The latter represents all possible component configurations of the target video-surveillance system that satisfy the specifications. Both sub-models will be kept throughout the system life. They are used while the system is running to adjust its configuration in response to context changes.

To achieve this dynamic adaptation, our framework sets up three collaborating modules as shown in Figure 25:

- the Run Time Component Manager (RTCM) captures low level events manifesting context changes (e.g., lighting changes); it forwards them to the Configuration Adapter which returns a

---

7Here the “features” correspond to selectable concepts of the systems; they can be at any abstraction level (a feature may correspond to an specification entity such as “Intrusion detection” or to a more concrete element such as “High frame rate”). The features are organized along a tree, with logical selection relations (optional, mandatory features, exclusive choices...) and some constraints that restrict the valid combinations of features (i.e., configurations)
new component configuration; the RTCM is then responsible for applying this configuration, that is to tune, add, remove, or replace components, and possibly to change the workflow itself.

- the **Configuration Adapter** (CA) receives change events from the RTCM, translates them into the feature formalism, and forwards the result to the Model Manager; in return, it obtains a sub-model of component configurations compatible with the change; this sub-model is a compact representation of a set of valid configurations and the CA is responsible to select one and to instruct the RTCM to apply it; this selection uses some heuristics, possibly based on a cost function such as minimizing the number of component changes in the processing chain or maximizing the quality of service (e.g., accuracy, responsiveness).

- the **Model Manager** (MM) manages the representation of the two specialized Feature Models corresponding to the specification and possible component assemblies of the current application together with their constraints; its role is to enforce configuration validity. It is also responsible of the set of rules relating run time events and (de)selection of features in both models. From the CA, the Model Manager receives information about incoming events; it uses the rules to select or deselect the corresponding features; it then applies constraints, rules, and model transformations to infer a component sub-model that represents a subset of valid component configurations and that it returns to the CA.

This year, we have tested our approach on simple applications using well-known libraries (OpenCV) on different scenarios. At the moment, 77 features and $10^5$ configurations are present in the specification model while 51 features and $10^6$ configurations are present in the component model. Once the video surveillance designer has selected the features required by an application, before deployment, the average number of features to consider at runtime in the component model is less than $10^4$. The configuration spaces is reduced by several orders of magnitude and enables the use of the other tools in the end-to-end engineering process, whereas it would not have been possible without. Our experiments show the feasibility of such an approach with a limited performance overhead (if any) compared to traditional run time control where ad hoc adaptation code is hardwired and does not rely on the run time availability of an abstract representation of the application and its context evolution [29]. The next step will be to test our approach on our SUP platform and to study intelligent configuration selection heuristics.

### 6.21. The Girgit Software

**Participants:** Leonardo Rocha, Sabine Moisan, Jean-Paul Rigault, Sagar Sen.

Girgit is a Python based framework to build context-aware self-adaptive software systems.

Girgit is a simple and small [1] framework that allows dynamic reconfiguration of data processing chains and accepts any set of components for the configuration as long as they have the corresponding Python wrapper.

The basic Idea of Girgit is to provide a platform to be able to make dynamic adaptive systems. It provides a dynamic adaptive engine that can deal with event/action pairs called rules and also provides an application programming interface to be able to use it.

As the system is designed to be interactive, any user can interact with it, be a human operator or a reasoning engine. Rules can be pre-loaded with the configuration, and components can launch events, this allows pre-configured rules to be loaded at launch time of the Girgit.

Girgit has been used to evaluate the performance on real time video applications and show the architecture. Three publications [47], [45], [46] where based on the framework.

### 6.22. Scenario Analysis Module

**Participants:** Sabine Moisan, Annie Ressouche, Jean Paul Rigault.
To generate activity recognition systems we supply a scenario analysis module (SAM) to express and recognize complex events from primitive events generated by SUP or others sensors. In this framework, this year we focus on recognition algorithm improvement in order to face the problem of large number of scenario instances recognition.

The purpose of this research axis is to offer a generic tool to express and recognize activities. Genericity means that the tool should accommodate any kind of activities and be easily specialized for a particular framework. In practice, we propose a concrete language to specify activities in the form of a set of scenarios with temporal constraints between scenarios. This language allows domain experts to describe their own scenario models. To recognize instances of these models, we consider the activity descriptions as synchronous reactive systems \cite{69} and we apply general modeling methods \cite{62} to express scenario behaviors. This approach facilitates scenario validation and allows us to generate a recognizer for each scenario model. The SAM module thus provides users with (1) a simulation tool to test scenario behaviors; (2) a generator of a recognizer for each scenario model; (3) an exhaustive verification of safety properties relying on model checking techniques our approach allows. The latter offers also the possibility to define safety properties to prove as “observers” \cite{63} expressed in the scenario language.

Last year we have completed SAM in order to address the life cycle of scenario instances. For a given scenario model there may exist several (possibly many) instances at different evolution states. These instances are created and deleted dynamically, according to the input event flow. The challenge is to manage the creation/destruction of this large set of scenario instances efficiently (in time and space), to dispatch events to expecting instances, and to make them evolve independently. This year, to face this challenge we first replace some operators of the language, by others having a more strict semantics. For instance, we replace the before operator whose semantics allowed that events can meet, by two operators, a strict before and a meet. Hence, the number of events a scenario instance reacts to decreases. Second, we now generate within the recognition engine, the expected events of the next step. This avoids to run the engine automatically with events that are not relevant for the recognition process.

Presently, we still rely on the existing synchronous language (Lustre \cite{62}) to express the equational semantics of scenario models and to generate recognizers because this language offers simulation and verification means. But, to improve efficiency, we plan to build our own compiler and to generate recognizers directly from the Boolean equation systems modeling scenario models. This implies that we must supply our own simulation tool and that we interface with a model checking tool as NuSMV \cite{57}.

Now the challenge is to take into account some uncertainty on the primitive events due to input sensor errors. In the family of synchronous languages, the Lutin language \cite{8} could be able to automate the generation of realistic input sequences of events, taking into account probabilistic distributions over primitive events. In other words, it could generate a set of input events for which a set a constraints can be verified. In complement, it offers also means to compute the real values verifying these constraints. Thus, we think to rely on Lutin to express uncertainty on primitive events and get input events to feed scenario recognition engines.

### 6.23. Multiple Services for Device Adaptive Platform for Scenario Recognition

**Participants:** Annie Ressouche, Jean-Yves Tigli.

The aim of this research axis is to federate the inherent constraints of an activity recognition platform like SUP (see section 6.19) with a service oriented middleware approach dealing with dynamic evolutions of system infrastructure. The Rainbow team (Nice-Sophia Antipolis University) proposes a component-based adaptive middleware (WComp \cite{76}, \cite{75}, \cite{64}) to dynamically adapt and recompose assemblies of components. These operations must obey the "usage contract" of components. The existing approaches don’t really ensure that this usage contract is not violated during application design. Only a formal analysis of the component behavior models associated with a well sound modeling of composition operation may guarantee the respect of the usage contract.

---

\footnote{http://www-verimag.imag.fr/Lutin.html}
The approach we adopted introduces in a main assembly, a synchronous component for each sub assembly connected with a critical component. This additional component implements a behavioral model of the critical component and model checking techniques apply to verify safety properties concerning this critical component. Thus, we consider that the critical component is validated.

When a critical component has multiple synchronous monitors corresponding to several concern managements in the application, we want to build an only synchronous model component which agrees with all these primitive synchronous monitors. To specify how output events sent by different synchronous monitors and connected to a critical component, we introduce a sound (with respect to our mathematical formalism) operation of composition under constraints of synchronous models (see figure 26). We proved that this operation preserves already separately verified properties of synchronous components. This operation is an answer to the multiple access to critical components. Actually, we supply a graphical interface to design both critical component behaviors and properties as observers in the synchronous language Lustre [62]. Then the validation of properties and the creation of the validated synchronous component is automatic [44], [53].

![Diagram](image)

Figure 26. Example of multiple access to alarm critical component: (a) SM - the synchronous monitors; (b) CSM - the composition under constraints operation

This year we focus on the main challenge of this approach which is to deal with the possibly very large number of constraints a user must specify. Indeed, each synchronous monitor has to tell how it combines with others, then we get a combinatorial number of constraints with respect to the number of synchronous monitors and inputs of the critical component. To be adaptive with efficiency, we must face this problem. We first introduced some default rules to avoid the user to express a large number of constraints. We also studied how Abstract Interpretation technique can help us to reduce this complexity. This approach works if we forbid some “non monotonic” constraints, but this is a strong limitation. Thus, it is still a challenge for us. This drawback is a popular challenge in adaptive middleware and some results exist relying on controller synthesis methods. We are not in the exact framework where these techniques apply, but we plan to study if we can rely on some extension of these techniques.

On another hand, we also want to complement our preservation result in studying how the proof of a global property can be decomposed into the proof of local ones. In general, this decomposition (known as assume-
guarantee paradigm) is difficult to apply but there is no communication between synchronous monitors and so the decomposition could be tractable. Moreover, some works address this problem and we can rely on them.

### 6.24. The Clem Workflow

**Participants:** Annie Ressouche, Daniel Gaffé.

This research axis concerns the theoretical study of a synchronous language LE with modular compilation and the development of a toolkit (see figure 6) around the language to design, simulate, verify and generate code for programs. The novelty of the approach is the ability to manage both modularity and causality. Indeed, only few approaches consider a modular compilation because there is a deep incompatibility between causality and modularity. Thus, relying on semantics to compile a language ensures a modular approach but requires to complete the compilation process with a global causality checking. To tackle this problem, we introduced a new way to check causality from already checked sub programs and the modular approach we infer. The equational semantics compute a Boolean equation system and we ensure both modularity and causality in computing all the partial orders valid for a system and we define a way to merge two partial orders. The algorithm which computes partial orders rely on the computation of two dependency graphs: the upstream (downstream) dependency graph computes the dependencies of each variable of the system starting from the input (output) variables. This way of compiling is the corner stone of our approach. We defined three different approaches to compute the partial orders valid for an equation system:

1. apply **PERT** method: inputs (resp. outputs) have date 0 and recursively increase of dates for each vertex in the upstream (resp downstream) dependencies graph;

2. apply **graph theory**:
   - compute the adjacency matrix \( U \) of upstream (resp. downstream) dependencies graph;
   - the length of the maximal path from a variable \( v \) to system inputs is characterized by the maximal \( k \) such that \( U^k[v, i] \neq 0 \) for all inputs \( i \).

3. apply **fix point theory**: the vector of earliest (resp. lastest) dates can be computed as the least fix point of a monotonic increasing function.

The fix point characterization helps us to prove that the merge algorithm is correct (i.e we get the same partial orders using the merge algorithm on two previously sorted equation systems or when sorting the union of the two equation systems considered).

To be modular, we defined a technique to compose two already sorted equation systems: first, we memorize the two dependency graphs of equation systems. Second, we define two merge algorithms relying on two different techniques:

1. propagation of common variables dates adjustment;

2. fix point characterization starting with the vectors of already computed dates and considering only the variables in the dependencies (upstream and downstream) of common variables

This year we began the implementation of a separated compilation of LE programs, according to these theoretical results. We define a new intermediate format (lea) to record partially compiled module, i.e module whose Boolean equation systems may be composed of non defined variables (we called them abstract). Then we are implementing a refinement operation which replaces these abstract variables by their definition and performs adjustment of the dates. According to our theoretical results, we know that the resulting sorting is the same as with a global approach. After the termination of this separated compilation of LE programs, the challenge will be to use Clem to design a large application in the domain of smart cards. The application needs more than forty LE automata in parallel and the compiled code will have more than 500 registers and thousands variables. Only a separated compilation will work.

The Clem toolkit is completely described in [28]
5. New Results

5.1. Plausible Image Rendering

5.1.1. Filtering Solid Gabor Noise

Participants: Ares Lagae, George Drettakis.

Figure 3. Filtering Solid Gabor Noise. Solid procedural textures generated with our new random-phase solid Gabor noise: (a) Granite chess pieces; (b) Granite dancer statuettes; (c) Wooden chess board. Our new noise function enables a variety of isotropic and anisotropic solid noise textures.

Solid noise is a fundamental tool in computer graphics. Surprisingly, no existing noise function supports both high-quality anti-aliasing and continuity across sharp edges. Existing noise functions either introduce discontinuities of the solid noise at sharp edges, which is the case for wavelet noise and Gabor noise, or result in detail loss when anti-aliased, which is the case for Perlin noise and wavelet noise. In this project, we therefore present a new noise function that preserves continuity over sharp edges and supports high-quality anti-aliasing. We show that a slicing approach is required to preserve continuity across sharp edges, and we present a new noise function that supports anisotropic filtering of sliced solid noise. This is made possible by individually filtering the slices of Gabor kernels, which requires the proper treatment of phase. This in turn leads to the introduction of the phase-augmented Gabor kernel and random-phase Gabor noise, our new noise function. We demonstrate that our new noise function supports both high-quality anti-aliasing and continuity across sharp edges, as well as anisotropy. Fig. 3 shows several solid procedural textures generated with our new random-phase solid Gabor noise.

This work was presented at ACM SIGGRAPH 2011 in Vancouver and published in ACM Transactions on Graphics [18] (SIGGRAPH paper).

5.1.2. Image-Guided Weathering for Flow Phenomena

Participants: Carles Bosch, Pierre-Yves Laffont, George Drettakis.

The simulation of weathered appearance is essential in the realistic modeling of urban environments. With digital photography and Internet image collections, visual examples of weathering effects are readily available. These images, however, mix the appearance of the weathering phenomena with the specific local context. In [12], we have introduced a new methodology to estimate parameters of a phenomenological weathering simulation from existing imagery, in a form that allows new target-specific weathering effects to be produced. In addition to driving the simulation from images, we complement the visual result with details and colors extracted from the images. This methodology has been illustrated using flow stains as a representative case, demonstrating how a rich collection of flow patterns can be generated from a small set of exemplars. In Fig. 4, we show the major components required for this approach.
This work was published in ACM Transactions on Graphics [12] and also presented at ACM SIGGRAPH 2011 in Vancouver (TOG paper).

Figure 4. From left to right: given a photograph of a stain effect, we extract its stain degree map, we estimate the information required for the simulation (color, simulation parameters, and high frequency details), and run simulations on new scenes using this information, resulting in a wide variety of similar-looking stains.

5.1.3. Relighting Photographs of Tree Canopies

Participants: Marcio Cabral, George Drettakis.

We present an image-based approach to relighting photographs of tree canopies. Our goal is to minimize capture overhead; thus the only input required is a set of photographs of the tree taken at a time of day, while allowing relighting at any other time. We first analyze lighting in a tree canopy, both theoretically and using simulations. From this analysis, we observe that tree canopy lighting is similar to volumetric illumination. We assume a single-scattering volumetric lighting model for tree canopies, and diffuse leaf reflectance. To validate our assumptions, we apply our method on several synthetic renderings of tree models, for which we are all able to compute all quantities involved.

Our method first creates a volumetric representation of the tree using 10-12 images taken at a single time of day and use a single-scattering participating media lighting model - these photos are taken from different viewpoints, around the tree. An analytical sun and sky illumination model, namely the Preetham model, provides consistent representation of lighting for the captured input and unknown target times.

We relight the input image by applying a ratio of the target and input time lighting representations. We compute this representation efficiently by simultaneously coding transmittance from the sky and to the eye in spherical harmonics. We validate our method by relighting images of synthetic trees and comparing to path-traced solutions. We also present results for photographs, validating with time-lapse ground truth sequences. An example is shown in Fig. 5. This work was published in the IEEE Transactions on Computer Graphics and Visualization [15], and was in collaboration with the past members of REVES, N. Bonneel and S. Lefebvre.

5.1.4. Silhouette-aware Warping for Image-based Rendering

Participants: Gaurav Chaurasia, George Drettakis.
Figure 5.

Mulberry tree Top row: input images and 4 target images with corresponding times of day. Middle row: 4 resulting relit images using our approach. Bottom row: $E_{in}$ and the four $E_{target}$ images.
Figure 6. Top: Overview of our approach starting from input images to rendered novel views along with novel view generated using current state of the art IBR. Bottom: Novel views rendered using our approach.
We have presented a novel solution for image-based rendering (IBR) of urban scenes based on image warping. IBR techniques allow capture and display of 3D environments using photographs. Modern IBR pipelines reconstruct proxy geometry using multi-view stereo, reproject the photographs onto the proxy and blend them to create novel views. The success of these methods depends on accurate 3D proxies, which are difficult to obtain for complex objects such as trees and cars. Large number of input images do not improve reconstruction proportionally; surface extraction is challenging even from dense range scans for scenes containing such objects. Our approach does not depend on dense accurate geometric reconstruction; instead we compensate for sparse 3D information by variational warping. In particular, we formulate silhouette-aware warps that preserve salient depth discontinuities. This improves the rendering of difficult foreground objects, even when deviating from view interpolation. We use a semi-automatic step to identify depth discontinuities and extract a sparse set of depth constraints used to guide the warp. On the technical side, our formulation breaks new ground by demonstrating how to incorporate discontinuities in variational warps. Our framework is lightweight and results in good quality IBR for previously challenging environments as shown in figure.

Applications. Robust image-based rendering can be used to generate photo-realistic visual content easily which can be very useful for virtual reality applications. Commercial products like Google StreetView and Microsoft PhotoSynth use rudimentary image-based rendering for large scale visualization of cities. Our work advances the state of the art by treating the hardest class of scenes.

This work is a collaboration with Prof. Dr. Olga Sorkine (ETH Zurich). The work was published in the special issue of the journal Computer Graphics Forum [16], presented at the Eurographics Symposium on Rendering 2011 in Prague, Czech Republic.

5.1.5. Real Time Rough Refractions

Participant: Adrien Bousseau.

Figure 7. Compared to an expensive ray-traced reference (a), our method produces plausible results in real-time (b).

We have proposed an algorithm to render objects made of transparent materials with rough surfaces in real-time, under environment lighting. Rough surfaces such as frosted and misted glass cause wide scattering as light enters and exits objects, which significantly complicates the rendering of such materials. We introduced two contributions to approximate the successive scattering events at interfaces, due to rough refraction: First, an approximation of the Bidirectional Transmittance Distribution Function (BTDF), using spherical Gaussians, suitable for real-time estimation of environment lighting using pre-convolution; second, a combination of cone tracing and macro-geometry filtering to efficiently integrate the scattered rays at the exiting interface of the object. Our method produces plausible results in real-time, as demonstrated by comparison against stochastic ray-tracing (Figure 7).
This work is a collaboration with Charles De Rousiers, Kartic Subr, Nicolas Holzschuch (ARTIS / INRIA Rhône-Alpes) and Ravi Ramamoorthi (UC Berkeley) as part of our CRISP associate team with UC Berkeley. The work was presented at the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games 2011 and won the best paper award.

5.1.6. Single view intrinsic images

**Participants:** Pierre-Yves Laffont, Adrien Bousseau, George Drettakis.

![Figure 8. Starting from multiple views of the scene (a), our method automatically decomposes photographs into three intrinsic components — the reflectance (b), the illumination due to sun (c) and the illumination due to sky (d). Each intrinsic component can then be manipulated independently for advanced image editing applications.](image)
This information is propagated to the remaining pixels. In a final step we adapt appropriate image propagation algorithms, by replacing manual user indications by data inferred from the geometry-driven shadow and reflectance estimation. This allows us to automatically extract intrinsic images from multiple viewpoints, thus allowing many types of manipulation in images.

As illustrated on Figure 8, our method can extract reflectance at each pixel of the input photographs. The decomposition also yields separate sun illumination and sky illumination components, enabling easier manipulation of shadows and illuminant colors in image editing software.

This work won the best paper award at “les journées de l’AFIG (Association Francophone d’Informatique Graphique) 2011”, and will appear in the French journal REFIG [21].

5.1.7. Single view intrinsic images with indirect illumination

Participants: Pierre-Yves Laffont, Adrien Bousseau, George Drettakis.

Building on top of our work published in [21], we extended our intrinsic image decomposition method to handle indirect illumination, in addition to sun and sky illumination. We introduced a new algorithm to reliably identify points in shadow based on a parameterization of the reflectance with respect to sun visibility, and proposed to separate illumination components using cascaded image-based propagation algorithm. We also demonstrated direct applications of our results for the end user in widespread image editing software.

This rich intrinsic image decomposition method has been submitted for publication.

5.1.8. Multi-lighting multi-view intrinsic images

Participants: Pierre-Yves Laffont, Adrien Bousseau, George Drettakis.

We compute intrinsic image decompositions using several images of the same scene under different viewpoints and lighting conditions. Such images can be easily gathered for famous landmarks, using photo sharing websites such as Flickr. Our method leverages the heterogeneity of photo collections (multiple viewpoints and lighting conditions) to guide the intrinsic image separation process. With this automatic decomposition, we aim to facilitate many image editing tasks and improve the quality of image-based rendering from photo collections.

This work is in collaboration with Frédo Durand (associate professor at MIT) and Sylvain Paris (researcher at Adobe).

5.1.9. Warping superpixels

Participants: Gaurav Chaurasia, Sylvain Duchêne, George Drettakis.

We are working on developing a novel representation of multi-view scenes in the form of superpixels. Such a representation segments the scene into semantically meaningful regions called superpixels which admit variational warps [16], thereby leveraging the power of shape-preserving warps while providing automatic silhouette and occlusion handling. This entails several contributions namely incorporating geometric priors in superpixel extraction, generating consistent warping constraints and a novel rendering pipeline that assembles warped superpixels into the novel view maintaining spatio-temporal coherence.

This work is a collaboration with Prof. Dr. Olga Sorkine (ETH Zurich).

5.1.10. State-of-the-art Report on Temporal Coherence for Stylized Animations

Participant: Adrien Bousseau.
Non-photorealistic rendering (NPR) algorithms allow the creation of images in a variety of styles, ranging from line drawing and pen-and-ink to oil painting and watercolor. These algorithms provide greater flexibility, control and automation over traditional drawing and painting. The main challenge of computer generated stylized animations is to reproduce the look of traditional drawings and paintings while minimizing distracting flickering and sliding artifacts present in hand-drawn animations. These goals are inherently conflicting and any attempt to address the temporal coherence of stylized animations is a trade-off. This survey is motivated by the growing number of methods proposed in recent years and the need for a comprehensive analysis of the trade-offs they propose. We formalized the problem of temporal coherence in terms of goals and compared existing methods accordingly (Figure 9). The goal of this report is to help uninformed readers to choose the method that best suits their needs, as well as motivate further research to address the limitations of existing methods.

Figure 9. The temporal coherence problem involves three goals represented by the axes of these diagrams. Fully ignoring one of them produces the opposite artifacts.

This work is a collaboration with Pierre Bénard and Joëlle Thollot (ARTIS / INRIA Rhône-Alpes) and was published in the journal Computer Graphics Forum [14].

5.1.11. Improving Gabor Noise

Participant: Ares Lagae.

We have recently proposed a new procedural noise function, Gabor noise, which offers a combination of properties not found in existing noise functions. In this project, we present three significant improvements to Gabor noise: (1) an isotropic kernel for Gabor noise, which speeds up isotropic Gabor noise with a factor of roughly two, (2) an error analysis of Gabor noise, which relates the kernel truncation radius to the relative error of the noise, and (3) spatially varying Gabor noise, which enables spatial variation of all noise parameters. These improvements make Gabor noise an even more attractive alternative for existing noise functions. Fig. 3 shows a procedural textures generated with our new improved Gabor noise. This work was published in IEEE Transactions on Computer Graphics and Visualization [19].

5.2. Visual Perception and Audio Rendering
5.2.1. Perception of Visual Artifacts in Image-Based Rendering of Façades

Participants: Peter Vangorp, Gaurav Chaurasia, Pierre-Yves Laffont, George Drettakis.

Image-based rendering (IBR) techniques allow users to create interactive 3D visualizations of scenes by taking a few snapshots (Figure 11 (left)). However, despite substantial progress in the field, the main barrier to better quality and more efficient IBR visualizations are several types of common, visually objectionable artifacts. These occur when scene geometry is approximate or viewpoints differ from the original shots, leading to parallax distortions (Figure 11 (mid)), blurring, ghosting (Figure 11 (right)) and popping errors that detract from the appearance of the scene. We argue that a better understanding of the causes and perceptual impact of these artifacts is the key to improving IBR methods.

We present a series of psychophysical experiments in which we systematically map out the perception of artifacts in IBR visualizations of façades as a function of the most common causes. We separate artifacts into different classes and measure how they impact visual appearance as a function of the number of images available, the geometry of the scene and the viewpoint. The results reveal a number of counter-intuitive effects in the perception of artifacts.

We summarize our results in terms of the following practical guidelines for improving existing and future IBR techniques:
• When the total number of available images is small, e.g., because of storage limitations, it is preferable to use a sudden transition with its associated popping artifact rather than a gradual blending transition with its associated ghosting artifact.
• Interestingly, the depth range of the façade does not affect the perceived parallax distortions, even though it clearly does affect the objective parallax distortions. Only the intended output viewing angle should be taken into account when capturing images.
• For Google Street View™-like visualizations, a shorter cross-fading transition would improve the perceived quality.

This work is a collaboration with Roland W. Fleming (Justus-Liebig-University Giessen, Germany). The work was published in the special issue of the journal Computer Graphics Forum [20] and presented at the Eurographics Symposium on Rendering 2011.

5.2.2. Perception of Slanted, Textured Façades

Participants: Peter Vangorp, Adrien Bousseau, Gaurav Chaurasia, George Drettakis.

In large-scale urban visualizations, buildings are often geometrically represented by simple boxes textured with images of the façades. Any depth variations in the façade, such as balconies, are perceived to have distorted angles when the viewer is not at the capture camera position. The retinal hypothesis provides the most likely prediction of the magnitude of the perceived distortion. We conduct psychophysical experiments to measure the perceived distortion, thereby validating the retinal hypothesis, and to measure the threshold for detecting any distortion. The result is a prediction of the valid range of viewer motion for a given capture.

This work is a collaboration with Martin S. Banks (UC Berkeley).

5.2.3. Binocular and Dynamic Cues to Glossiness

Participants: Peter Vangorp, George Drettakis.

Recent advances in display technology have made it possible to present high quality stereoscopic imagery with accurate head tracking. This improves not only depth perception but also affects the perception of glossy materials. Previous work has shown that these conditions can increase the perceived gloss by a small amount. We conduct psychophysical experiments to measure this effect quantitatively.

This work is a collaboration with Roland W. Fleming (Justus-Liebig-University Giessen, Germany) and Martin S. Banks (UC Berkeley).

5.2.4. Sound Particles

Participants: Charles Verron, George Drettakis.

This research deals with a sound synthesizer dedicated to particle-based environmental effects, and intended to be used in interactive virtual environments. The synthesis engine is based on five physically-inspired basic elements (sound atoms) that can be parameterized and stochastically distributed in time and space. Physically-inspired controls simultaneously drive graphics particle models (e.g., distribution of particles, average particles velocity etc.) and sound parameters (e.g., distribution of sound atoms, spectral modifications etc.). The simultaneous audio/graphics controls result in an intricate interaction between the two modalities that enhances the naturalness of the scene. The approach is currently illustrated with three environmental phenomena: fire, wind, and rain.

5.2.5. Sound Synthesis for Crowds

Participants: Charles Verron, George Drettakis.

We are currently investigating new methods for synthesis of crowd sounds in virtual environments. Crowd sounds are constituted of many overlapping voices spatialized at different positions in the environment. A novel level of detail for crowd sounds is desirable: the cost of spatializing many individual voice sounds can be replaced by an efficient babble noise synthesis model. Furthermore, high-level control should be proposed to modify the crowd sounds by semantic parameters, related to the crowd emotional state (e.g., calm, angry...). This research should result in a new real-time crowd sound synthesizer with semantic controls for virtual environments.
5.3. Interaction and Design for Audiovisual Virtual Environments

5.3.1. Lighting Design for Material Depiction

Participants: Adrien Bousseau, Emmanuelle Chapoulie.

Shading, reflections and refractions are important visual features for understanding the shapes and materials in an image. While well designed lighting configurations can enhance these features and facilitate image perception (Figure 12 b), poor lighting design can lead to misinterpretation of image content (Figure 12 a).

![Figure 12. Our method (a) automatically optimizes the lighting to enhance material-specific visual features. The lighting reveals the thin and thick parts of the subsurface scattering wax candle, it accentuates the Fresnel reflections along the side of the porcelain vase and it adds strong specular highlights to emphasize the shiny chrome metal of the sculpture. Poorly designed lighting (b) diminishes these characteristic visual features of the materials. The candle appears more like solid plastic, the vase appears to be made of diffuse clay and the sculpture no longer looks like it is made of chrome.](image)

We have presented an automated system for optimizing and synthesizing environment maps that enhance the appearance of materials in a scene. We first identified a set of lighting design principles for material depiction. Each principle specifies the distinctive visual features of a material and describes how environment maps can emphasize those features. We then proposed a general optimization framework to solve for the environment map that best fulfill the design principles. Finally we described two techniques for transforming existing photographic environment maps to better emphasize materials. Our approach generates environment maps that enhance the depiction of a variety of materials including glass, metal, plastic, marble and velvet.

This work is a collaboration with Ravi Ramamoorthi and Maneesh Agrawala (UC Berkeley) as part of our CRISP associate team with UC Berkeley. The work was published in the special issue of the journal Computer Graphics Forum, presented at the Eurographics Symposium on Rendering 2011.

5.3.2. A Multimode Immersive Conceptual Design System for Architectural Modeling and Lighting

Participants: Marcio Cabral, Peter Vangorp, Gaurav Chaurasia, Emmanuelle Chapoulie, Martin Hachet, George Drettakis.

We developed a system which allows simple architectural design in immersive environments. The user is able to define the initial conceptual design of the model and can take into account the effects of daylight. Our system allows the manipulation of simple elements such as windows, doors and rooms while the overall model is automatically adjusted to the manipulation. The system runs on a four-sided stereoscopic, head-tracked immersive display. We also provide simple lighting design capabilities, with an abstract representation of sunlight and its effects when shining through a window. Our system provides three different modes
of interaction: a miniature-model \textit{table mode}, a fullscale \textit{immersive mode} and a combination of table and immersive which we call \textit{mixed mode} (see Figure 13). Our goal is to study direct manipulation for basic 3D modeling in an immersive setting, in the context of conceptual or initial design for architecture.

We performed an initial pilot user test to evaluate the relative merits of each mode for a set of basic tasks such as resizing and moving windows or walls, and a basic light-matching task. The study indicates that users appreciated the immersive nature of the system, and found interaction to be natural and pleasant. In addition, the results indicate that the mean performance times seem quite similar in the different modes, opening up the possibility for their combined usage for effective immersive modeling systems for novice users.

![Figure 13. (a) Table mode, (b) Immersive mode, (c) Mixed mode](image)

### 5.3.3. Walking in a Cube: Novel Metaphors for Safely Navigating Large Virtual Environments in Restricted Real Workspaces

**Participants:** Peter Vangorp, Emmanuelle Chapoulie, George Drettakis.

Immersive spaces such as 4-sided displays with stereo viewing and high-quality tracking provide a very engaging and realistic virtual experience. However, walking is inherently limited by the restricted physical space, both due to the screens (limited translation) and the missing back screen (limited rotation). We propose three novel navigation techniques that have three concurrent goals: keep the user safe from reaching the translational and rotational boundaries; increase the amount of real walking; and finally, provide a more enjoyable and ecological interaction paradigm compared to traditional controller-based approaches. We notably introduce the “Virtual Companion”, which uses a small bird to guide the user through VEs larger than the physical space. We evaluate the three new techniques through a user study with pointing and path following tasks. The study provides insight into the relative strengths of each new technique for the three aforementioned goals.

This work is a collaboration with Gabriel Cirio, Maud Marchal, and Anatole Lécuyer (VR4I / INRIA Rennes) in the context of ARC NIEVE (Section 6.2.1) and has been accepted for publication [17].

### 5.3.4. Inferring Normals Over Design Sketches

**Participant:** Adrien Bousseau.

We are currently working on a sketch-based tool to infer normals over a 2D drawing. Our tool should allow users to apply realistic and non-photorealistic shading over the drawing, with applications in product design. This work is a collaboration with Alla Sheffer (University of British Columbia), Cloud Shao and Karan Singh (University of Toronto).
5.3.5. Using natural gestures into virtual reality immersive space

Participants: Emmanuelle Chapoulie, George Drettakis.

We are studying the use of gestures which are as natural as possible in a context of virtual reality environments. We define a scenario which is a sequence of tasks (hiding, finding, pushing, pulling, grabbing, picking up, putting down objects) that the users will perform with hands and with wands, in order to evaluate the usability of our interaction approach. Each task will be used to evaluate a specific criterion.
6. New Results

6.1. Data and Metadata Management

6.1.1. Uncertain Data Management

Participants: Reza Akbarinia, Patrick Valduriez, Guillaume Verger.

Data uncertainty in scientific applications can be due to many different reasons: incomplete knowledge of the underlying system, inexact model parameters, inaccurate representation of initial boundary conditions, inaccuracy in equipments, etc. For instance, in the monitoring of plant contamination, sensors generate periodically data which may be uncertain. Instead of ignoring (or correcting) uncertainty, which may generate major errors, we need to manage it rigorously and provide support for querying.

In [46], we address the problem of aggregate queries that return possible sum values and their probabilities. This kind of query which, we call ALL-SUM, is also known as sum probability distribution. The results of ALL-SUM can be used for answering many other type of queries over probabilistic data. In general, the problem of ALL-SUM query execution is NP-complete. We propose pseudo-polynomial algorithms that are efficient in many practical applications, e.g. when the aggr attribute values are small integers or real numbers with small precision, i.e. small number of digits after decimal point. These cases cover many practical attributes, e.g. temperature, blood pressure, needed human recourses per patient in medical applications.

We have started to develop a probabilistic database prototype, called ProbDB (Probabilistic Database), on top of an RDBMS. ProbDB divides a query into two parts: probabilistic and deterministic (i.e. non probabilistic). The deterministic part is executed by the underlying RDBMS, and the rest of work is done by our probabilistic query processing algorithms that are executed over the data returned by the RDBMS. In [51], we demonstrated the efficient execution of aggregate queries with the first version of ProbDB.

6.1.2. Metadata Integration

Participants: Zohra Bellahsène, Rémi Coletta, Duy Hoa Ngo.

Due to the various types of heterogeneity of ontologies, ontology matching must exploit many features of ontology elements in order to improve matching quality. For this purpose, numerous similarity metrics have been proposed to deal with ontology semantics at different levels: elements level, structural level and instance level.

Elements level metrics can be categorized in three groups: (1) terminological, (2) structural and (3) semantic. Metrics of the first group exploit text features such as names, labels and comments to compute the similarity score between entities. Whereas metrics of the last two groups exploit the hierarchy and semantic relationship features. Our approach consists in first using terminological metrics. Then, during the matching process, mappings discovered by terminological metrics are used as input mappings to other metrics of the second and third groups. Obviously, the more precise results terminological metrics are, the more accurate results structural and semantic metrics have.

However, finding a good combination of different metrics is very difficult and time consuming. We proposed YAM++ (not Yet Another Matcher), an approach that uses machine learning to combine similarity metrics. Our main contributions are: the definition of new metrics dealing with terminological and context profile features of entities in ontologies [37], and the use of a decision tree model to combine similarity metrics [38].

To improve matching quality of YAM++, we exploit instances accompanying ontologies. We then apply similarity flooding propagation algorithm to discover more semantic mappings. At the 2011 competition of the Ontology Alignment Evaluation Initiative (http://oaei.ontologymatching.org), YAM++ achieved excellent results: first position on the Conference track and second position on the Benchmark track [39].
6.2. Data and Process Sharing

6.2.1. Social-based P2P Data Sharing

Participants: Hinde Bouziane, Michèle Cart, Esther Pacitti, Didier Parigot, Guillaume Verger.

This work focuses on P2P content recommendation for on-line communities. In [20], we propose P2Prec, a recommendation service for P2P content sharing systems that exploits users’ social data. Given a query, P2Prec finds peers that can recommend high quality documents that are relevant for the query. A document is relevant to a query if it covers the same topics. It is of high quality if relevant peers have rated it highly. P2Prec finds relevant peers through a variety of mechanisms including advanced content-based and collaborative filtering. The topics each peer is interested in are automatically calculated by analyzing the documents the peer holds. Peers become relevant for a topic if they hold a certain number of highly rated documents on this topic.

To efficiently disseminate information about peers’ topics and relevant peers, we proposed new semantic-based gossip protocols. In our experimental evaluation, using the TREC09 dataset, we showed that using semantic gossip increases recall by a factor of 1.6 compared to well-known random gossiping. Furthermore, P2Prec has the ability to get reasonable recall with acceptable query processing load and network traffic. P2Prec was demonstrated in [31] and [47].

In [30], we exploit social relationships between users as a parameter to increase the trust of recommendation. We propose a novel P2P recommendation approach (called F2Frec) that leverages content and social-based recommendation by maintaining a P2P and friend-to-friend network. This network is used as a basis to provide useful and high quality recommendations. Based on F2Frec, we propose new metrics, such as usefulness and similarity (among users and their respective friend network). We define our proposed metrics based on users’ topic of interest and relevant topics that are automatically extracted from the contents stored by each user. Our experimental evaluation, using the TREC09 dataset and Wiki vote social network, shows the benefits of our approach compared to anonymous recommendation. In addition, we show that F2Frec increases recall by a factor of 8.8 compared with centralized collaborative filtering.

6.2.2. Satisfaction-based Query Replication

Participant: Patrick Valduriez.

In a large-scale Internet-based distributed, participants (consumers and providers) who are willing to share data are typically autonomous, i.e. they may have special interests towards queries and other participants’ data. In this context, a way to avoid a participant to voluntarily leave the system is satisfying its interests when allocating queries. However, participants’ satisfaction may also be negatively affected by the failures of other participants. Query replication can deal with providers failures, but, it is challenging because of autonomy: it cannot only quickly overload the system, but also dissatisfaction participants with uninteresting queries. Thus, a natural question arises: should queries be replicated? If so, which ones? and how many times?

In [25], we answer these questions by revisiting query replication from a satisfaction and probabilistic point of view. We propose a new algorithm, called S b QR, that decides on-the-fly whether a query should be replicated and at which rate. As replicating a large number of queries might overload the system, we propose a variant of our algorithm, called S b QR+. The idea is to voluntarily fail to allocate as many replicas as required by consumers for low critical queries so as to keep resources for high critical queries during query-intensive periods. Our experimental results demonstrate that our algorithms significantly outperform the baseline algorithms from both the performance and satisfaction points of view. We also show that our algorithms automatically adapt to the criticality of queries and different rates of participant failures.

6.2.3. View Selection in Scientific Data Warehousing

Participants: Zohra Bellahsène, Rémi Coletta, Imen Mami.
Scientific data generate large amounts of data which have to be collected and stored for analytical purpose. One way to help managing and analyzing large amounts of data is data warehousing, whereby views over data are materialized. However, view selection is an NP-hard problem because of many parameters: query cost, view maintenance cost and storage space. In [36], we propose a new solution based on constraint programming, which has proven efficient at solving combinatorial problems. This allows using a constraint programming solver to set up the search space by identifying a set of views that minimizes the total query cost. We address view selection under two cases: (1) only the total view maintenance cost needs be minimized, assuming unlimited storage space (meaning that it is not a critical resource anymore); (2) both storage space and maintenance cost must be minimized. We implemented our approach and compared it with a randomized method (i.e., genetic algorithm). We experimentally show that our approach provides better performance resulting from evaluating the quality of the solutions in terms of cost savings. Furthermore, our approach scales well with the query workload.

6.2.4. Scientific Workflow Management

Participants: Ayoub Ait Lahcen, Eduardo Ogasawara, Didier Parigot, Patrick Valduriez.

Scientific workflows have emerged as a basic abstraction for structuring and executing scientific experiments in computational environments. In many situations, these workflows are computationally and data-intensive, thus requiring execution in large-scale parallel computers. However, parallelization of scientific workflows remains low-level, ad-hoc and labor-intensive, which makes it hard to exploit optimization opportunities.

To address this problem, we propose in [23] an algebraic approach (inspired by relational algebra) and a parallel execution model that enable automatic optimization of scientific workflows. With our scientific workflow algebra, data is uniformly represented by relations and workflow activities are mapped to operators that have data aware semantics. Our workflow execution model is based on the concept of activity activation, which enables transparent distribution and parallelization of activities;

We conducted a thorough validation of our approach using both a real oil exploitation application and synthetic data scenarios. The experiments were run in Chiron, a data-centric scientific workflow engine implemented at UFRJ to support our algebraic approach. Our experiments demonstrate performance improvements of up to 226% compared to an ad-hoc workflow implementation. This work was done in the context of the Equipe Associée Sarava and the CNPq-INRIA project DatLuge.

In the context of SON, we also proposed a declarative workflow language based on service/activity rules [41]. This language makes it possible to infer a dependency graph for SON applications that provides for automatic parallelization.

6.3. Scalable Data Analysis

6.3.1. Massive Graph Management

Participant: Patrick Valduriez.

Traversing massive graphs as efficiently as possible is essential for many scientific applications. Many common operations on graphs, such as calculating the distance between two nodes, are based on the Breadth First Search (BFS) traversal. However, because of the exhaustive exploration of all the nodes and edges of the graph, this operation might be very time consuming. A possible solution is partitioning the graph among the nodes of a shared-nothing parallel system. However, partitioning a graph and keeping the information regarding the location of vertices might be unrealistic for massive graphs because of much inter-node communication. In [28], we propose ParallelGDB, a new graph database system based on specializing the local caches of any node in this system, providing a better cache hit ratio. ParallelGDB uses a random graph partitioning, avoiding complex partition methods based on the graph topology, that usually require managing extra data structures. This proposed system provides an efficient environment for distributed graph databases.

6.3.2. Top-k Query Processing in Unstructured P2P Systems

Participants: Reza Akbarinia, William Kokou Dedzoe, Patrick Valduriez.
Top-k query processing techniques are useful in unstructured P2P systems to avoid overwhelming users with too many results and provide them with the best ones. However, existing approaches suffer from long waiting times, because top-k results are returned only when all queried peers have finished processing the query. As a result, response time is dominated by the slowest queried peer. We proposed to revisit the problem of top-k query processing. In [29] we address the problem of reducing user waiting time of top-k query processing, in the case of unstructured P2P systems with overloaded peers. We propose a new algorithm, called QUAT, in which each peer maintains a semantic description of its local data and the semantic descriptions of its neighborhood (i.e. the semantic descriptions of data owned locally by its direct neighbors and data owned locally by these neighbors direct neighbors). These semantic descriptions allow peers to prioritize the queries that can provide high quality results, and to forward them in priority to the neighbors that can provide high quality answers. We validated our solution through a thorough experimental evaluation using a real-world dataset. The results show that QUAT significantly outperforms baseline algorithms by returning faster the final top-k results to users.

6.3.3. **Top-k Query Processing Over Sorted Lists**

**Participants:** Reza Akbarinia, Esther Pacitti, Patrick Valduriez.

The problem of answering top-k queries can be modeled as follows. Suppose we have $m$ lists of $n$ data items such that each data item has a local score in each list and the lists are sorted according to the local scores of their data items. Each data item has an overall score computed based on its local scores in all lists using a given scoring function. Then, the problem is to find the $k$ data items whose overall scores are the highest. This problem model is a general model for top-k queries in many centralized, distributed and P2P applications. For example, in IR systems one of the main problems is to find the top-k documents whose aggregate rank is the highest wrt. some given keywords. To answer this query, the solution is to have for each keyword a ranked list of documents, and return the $k$ documents whose aggregate rank in all lists are the highest.

In [16], we propose an extension of our best position algorithms (BPA) which had been proposed for top-k query processing over sorted lists model. The BPA algorithms have been shown to be more efficient than the well known TA Algorithm. We propose several techniques using different data structures for managing best positions that are crucial for efficient execution of top-k algorithms. We also provide a complete discussion on the instance optimality of TA algorithm (TA was considered so far as optimal over any database of sorted lists). We illustrate that, the existence of deterministic algorithms such as BPA shows that if we are aware of positions of seen data, then one of the main arguments used for proving the instance optimality of TA is invalidated. Therefore, in this case the proof of TA’s instance optimality is incorrect, and must be revisited.

6.3.4. **Satellite Image Mining**

**Participant:** Florent Masseglia.

Satellite Image Time Series (SITS) provide us with precious information on land cover evolution. By studying SITS, we can both understand the changes of specific areas and discover global phenomena that spread over larger areas. Changes that can occur throughout the sensing time can spread over very long periods and may have different start time and end time depending on the location, which complicates the mining and the analysis of series of images. In [45], we propose a frequent sequential pattern mining method for SITS analysis. Designing such a method called for important improvements on the data mining principles. First, the search space in SITS is multi-dimensional (the radiometric levels of different wavelengths correspond to infra-red, red, etc.). Furthermore, the non evolving regions, which are the vast majority and overwhelm the evolving ones, challenge the discovery of these patterns. Our framework enables discovery of these patterns despite these constraints and characteristics. We introduce new filters in the mining process to yield important reductions in the search space by avoiding consecutive occurrences of similar values in the sequences. Then, we propose visualization techniques for results analysis (where modified regions are highlighted). Experiments carried out on a particular dataset showed that our method allows extracting repeated, shifted and distorted temporal behaviors. The flexibility of this method makes it possible to capture complex behaviors from multi-source, noisy and irregularly sensed data.
6.3.5. Distributed Approximate Similarity Join
Participant: Alexis Joly.

Efficiently constructing the KNN-graph of large and high dimensional feature datasets is crucial for many data intensive applications involving feature-rich objects, such as image features, text features or sensor’s features. In this work we investigate the use of high dimensional hashing methods for efficiently approximating the full knn graph of large collections, in particular, in distributed environments. We first analyzed and experimented what seems to be the most intuitive hashing-based approach: constructing several Locality Sensitive Hashing (LSH) tables in parallel and computing the frequency of all emitted collisions. We show that balancing issues of classical LSH functions strongly affect the performance of this approach. On the other side, we show that using an alternative data-dependent hashing function (RMMH), that we introduced recently [34], can definitely change that conclusion. The main originality of RMMH hash function family is that it is based on randomly trained classifiers, allowing to learn random and balanced splits of the data instead of using random splits of the feature space as in LSH. We show that the hash tables constructed through RMMH are much more balanced and that the number of emitted collisions can be strongly reduced without degrading quality. In the end, our hashing-based filtering algorithm of the all-pairs graph is two orders of magnitude faster than the one based on LSH. An efficient distributed implementation of the method was implemented within the MapReduce framework (and is the basis of the SimJoin prototype). This work is done in the context of the supervision of a PhD student working at INRIA Imedia (Riadh Mohamed Trad).

6.3.6. Visual objects mining
Participant: Alexis Joly.

State-of-the-art content-based object retrieval systems have demonstrated impressive performance in very large image datasets. These methods, based on fine local descriptions and efficient matching techniques, can detect accurately very small rigid objects with unambiguous semantics such as logos, buildings, manufactured objects, posters, etc. Mining such small objects in large collections is however difficult. Constructing a full local matching graph with a naïve approach would indeed require to probe all candidate query leading to an intractable algorithm complexity. In this work, we first introduce an adaptive weighted sampling scheme, starting with some prior distribution and iteratively converging to unvisited regions [35]. We show that the proposed method allows to discover highly interpretable visual words while providing excellent recall and image representativity. We then focused on mining visual objects on top of the discovered visual words. We therefore developed an original shared nearest-neighbors clustering method, working directly on the generated bi-partite graph. This work is in the context of the supervision of two PhD students, one working jointly with INA and INRIA and who will join the Zenith team next year (Pierre Letessier), one working at INRIA Imedia (Amel Hamzaoui).

6.3.7. Visual-based plant species identification from crowdsourced data
Participant: Alexis Joly.

Inspired by citizen sciences, the main goal of this work is to speed up the collection and integration of raw botanical observation data, while providing to potential users an easy and efficient access to this botanical knowledge. We therefore designed and developed an original crowdsourcing web application dedicated to the access of botanical knowledge through automated identification of plant species by visual content. Technically, the first side of the application deals with content-based identification of plant leaves. Whereas state-of-the-art methods addressing this objective are mostly based on leaf segmentation and boundary shape features, we developed a new approach based on local features and large-scale matching. This approach obtained the best results within ImageCLEF 2011 plant identification benchmark [48]. The second side of the application deals with interactive tagging and allows any user to validate or correct the automatic determinations returned by the system. Overall, this collaborative system allows to enrich automatically and continuously the visual botanical knowledge and thus to increase progressively the accuracy of the automated identification. A demo of the developed application was presented at the ACM Multimedia conference [33]. This work is done in collaboration with INRIA Imedia and with the botanists of the AMAP UMR team (CIRAD). It is also closely related to a citizen science project around plant’s identification that we develop with the support of the TelaBotanica social network.