Activity Report 2012

Section New Results

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

5. New Results

5.1. Modeling Interfaces and Contacts

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

5.1.1. *Modeling Macro-molecular Complexes : a Journey Across Scales*
Participants: Frédéric Cazals, Tom Dreyfus.

In collaboration with C. Robert (IBPC / CNRS, Paris, France).

While proteins and nucleic acids are the fundamental components of an organism, Biology itself is based on the interactions they make with each other. Analyzing macromolecular interactions typically requires handling systems involving from two to hundreds of polypeptide chains. After a brief overview of the modeling challenges faced in computational structural biology, the text [16] reviews concepts and tools aiming at improving our understanding of the link between the static structures of macromolecular complexes and their biophysical/biological properties. We discuss geometrical approaches suited to atomic-resolution complexes and to large protein assemblies; for each, we also present examples of their successful application in quantifying and interpreting biological data. This methodology includes state-of-the-art geometric analyses of surface area, volume, curvature, and topological properties (isolated components, cavities, voids, cycles) related to Voronoi constructions in the context of structure analysis. On the applied side, we present novel insights into real biological problems gained thanks to these modeling tools.

5.1.2. *CSA: Comprehensive Comparison of Pairwise Protein Structure Alignments*
Participant: Noël Malod-Dognin.

In collaboration with I. Wohlers (CWI / VU University Amsterdam, Netherlands), R. Andonov (Irisa / Rennes University, France), G.W. Klau (CWI / VU University Amsterdam, Netherlands).

Protein structural alignment is a key method for answering many biological questions involving the transfer of information from well-studied proteins to less well-known proteins. Since structures are more conserved during evolution than sequences, structural alignment allows for the most precise mapping of equivalent residues. Many structure-based scoring schemes have been proposed and there is no consensus on which scoring is the best. Comparative studies also show that alignments produced by different methods can differ considerably. Based on the alignment engine derived from A_purva, we designed CSA (Comparative Structural Alignment), the first web server for computation, evaluation and comprehensive comparison of pairwise protein structure alignments at single residue level [15]. It offers the exact computation of alignments using the scoring schemes of DALI, Contact Map Overlap (CMO), MATRAS and PAUL. In CSA, computed or uploaded alignments can be explored in terms of many inter-residue distances, RMSD, and sequence-based scores. Intuitive visualizations also help in grasping the agreements and differences between alignments. The user can thus make educated decisions about the structural similarity of two proteins and, if necessary, post-process alignments by hand. CSA is available at http://csa.project.cwi.nl.

Upon publication [15], CSA was selected by *Nucleic Acids Research* as featured article of July 2012 (top 5% of papers in terms of originality, significance and scientific excellence.

5.2. Modeling Macro-molecular Assemblies

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

5.2.1. *Stoichiometry Determination for Mass-spectrometry Data: the Interval Case*

Participants: Deepesh Agarwal, Frédéric Cazals, Noël Malod-Dognin.

In structural proteomics, given the individual masses of a set of protein types and the exact mass of a protein complex, the *exact stoichiometry determination problem (SD)*, also known as the money-change problem, consists of enumerating all the stoichiometries of these types which allow to recover the target mass. If the target mass suffers from experimental uncertainties, the *interval SD problem* consists of finding all the stoichiometry vectors compatible with a target mass within an interval.

We make contributions in two directions [18]. From a theoretical standpoint, we present a constant-memory space algorithm (`DIOPHANTINE`) and an output sensitive dynamic programming based algorithm (`DP++`), both inherently addressing the interval SD problem. From an applied perspective, we raise three points. First, we show that `DIOPHANTINE` and `DP++` yield an improvement from 3 to 4 orders of magnitude over state-of-the-art exact SD algorithms, for typical protein complexes facing uncertainties on the target mass in the range 0.1-1%. Second, we show that `DIOPHANTINE` behaves like an output-sensitive algorithm—especially when the interval width increases, albeit such a property cannot be expected in general. Third, from a biological perspective, using a panel of biological complexes (eukaryotic translation factor, yeast exosome, 19S proteasome sub-unit, nuclear pore complex), we stress the importance of enumeration, even at a null noise level.

The programs accompanying this paper are available from http://team.inria.fr/abs/addict/.

5.3. Algorithmic Foundations

Voronoi diagrams, α-shapes,

The work undertaken in this vein in 2012 will be finalized in 2013.

5.4. Immunology

Immune response, infection, antibodies, complementarity determining region (CDR)

5.4.1. *Teleost Fish Mount Complex Clonal IgM and IgT Responses in Spleen Upon Systemic Viral Infection*

Participant: Frédéric Cazals.

In collaboration with

- *R. Castro, L. Journeau, A. Benmansour and P. Boudinot (INRA Jouy-en-Josas, France)*
- *H.P. Pham and A. Six (Univ. of Paris VI, France)*
- *O. Bouchez (INRA Castanet Tolosan, France)*
- *V. Giudicelli and M-P. Lefranc (IMGT / CNRS, Montpellier, France)*
- *E. Quillet (INRA Jouy-en-Josas, France)*
- *S. Fillatreau (Leibniz Institute, Berlin, Germany)*
- *O. Sunyer (Univ. of Pennsylvania, USA)*

Upon infection, B-lymphocytes expressing antibodies specific for the intruding pathogen develop clonal responses triggered by pathogen recognition via the B-cell receptor. The constant region of antibodies produced by such developing clones dictates their functional properties. In teleost fish, the clonal structure of B-cell responses and the respective contribution of the three isotypes IgM, IgD, and IdT remains unknown. The expression of IgM and IgT are mutually exclusive, leading to the existence of two B-cell subsets expressing either both IgM and IgD or only IgT. In [12], we undertook a comprehensive analysis of the variable heavy chain (VH) domain repertoires of the IgM, IgD and IgT in spleen of homozygous isogenic rainbow trout (Onchorhynus mykiss), before and after challenge with a rhabdovirus, the Viral Hemorrhagic Septicemia Virus (VHSV), using CDR3-length spectratyping and pyrosequencing of immunoglobulin (Ig) transcripts. In healthy fish, we observed distinct repertoires for IgM, IgD and IgT respectively, with a few amplified μ

and τ junctions, suggesting the presence of IgM and IgT secreting cells in the spleen. In infected animals, we detected complex and highly diverse IgM responses involving all VH subgroups, and dominated by a few large public and private clones. A lower number of robust clonal responses involving only a few VH were detected for the mucosal IgT, indicating that both IgM$^+$ and IgT$^+$ spleen B cells responded to systemic infection but at different degrees. In contrast, the IgD response to the infection was faint. Although the IgD and IgT present different structural features and evolutionary origin compared to mammalian IgD and IgA respectively, their implication in the B-cell response evokes these mouse and human counterparts. Thus, it appears that the general properties of antibody responses were already in place in common ancestors of fish and mammals, and were globally conserved during evolution with possible functional convergences.

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

6. New Results

6.1. Logical time in Model-Driven Engineering embedded design

Participants: Charles André, Frédéric Mallet, Julien Deantoni, Marie-Agnès Peraldi Frati, Arda Goknil, Nicolas Chleq.

6.1.1. TimeSquare

We progressed our work on the foundations of logical time modeling as present in MARTE Time Model and our CCSL clock constraint specification language, while continuing the development of the TimeSquare tool environment which supports this in practice. A technical position paper was presented to the international TOOLS conference [22].

6.1.2. ECL (Event Constraint Language

Our contributions on CCSL and Time Model to the MARTE profile are part of the standard, but so far expressed in a syntax that is clearly distinct of the former UML notations. On the other hand, UML provides a textual language, named OCL, to express well-formedness constraints on diagram models and metamodels. While the original objectives were quite different, it seemed tempting to extend or adapt the general OCL philosophy, and to apply it then to timing and performance constraints as targeted by CCSL. The goal is to able the description of MoCs in an appropriate syntax, at metamodeling level. The result was a new syntax, called ECL for event contraint language, endowed with the well-established, sound timing interpretation as in CCSL. This work was reported in [40].

6.1.3. Logical time clocks to schedule data-flow models

Data-flow models can be used to capture data dependencies from applications, execution platforms and allocations. Most of the time such data dependencies impose only a partial order on the execution of application elements onto the execution platform and allow several allocation schemes. In [38], we have shown how to use logical time and CCSL constraints to capture explicitly the partial order imposed by the data-dependencies without imposing a total order. This work of representation expressivity then paved the way for analysis studies on time refinement, described in 6.3 .

6.1.4. Timing requirement modeling

One of the weak points of UML regarding a complete system design flow is its poor treatment of requirement capture (although this is partly corrected in the SysML profile). When requirements are made on timing aspects and logical time (as in our advocated approach), the relevant syntactic expressivity must be provided. We worked on the definition of a Domain-Specific Language (DSL for Timing Requirements engineering. The results were presented in [24], then applied to system specification in the context of the work described in section 6.6 .

6.2. Semantic translation of CCSL constraints into appropriate Büchi automata for trace recognition

Participants: Frédéric Mallet, Julien Deantoni, Robert de Simone, Ling Yin.

Our CCSL language expresses timing and scheduling constraints for a system, based on the notion of abstract logical clocks providing time events, and contraints linking them with relations of "asynchronous" nature (precedence, faster than) or of "synchronous" origin (subclocking, included in). Of course in a large system design both types coexist, and functional definitions also live next to declarative specifications to allow several timing solutions. Such a solution, called a schedule, must enforce that each logical clock either ticks endlessly, or terminates properly, in a way that globally respects the constraints. In previous works we have shown how a large variety of semantic scheduling constraints from the literature could soundly be represented in CCSL.

This year we focused on the semantic foundation of our CCSL language, by defining a structural operational semantic translation into a specific type of transition systems. Because we deal with infinite traces we had to consider acceptance mechanisms such as Büchi repeated states (as already used for translation of LTL temporal logic formulae in classical model-checking). Next we found out that, while state-labeled acceptance conditions were fine to obtain a direct and intuitive translation of individual constraints, building the composition of such models when dealing with multiple constraints was much easier in the case of *transition-labeled* Büchi automata (with repeated acceptance criteria now on transitions); the theory carries over to such case quite naturally, and has already been studied in the past. Finally, because traces must include infinite occurrences for *each* clock, we had to move to so-called *extended* Büchi automata, again a model already studied previously. We provided a complete semantic translation for all CCSL kernel constructs. Most importantly, we provided an efficient and simple fix-point algorithm to check the existance of a valid schedule, based on the type of automata just defined. This is (we believe) a genuine improvement on existing results, with potential applications outside our direct scope. These results are presented in a technical report, submitted for publication [46].

6.3. Timing refinement for multidimensional dataflow models using MARTE Time Model

Participants: Frédéric Mallet, Julien Deantoni, Jean-Vivien Millo.

Extensions of dataflow process networks have been proposed (as multidimensional SDF) to combine task parallelism (as in traditional process networks) with intensive data parallelism (as proposed in the Array-OL/Gaspard2 formalism developed in the DaRT EPI, for instance). The prospect of scheduling (seen as precise time cycle allocation) is here more complex, because of possible trade-offs between the granularity of treatments at task level *vs.* the size of data arrays that are handled uniformally in parallel inside each task. We considered how these phenomena could be represented (if not solved) inside the framework of MARTE Time Model and logical clocks, so as to handle such design issues in a well-defined MDE approach. Additionally, we used the MARTE platform description to specify how the previous models are refined through mapping allocation. The resulting modeling framework was presented in a journal article [19]. This work was conducted jointly with P. Boulet, from DaRT EPI, and C. Glitia, former DaRT PhD and Aoste postdoc student.

6.4. Process Network analysis

Participants: Robert de Simone, Jean-Vivien Millo.

6.4.1. *K-periodic routing schemes for Network-on-Chip data traffic*

This year we considered more specifically the issue of exploiting the predictable routing schemes of our KRG models, expressed as infinite binary words to indicate the sucessive branching directions at merge/select switch nodes, in order to encode data traffic patterns expanded at compile time, when mapping applications expressed under the form of dataflow process networks onto processor arrays in manycore architectures based on network-on-chip interconnects. To show the potential impact of such predicatble compile-time routing patterns, we stdudied as a typical example a fulll (all-to-all) broadcast algorithm on a mesh topology, connecting mode-less computation nodes as in the theory of cellular automata. This resulted in a precise recursive definition of routing patterns, which achieve an optimal data propagation (broadcast implemented as multicast), given the availability of actual links in the NoC topology. This result was presented at the Autamata'2012 conference [30], and an expanded version is available as technical report [44].

A wider view of the approach, and its potential benefits, are described in a technical report [43], submitted for publication.

6.4.2. Optimal data placement for process network scheduling

The topic of efficient scheduling of dataflow process network traffic to optimize both throughput and buffer queue sizing has given rise to a huge literature starting with seminal works in [49], [47], [56]. It has recently been given new impulse due to the advent of manycore architectures (see above). We conducted a number of theoretical works, to establish how such optimal computation scheduling can be best achieved in configurations where data are evenly distributed and streched in time across the (process) network. While this result is intuitively obvious, we formalized precisely what evenly distributed technically means, with the notion of balanced/mechanical words going a long way back in formal language theory, and we demonstrated that under such assumptions optimal schedules could be constructed *in a fully analytical way*, without any symbolic simulation steps or behavior expansion. The result was accepted for publication in a journal article [20].

6.5. Transformation from MARTE Time Model and CCSL to formal analysis models

Participants: Frédéric Mallet, Ling Yin.

This work was conducted in the context of an on-going collaboration with the Software Engineering Institute (SEI) of East Normal China University (ECNU) at Shanghai, which led altogether in part to the DAESD Associated-team, followed by a LIAMA joint project proposal recently submitted (HADES), and the co-supervision by Frédéric Mallet (together with Professor Jing Liu from ECNU) of the PhD thesis of Yin Ling. Yin Ling spent a one-year visit in our team, funded on a chinese governemental grant.

We studied the efficient and sound formal translation of a subset of CCSL contraints into the PROMELA/SPIN formalism, to benefit from model-checking formal analysis features in this environment. The translation is not completely direct, as synchronous simultaneity is not a native notion of PROMELA, and has to be encoded as atomicity. The motivating principles and translation details are provided in [42]. A similar attempt could be considered in the future, this time with the synchronous model-checker SMV, which allows coumpound instantaneous atomic behaviors.

Another line of research was initiated at ECNU to consider *logical continuous time*, while most of our current work considers only discrete time (while MARTE Time Model considers both). Considerations on *hybrid state diagrams* , inviting the expressive power of formal hierarchical hybrid automata models into the MDE design space of UML MARTE, were investigated in [27].

6.6. Use of MARTE Time Model and Logical Time in automotive design and AUTOSAR/TADL

Participants: Marie-Agnès Peraldi Frati, Julien Deantoni, Arda Goknil.

Precise timing constraint modeling and analysis [26], [33] is a key point for the correct development of automotive electronics. EAST-ADL and AUTOSAR has been adopted as standards in automotive industry. The timing model (TADL :Time augmented Description Language) of these standards raises different issues, mainly concerning the precise modeling of the multi clock characteristics of distributed systems together with parameterized timing expressions. In the ITEA TIMMO-2-USE project [35] 8.3.2.1 , we conducted a work [34], [35], on extending TADL with an explicit notion of multiple time bases for modeling the various temporal referentials used in an automotive design (clocks from different ECUs, motor position, etc). Additionally, timing constraints are augmented with parameters, which can be free at the highest abstraction level and then progressively defined during the design process. As a result, a symbolic timing expression in TADL2 is possibly made of a suitable set of arithmetic operators mixing symbolic identifiers (not necessarily set variables) and referring to different time bases. One typical use of this feature is to capture unknown configuration parameters for time budgeting; another one is to relate constraints in different time-bases to each other. Inherent to this work is also the study of the allowable ranges for symbolic values that are dictated by a set of constraints.

6.7. Multiview modeling and power intent in Systems-on-chip

Participants: Carlos Gomez Cardenas, Ameni Khecharem, Jean-François Le Tallec, Frédéric Mallet, Julien Deantoni, Robert de Simone.

6.7.1. *High-level power management modeling*

One of the concern of the UML MARTE profile is to allow non-functional property modeling, so that the same system bare description can be annotated in a number of views. In our case, combined with our logical time framework, such properties can be made as time-depending, inside potentially distinct views. We examplified this approach by dealing to a large extent with the example of low-power design and energy modeling in the case of Systems-on-Chip (SoC) in the mobile phone domain. Pure power/thermal modeling can be realized, based on the system global architecture, then made operational with the use of logical time controllers triggering power management functionalities.

Thermal/power simulation models are usually relying on continuous time. Therefore we considered the issue of *logical continuous* time, in an early attempt at combining simulation of continuous time power/thermal models with intrinsically discrete functional aspects. A prototype was realized in Scicos, as part of Ameni Khecharem master internship.

This work was conducted in the context of Carlos Gomez PhD thesis, and in collaboration with several partners inside the ANR HeLP project. It should be continued in the forthcoming PhD thesis of Ameni Khecharem, just started in the context of the follow-up ANR HOPE project, which will consider specific issues of hierachical power modeling and compositional power management (as an example of incremental multiview aspects).

6.7.2. *IP-XACT*

In this context of high-level power modeling and multiview concerns, we considered the emerging Accelera standard IP-XACT, made to provide easy-to-plug interfaces and Architecture Description Language (ADL) to allow simple assembly of hardware IP components into well-behaved SoCs. More specifically we provided means to annotate such interface with extra informations, directly borrowed from UML MARTE NFP properties, to handle power and thermal aspects. A number of model transformations back and forth between MARTE and (extended) IP-XACT were realized, and extraction of IP-XACT compliant interfaces from proprietary SystemC code describing the elementary IP component tehmselves has been defined and implemented as well.

This work was initiated as part of a project with STMicroelectronics, inside the nano2012 programme (ended 2011), and continued as part of the ANR HeLP collaboration. It resulted in the PhD thesis of Jean-François Le Tallec (who remained in the team for a couple of months later to complete the prototype implementation) [16].

6.8. Correct and efficient implementation of polychronous formalisms

Participants: Thomas Carle, Manel Djemal, Dumitru Potop Butucaru, Robert de Simone, Yves Sorel.

We extended our work on 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) and non-functional constraints (partitioning).

6.8.1. *Time-Triggered Platform targets*

Our first result this year concerns the automatic scheduling and code generation for time-triggered platforms. We extended our previous results in two significant ways. First, we designed a novel approach for specification of real-time features of time-triggered systems, with deadlines longer than periods; this allows a faithful representation of complex end-to-end flow requirements. Second, we provided new algorithms for off-line pipelined scheduling of these specifications onto partitioned time-triggered architectures *à la* ARINC 653; allocation of time slots/windows to partitions can be either complete or partially provided, or synthesized by our tool. Automatic allocation and scheduling onto multi-processor (distributed) systems with a global time base becomes feasible, taking into account communication costs. For single processors, we allow the generation of fully compliant ARINC653/APEX implementation code.

This work was mainly carried out inside the FUI Parsec 8.2.2.2 (which funds the PhD thesis of T. Carle) and P 8.2.2.1 projects, as well as a collaboration with ASTRIUM Space Transportation. First results are presented in a technical report, submitted for publication [39].

6.8.2. *Multi-Processor System-on-Chip (MP-SoC) targets*

Our second contribution concerns the automatic allocation and real-time scheduling over MPSoC (multi-processor on chip) architectures with NoC (network-on-chip) interconnect. One must take into account the specific 2D mesh network-on-chip topology, and synthesize the NoC routing patterns. This work provides operational execution support for the contributions described in 6.9 .

6.8.3. *The LoPhT tool*

Our recent work on extending the AAA methodology with better handling of execution conditions, with pipelining and pipelined scheduling, and with specific real-time scheduling and code generation techniques for time-triggered/partitioned and MPSoC platforms resulted in the development of a new scheduling and code generation toolbox, called LoPhT (for Logical to Physical Time Compiler).

6.9. Programmable On-Chip Networks

Participants: Thomas Carle, Manel Djemal, Dumitru Potop Butucaru, Robert de Simone, Zhen Zhang.

Modern computer architectures are increasingly relying on multi-processor systems-on-chip (MPSoCs), with data transfers between cores and memories managed by on-chip networks (NoC). This reflects in part a convergence between embedded, general-purpose PC, and high-performance computing (HPC) architecture designs.

Efficient compilation of applications onto MPSoCs remains largely an open problem, with the issue of best mapping of computation parts (threads, tasks,...) onto processing resources amply recognized, while the issue of best use of the interconnect NoC to route and transfer data still less commonly tackled. In the most general case, dynamic allocation of applications and channel virtualization can be guided by user-provided information under various forms, as in OpenMP, CUDA, OpenCL and so on. But then there is no clear guarantee of optimality, and first attempts by non-experts often show poor performances in the use of available computing power. Conversely there are consistent efforts, in the domains of embedded and HPC computing, aiming at automatic parallelization, compile-time mapping and scheduling optimization. They rely on the fact that applications are often known in advance, and deployed without disturbance from foreign applications, and without uncontrolled dynamic creation of tasks. Our contribution follows this "static application mapping" approach.

An optimal use of the NoC bandwidth should authorize data transfers to be realized according to (virtual) channels that are temporarily patterned to route data "just-in-time". Previous works have identified the need for Quality of Service (QoS) in "some" data connections across the network (therefore borrowing notions from macroscopic networks, say internet and its protocols). But our experience with the AAA methodology strongly suggests that optimal NoC usage should result from a global optimization principle (embodied in a form of the AAA methodology), as opposed to a collection of local optimizations of individual connections. Indeed, various data flows with distinct sources and targets will nevertheless be highly concerted, both in time and space, like in a classical pipelined CPU, where the use of registers (replaced in our case with a complex NoC) is strongly synchronized with that of the functional units.

One main problem in applying such a global optimization approach is to provide the proper hardware infrastructures allowing the implementation of optimal computation and communication mappings and schedules. Our thesis is that optimal data transfer patterns should be encoded using simple programs configuring the router nodes (each router being then programmed to act its part in the global concerted computation and communication scheme).

We addressed this problem in the framework of our collaboration 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. We concretely supported our proposed approach by extending the DSPIN 2D mesh network-on-chip (NoC) developed at UPMC- LIP6. In this NoC, we replace the fair arbitration modules of the NoC routers with static, micro-programmable modules that can enforce a given packet routing sequence, as specified by small programs. The design of such simple routing schemes can, for instance, be extracted from our results in section 6.4 .

We advocate the desired level of expressiveness/complexity for such simple configuration programs, and provide experimental data (cycle-accurate simulations) supporting our choices. We also wrote an architecture synthesis tool that allows simple architectural exploration of MPSoCs using the new DSPINPro NoC. First results in this direction have been presented in the DASIP 2012 conference, where our paper [23] has been short-listed for best paper award.

6.10. Uniprocessor Real-Time Scheduling

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

6.10.1. Combination of Non-Preemptive and Preemptive Tasks

We focused on fixed priority scheduling for a combination of non-preemptive strict periodic tasks in conjunction with preemptive sporadic tasks, that we extended to software fault tolerance [29]. We first investigated the transient phase for non-preemptive strict periodic tasks and we proved that its length is smaller than the transient phase for preemptive periodic tasks. Then, we determined the worst case scenario for preemptive sporadic tasks where the Worst Case Response Time (WCRT) can be obtained in the presence of strict periodic tasks. We proved that these release times belong only to the permanent phase of strict periodic tasks, and thus that the schedulability analysis for sporadic tasks can be restricted to the permanent phase. For preemptive sporadic tasks, we extended the classical necessary and sufficient schedulability condition based on the worst case response time computation to take into account non-preemptive strict periodic tasks. Finally, we considered software fault tolerance in the particular case where each primary strict periodic task has an alternate sporadic task which is released when the primary task fails. The schedulability analysis guarantees that even if all strict periodic tasks fail then all their respective alternate tasks will meet their deadlines.

6.10.2. Formal Proofs of Real-Time Scheduling Theorems

We completed two formal proofs of theorems in Coq on scheduling of fixed priority real-time preemptive tasks: one dealing with the sizes of busy periods (about 3500 lines of Coq), and another one dealing with response time (about 5200 lines of Coq). A monograph about these proofs, together with the formal check in Coq of scheduling conditions of strict periodicity, presented in the conference JFLA 2012 [37], have been started (currently about 70 pages).

6.11. Multiprocessor Real-Time Scheduling

Participants: Abderraouf Benyahia, Laurent George, Mohamed Marouf, Falou Ndoye, Simon Nivault, Yves Sorel, Cécile Stentzel, Meriem Zidouni.

6.11.1. Non-Preemptive Partitioned Fault Tolerant Scheduling

We addressed partitioned multiprocessor scheduling of non-preemptive strict periodic tasks which is extended thereafter to hardware fault tolerance [17].

In order to schedule a task set of non-preemptive strict periodic tasks on a multiprocessor platform, we partitioned this task set into subsets of tasks, each one is scheduled on a single processor using our proposed uniprocessor scheduling algorithm. The partition is carried out according to an enhanced "First Fit" algorithm that balances the load of the tasks on all the processors. However, inter-processors communications can lead to delay task execution. Thus, we determined the start time of each task taking into account the communication delay between this latter task and its predecessor tasks. Also, as inter-processor communications may generate a transient phase, we computed the length of the transient phase.

We proposed a fault tolerant real-time scheduling algorithm which allows hardware processors and/or buses faults, and conserves the strict periodicity of each task. We also proposed a graph transformation algorithm, applied on the task graph, which generates redundancies of tasks as well as dependencies. The transformation adds also selector tasks which choose data coming from the non failing processors and buses. That algorithm is based on exclusion relations to assign redundant tasks (resp. dependencies) to different processors (resp. busses). Then, we extended the previous partitioned multiprocessor scheduling algorithm to manage fault tolerance taking into account these exclusion relations.

This approach was successfully implemented on a CyCabs electric vehicle in a real-time fault tolerant tracking application where some processor or some bus could fail without any consequence on the proper execution of the application, i.e. same functional behaviour and real-time constraints satisfied.

6.11.2. Partitioned Scheduling with Exact Preemption Cost

Preemption allows a better scheduling success ratio but has a cost that must not be neglected in safety critical applications of domains such as avionic, automotive, etc. We focused on partitioned multiprocessor scheduling of independent preemptive periodic real-time tasks, while taking into account the exact preemption cost with the $\oplus$ operation formerly proposed by Meumeu and Sorel [10]. We improved the "greedy" heuristic proposed last year and compared it with the "Best-Fit" (BF) and "Worst-Fit" (WF) heuristics classically used in partitioned multiprocessor scheduling, but extended to take into account the exact preemption cost. We also compared our heuristic with an exact "Branch and Bound" algorithm with the same extension. The first comparison shows that the task allocation found by our heuristic gives a better response time than those found by WF and BF. This is due to the fact that the execution of the tasks is better parallelized. On the other hand, BF and WF heuristics execute a bit faster than our heuristic because they do not use all the available processors contrary to our heuristic which has the advantage to improve the load balancing of the tasks on all the processors.

Then, we addressed the scheduling of preemptive periodic real-time tasks with dependence constraints involving task precedences and data dependences. We considered harmonic tasks, i.e. periods of tasks are multiple or equal, to avoid loss of data. In order to satisfy data dependence constraints, we modified the release dates and deadlines of the dependent tasks according to the reception date of the data. In addition, data dependences between tasks mean to share data between dependent tasks which can cause deadlock and priority inversion problems. In order to solve these problems while taking into account the preemption cost, we proposed a new schedulability condition based on an extension of the $\oplus$ operation. We plan to propose a mutiprocessor scheduling heuristic based on that condition applied on tasks with modified release dates and deadlines.

6.11.3. Semi-partitioned Scheduling

Semi-partitioned multiprocessor scheduling stands between partitioned and global scheduling, the latter allowing migrations. We mainly addressed the semi-partitioned scheduling approach 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. A job is migrated at its local deadline, computed from the deadline of the task it belongs to. We have studied this approach in the context of a fork/join task model with thread parallelism. A task is composed of a sequence of segments that can be parallelized in threads, if needed. The local deadlines depends on the number of parallel threads assigned to each segment.

6.11.4. Code Generation for Multicore

This work was carried out in the OPENPROD ITEA project 8.3.2.2 . xMod developed by IFPEN (IFP Energies Nouvelles), is an heterogeneous model integration environment that allows model importation from specific tools such as Simulink, AMSIM, etc. It also provides as a virtual instrumentation laboratory. In order to make xMod being able to run simulations with hardware-in-the-loop environment, we developed a new SynDEx executive kernel based on the kernel, dedicated to Windows/RTX, developed last year. That executive kernel is used with the macro-code generated by SynDEx to produce a real-time executable code that can drive the execution (real-time multi-core distribution and synchronized execution) of the models imported by xMod

and simulated in the virtual instrumentation laboratory. This prototype as well as the report describing the corresponding achieved works, are the final deliverable of the OPENPROD project.

Furthermore, a French and English SynDEx code generation reference manual has been written to help future SynDEx users and maintainers to generate real-time code for already supported architectures or new ones.

6.11.5. Gateway with Modeling Languages for Certified Code Generation

This work was carried out in the P FUI project 8.2.2.1 . We provide inside the project expertise mainly on schedulability analysis and automatic generation of distributed real-time code. In this context, we developed a gateway between UML/MARTE and SynDEx. From a model specified with UML (Activity Diagram to specify algorithms and Composite Structure Diagram to specify multicomponent architectures) and refined with the UML profile MARTE (Modeling and Analysis of Real-Time Embedded Systems), we use the gateway to generate automatically distributed real-time application specified in the SynDEx format. Currently, we intend to provide a gateway between the GeneAuto language and SynDEx. The GeneAuto language is a subset of the future pivot P language. We presently deal with the part of the GeneAuto language corresponding to Simulink for data-flow modeling and we plan to deal soon with the part corresponding to Stateflow for control-flow modeling (composition of automata).

6.11.6. SynDEx Updates

We continued the software developments for the 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 completed the tests on the new automatic code generator for the current version 7 of SynDEx. This new generator produces code for mono-periodic and multi-periodic applications with condition and repetitive control structures, for the different hardware architectures supported by SynDEx. We developed a checker for the generated code that was integrated in the new generator. This checker verifies the correct use of semaphores and consequently the absence of deadlocks in the real-time code. Deadlocks are the most difficult part when dealing with distributed architectures. We achieved also a maintenance report describing the structure and the main features of code generator, as well as the technical choices we did.

6.12. Variability of program execution times on multicore processors

Participants: Sid-Ahmed-Ali Touati, Matias Vara Larsen, Abdelhafid Mazouz.

The activity described here represents the finalization of previous efforts conducted by Sid Touati and members of his groups, initiated before he joined the AOSTE EPI, and which are progressively merged with our own objectives, for results to be reported hopefully next year).

With the massive introduction of multicore platforms on embedded systems, parallel applications gained in performance. However, we showed in previous studies that the performance gain comes with high instability: program execution times vary in important way. We investigated the reasons for this variations and tried to understand the factors that influence program performance variability, that we decompose intro multiple families: factors from the application itself (implemented algorithms, coding technique, synchronization barriers, etc.), factors from the execution environment (OS effects, thread scheduling, Input/Ouput operations) and factors from the underlying hardware (micro-architecture, memory hierarchy, speculative execution, hardware data prefectching, etc.). Now, we have better understanding to these factors thanks to the work of two students:

1. Mr. Abdelhafid Mazouz who defended his PhD under the direction of Sid Touati at the university of Versailles in 11th of December 2012. The title of his PhD is "An Empirical Study of Program Performance of OpenMP Applications on Multicore Platforms".

2. Mr. Matias Vara Larsen, intern under the supervision of Sid Touati from February to June 2012, inside the Aoste EPI in Sophia-Antipolis, co-funded under a grant from Inria international internship program). The topic of his internship was to study the influence of he Linux kernels (multiple versions) on the stability of parallel applications.

Last, we published a rigorous statistical protocol in [21] called the Speedup-Test. It is used to analyze valid speedups (performance gain) in presence of performance instability: The Speedup-Test protocol is implemented and distributed as an open source tool based on R software. Our statistical methodology defines a consistent improvement compared with the usual performance analysis method in high-performance computing.

APICS Project-Team

6. New Results

6.1. Source recovery problems

Participants: Laurent Baratchart, Sylvain Chevillard, Juliette Leblond, Ana-Maria Nicu.

The works presented here are done in collaboration with Maureen Clerc and Théo Papadopoulo from the Athena EPI, with Doug Hardin and Edward Saff from Vanderbilt University (Nashville, USA), and with Abderrazek Karoui (Univ. Bizerte, Tunisie) and Jean-Paul Marmorat (Centre de mathématiques appliquées (CMA), École des Mines de Paris).

This section in dedicated to inverse problems for 3-D Poisson-Laplace equations. Though the geometrical settings differ in the 2 sections below, the characterization of silent sources (that give rise to a vanishing potential at measurement points) is a common problem to both which has been recently achieved, see [37],[29], [39].These are sums of (distributional) derivatives of Sobolev functions vanishing on the boundary.

6.1.1. *Application to EEG*

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 [4] 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 milestone in a long-haul research on the behaviour of poles of best rational approximants of fixed degree to functions with branch points has been reached this year [14], which shows that the technique carries over to finitely many sources (see section 4.2). In this connection, a dedicated software "FindSources3D" (see section 5.6) has been developed, in collaboration with the team Athena [16], [26].

Further, it appears that in the rational approximation step of these schemes, *multiple* poles possess a nice behaviour with respect to the branched singularities. This is due to the very basic physical assumptions on the model (for EEG data, one should consider *triple* poles). Though numerically observed in [16], there is no mathematical justification so far why these multiple poles have such strong accumulation properties, which remains an intriguing observation.

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 [13], [67]), while other techniques should be considered as well.

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

It turns out that discretization issues in geophysics can also be approached by these approximation techniques. Namely, in geodesy or for GPS computations, one may need to get a best discrete approximation of the gravitational potential on the Earth's surface, from partial data collected there. This is the topic of a beginning collaboration with a physicist colleague (IGN, LAREG, geodesy). Related geometrical issues (finding out the geoid, level surface of the gravitational potential) are worthy of consideration as well.

6.1.2. *Magnetization issues*

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 (*e.g.* meteorites or stalactites) from fine field measurements 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. The problem of recovering a thin plate magnetization distribution from measurements of the field in a plane above the sample lead us to an analysis of the kernel of this operator, which we characterized in various function and distribution spaces (arbitrary compactly supported distributions or derivatives of bounded functions). For this purpose, we introduced a generalization of the Hodge decomposition in terms of Riesz transforms and showed that a thin plate magnetization is "silent" (i.e. in the kernel) if the normal component is zero and the tangential component is divergence free. In particular, we show that a unidirectional non-trivial magnetization with compact support cannot be silent. The same is true for bidirectional magnetizations if at least one of the directions is nontangential. We also proved that any magnetization is equivalent to a unidirectional. We did introduce notions of being silent from above and silent from below, which are in general distinct. These results have been reported in a paper to appear [37].

We currently work on Fourier based inversion techniques for unidirectional magnetizations, and Figures 5 , 6 , 7 and 8 show an example of reconstruction. A joint paper with our collaborators from VU and MIT is being written on this topic.
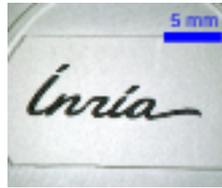

Figure 5. Inria's logo were printed on a piece of paper. The ink of the letters "In" were magnetized along a direction D_1. The ink of the letters "ria" were magnetized along another direction D_2 (almost orthogonal to D_1).

Figure 6. The Z-component of the magnetic field generated by the sample is measured by a SQUID microscope. The measure is performed $200\mu m$ above the sample.

For more general magnetizations, the severe ill-posedness of reconstruction challenges discrete Fourier methods, one of the main problems being the truncation of the observations outside the range of the SQUID measurements. We look forward to develop extrapolation techniques in the spirit of step 1 in section 3.1 .

6.2. Boundary value problems, generalized Hardy classes

Participants: Laurent Baratchart, Slah Chaabi, Juliette Leblond, Dmitry Ponomarev.

This work has been performed in collaboration with Yannick Fischer from the Magique3D EPI (Inria Bordeaux, Pau).

Figure 7. The field measured in Figure 6 is inversed, assuming that the sample is unidimensionally magnetized along the direction D_1. The letters "In" are fairly well recovered while the rest of the letters is blurred (because the hypothesis about the direction of magnetization is false for "ria").

Figure 8. The field measured in Figure 6 is inversed, assuming that the sample is unidimensionally magnetized along the direction D_2. The letters "ria" are fairly well recovered while the rest of the letters is blured (because the hypothesis about the direction of magnetization is false for "In").

In collaboration with the CMI-LATP (University Aix-Marseille I), 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 [5].

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 3.3.1 .

We generalized the construction to finitely connected Dini-smooth domains and $W^{1,q}$-smooth conductivities, with $q > 2$ [35]. The case of an annular geometry is the relevant one for the application to plasma shaping mentioned below [58], [35]. 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).

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 [54]. 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 [35].

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 provides a computational setting to solve the extremal problems mentioned before, and was the topic of the PhD thesis of Y. Fischer [58], [27]. 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, and the topic of the PhD thesis of D. Ponomarev.

The PhD work of S. Chaabi is devoted to further aspects of Dirichlet problems for the conjugate Beltrami equation. On the one hand, a method based on Foka's approach to boundary value problems, which uses Lax pairs and solves for a Riemann-Hilbert problem, has been devised to compute in semi explicit form solutions to Dirichlet and Neumann problems for the conductivity equation satisfied by the poloidal flux. Also, for more general conductivities, namely bounded below and lying in $W^{1,s}$ with $s \geq 2$, parameterization of solutions to Dirichlet problems on the disk by Hardy function was achieved through Bers-Nirenberg factorization. Note the conductivity may be unbounded when $s = 2$, which is completely new. Two papers are being prepared reporting on these topics.

Finally, note that the conductivity equation can be expressed like a static Schrödinger equation, for smooth enough conductivity coefficients. This provides a link with the following results recently set up by D. Ponomarev, who recently join the team for his PhD. A description of laser beam propagation in photopolymers can be crudely approximated by a stationary two-dimensional model of wave propagation in a medium with negligible change of refractive index. In such setting, Helmholtz equation is approximated by a linear Schrödinger equation with one of spatial coordinates being an evolutionary variable. Explicit comparison of the solutions in the whole half-space allows to establish global justification of the Schrodinger model for sufficiently smooth pulses [73]. This phenomenon can also be described by a nonstationary model that relies on the spatial nonlinear Schrödinger (NLS) equation with the time-dependent refractive index. A toy problem is considered in [71], when the rate of change of refractive index is proportional to the squared amplitude of the electric field and the spatial domain is a plane. The NLS approximation is derived from a 2-D quasi-

linear wave equation, for small time intervals and smooth initial data. Numerical simulations illustrate the approximation result in the 1-D case.

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

Participant: Fabien Seyfert.

This work has been done in collaboration with Smain Amari (Royal Military College, Kingston, Canada), Jean Charles Faugère (SALSA EPI, Inria Rocquencourt), Giuseppe Macchiarella (Politecnico di Milano, Milan, Italy), Uwe Rosenberg (Design and Project Engineering, Osterholz-Scharmbeck, Germany) and Matteo Oldoni (Politecnico di Milano, Milan, Italy).

We continued our work on the circuit realizations of filters' responses with mixed type (inductive or capacitive) coupling elements and constrained topologies [1]. For inline circuits, methods based on sequential extractions of electrical elements are best suited due to their computational simplicity. On the other hand, for circuits with no inline topology ,such methods are inefficient while algebraic methods (based on a Groebner basis) can be used, but at high computational cost. In order to tackle large order circuits, our approach is to decompose them into connected inline sections, which can be directly realized by extraction techniques, and into complex sections, where algebraic methods are needed for realization. In order to do this, we started studying the synthesis of filter responses by means of circuits with reactive non-resonating nodes (dangling resonators) [22]. Links of this topic with Potapov's factorization of J-inner functions are currently being investigated.

In this connection, sensitivity analysis of the electrical response of a filter with respect to the electrical parameters of the underlying circuit has been published in collaboration with the University of Cartagena and ESA [20]. We essentially proved that the total electrical sensitivity of a filters' response does not depend on the coupling topology of the underlying circuit: the latter however controls the distribution of this sensitivity within each resonator.

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

Participants: Martine Olivi, Sanda Lefteriu, Fabien Seyfert.

This work has been done in collaboration with Stéphane Bila (Xlim, Limoges, France), Hussein Ezzedin (Xlim, Limoges, France), Damien Pacaud (Thales Alenia Space, Toulouse, France), Giuseppe Macchiarella (Politecnico di Milano, Milan, Italy, and Matteo Oldoni (Politecnico di Milano, Milan, Italy).

6.4.1. Synthesis of compact multiplexers

We focused our research on multiplexer with a star topology. These are comprised of a central N-port junction, and of filters plugged on all but common ports (see Figure 9). A possible approach to synthesis of the multiplexer's response is to postulate that each filter channel has to match the multiplexer at n_k frequencies (n_k being the order of the filter) while rejecting the energy at m_k other frequencies (m_k being the order the transmission polynomial of the filter). The desired synthesis can then be cast into computing of a collection of filter's responses matching the energy as prescribed and rejecting it at specified frequencies when plugged simultaneously on the junction. Whether such a collection exists is one of the main open issues facing co-integration of systems in electronics. Investigating the latter led us to consider the simpler problem of matching a filter, on a frequency-varying load, while rejecting energy at fixed specified frequencies. If the order of the filter is n this amounts to fix a given transmission polynomial r and to solve for a unitary polynomial p meeting interpolation conditions of the form:

$$j = 1..n, \qquad \frac{p}{q}(w_j) = \gamma_j$$

where q is the unique monic Hurwitz polynomial satisfying the Feldtkeller equation

$$qq^* = pp^* + rr^*.$$

This problem can be seen as an extended Nevanlinna-Pick interpolation problem, which was considered in [62] when the interpolation frequencies lie in the *open* left half-plane. We conjecture that existence and uniqueness of the solution still holds in our case, where interpolation takes place on the boundary, provided r has no roots on the imaginary axis. Numerical experiments based on continuation techniques tend to corroborate our belief: efforts now focus on a mathematical proof. The derived numerical tools have already been used to successfully to design multiplexer's responses in collaboration with CNES and Xlim, thereby initiating a collaboration with Xlim on co-integration of filters and antennas.

6.4.2. De-embedding of multiplexers

Let S be the measured scattering matrix of a multiplexer composed of a N-port junction with response T and $N-1$ filters with responses $F_1, \cdots F_{N-1}$ as plotted on Figure 9 . The de-embedding question we raise is the following: given S and T, is it possible to retrieve the F_k's ? The answer to this question depends of course of the admissible class of filters. For the simplest case where no assumption is made (except reciprocity), we showed that the problem has a unique solution for $N > 3$ and for generic T, while for $N = 2$ the solution space at each frequency point has real dimension 2. This redundancy can be explained by the existence of "ghost" or "silent" components that can hide behind the junction: when being chained to the junction these components do not affect its response. We also experienced that the generic behaviour for $N > 3$ is rather theoretical, as usual junctions are often made of chained T-junctions: in this non generic case (which is rather generic in practice !) some "silent" component still exists for $N > 3$. Additional hypotheses, such as rationality with prescribed degree for F_k, are currently being studied and already yielded results for the case $N = 3$ [21].

This work is pursued in collaboration with Thales Alenia Space, Politecnico di Milano, Xlim and CNES in particular within the contract CNES-Inria on compact N-port synthesis (see section 7.1).
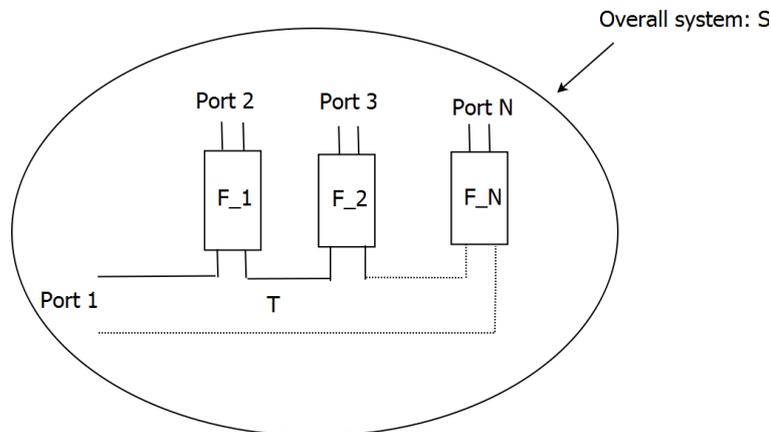

Figure 9. Multiplexer made of a junction T and filtering devices $F_1, F_2 \cdots F_N$

6.5. Detection of the instability of amplifiers

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

This work is conducted in collaboration with Jean-Baptiste Pomet from the McTao team. It is a continuation of a collaboration with CNES and the University of Bilbao.The goal is to help developing 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, based on properties of transfer functions of such amplifiers which belong to particular classes of analytic functions.

In non-degenerate cases, non-linear electrical components can be replaced by their first order approximation when studying stability to small perturbations. Using this approximation, diodes appear as perfect negative resistors and transistors as perfect current sources controlled by the voltages at certain points of the circuit.

In 2011, we had proved that the class of transfer functions which can be realized with such ideal components and standard passive components (resistors, selfs, capacitors and transmission lines) is rather large since it contains all rational functions in the variable and in the exponentials thereof.

In 2012, we focused on the kind of instabilities that these ideal systems can exhibit. We showed that a circuit can be unstable, although it has no pole in the right half-plane. This remains true even if a high resistor is put in parallel of the circuit, which is rather unusual. This pathological example is unrealistic, though, because it assumes that non-linear elements continue to provide gain even at very high frequencies. In practice, small capacitive and inductive effects (negligible at moderate frequencies) make these components passive for very high frequencies. Under this simple assumption, we proved that the class of transfer functions of realistic circuits is much smaller than in previous situation. In fact, a realistic circuit is unstable if and only if it has poles in the right half-plane. Moreover, there can only be finitely many of them. An article is currently being written on the subject.

6.6. Best constrained analytic approximation

Participants: Laurent Baratchart, Sylvain Chevillard, Juliette Leblond, Dmitry Ponomarev, Elodie Pozzi.

This work is performed in collaboration with Jonathan Partington (Univ. Leeds, UK).

Continuing effort is being paid by the team to carry over the solution to bounded extremal problems of section 3.3.1 to various settings. We mentioned already in section 6.2 the extension to 2-D diffusion equations with variable conductivity for the determination of free boundaries in plasma control and the development of a generalized Hardy class theory. We also investigate the ordinary Laplacian in $\mathbb{R}^3$, where targeted applications are to data transmission step for source detection in electro/magneto-encephalography (EEG/MEG, see section 6.1).

Still, questions about the behaviour of solutions to the standard bounded extremal problems (P) of section 3.3.1 deserve attention. We realized this year that Slepian functions are eigenfunctions of truncated Toeplitz operators in 2-D. This can be used to quantify robustness properties of our resolution schemes in H^2 and to establish error estimates, see [25]. Moreover we considered additional interpolation constraints [28], as a simplified but already interesting issue, before getting at extremal problems for generalized analytic functions in annular non-smooth domains. The latter arise in the context of plasma shaping in tokamaks like ITER, and will be the subject of the PhD thesis of D. Ponomarev.

In another connection, weighted composition operators on Lebesgue, Sobolev, and Hardy spaces appear in changes of variables while expressing conformal equivalence of plane domains. A universality property related to the existence of invariant subspaces for these important classes of operators has been established in [19]. Additional density properties also allow one to handle some of their dynamical aspects (like cyclicity).

6.7. Rational Approximation for fitting Non-Negative EPT densities

Participants: Martine Olivi, Fabien Seyfert.

This work has been done in collaboration with Bernard Hanzon and Conor Sexton from Univ. Cork.

The problem is to fit a probability density function on a large set of financial data. The model class is the set of non-negative EPT (Exponential-Polynomials-Trigonometric) functions which provides a useful framework for probabilistic calculation as illustrated in the link http://www.2-ept.com/2-ept-literature.html. Moreover, an EPT function can alternatively be interpreted as the impulse response of a continuous time stable system whose Laplace transform is a rational transfer function. This interpretation allows us to approach this problem using approximation tools developed by the team. The very context brings up a classical, as yet essentially unsolved difficulty in rational approximation, namely preservation of positivity. This is known to be a hard issue. Our work, initiated in 2011, resulted this year in an improved approach for checking non-negativity of an EPT function. These results have been presented at the 16th IFAC Conference on System Identification [23]. The proposed method was demonstrated on the positive daily Dow Jones Industrial Average (DJIA) log returns over 80 years.

6.8. Rational and meromorphic approximation

Participant: Laurent Baratchart.

This work has been done in collaboration with Herbert Stahl (TFH Berlin) and Maxim Yattselev (Univ. Oregon at Eugene, USA).

We completed and published this year the proof of an important result in approximation theory, namely the counting measure of poles of best H^2 approximants (more generally: of critical points) of degree n to a function analytically continuable, except over finitely many branchpoints lying outside the unit disk, converges to the Green equilibrium distribution of the compact set of minimal Green capacity outside of which the function is single valued [14]. The proof requires showing existence and uniqueness of a compact set of minimal weighted logarithmic capacity in a field, outside of which the function is single-valued. Structure of this contour, along with error estimates, also come out of the proof. The result is in fact valid for functions that are Cauchy integrals of Dini-smooth functions on such a contour. We rely in addition on asymptotic interpolation estimates from [63].

This result warrants source recovery techniques used in section 6.1.1 .

We also studied partial realizations, or equivalently Padé approximants to transfer functions with branchpoints. Identification techniques based on partial realizations of a stable infinite-dimensional transfer function are known to often provide unstable models, but the question as to whether this is due to noise or to intrinsic instability was not clear. In the case of 4 branchpoints, expressing the computation of Padé approximants in terms of the solution to a Riemann-Hilbert problem on the Riemann surface of the function, we proved that the pole behaviour generically shows deterministic chaos [49].

6.9. Tools for numerically guaranteed computations

Participant: Sylvain Chevillard.

The overall and long-term goal is to enhance the quality of numerical computations. The progress made during year 2012 is the following:

- Publication of a work about the implementation of functions erf and erfc in multiple precision and with correct rounding [15]. It corresponds to a work initially begun in the Arénaire team and finished in the Caramel team. The goal of this work is to show on a representative example the different steps of the rigorous implementation of a function in multiple precision arithmetic (choice of a series approximating the function, choice of the truncation rank and working precision used for the computation, roundoff analysis, etc.). The steps are described in such a way that they can easily be reproduced by someone who would like to implement another function. Moreover, it is showed that the process is very regular, which suggests that it (or at least large parts of it) could be automated.

- In the same field of multiple precision arithmetic, and with Marc Mezzarobba (Aric team), we proposed an algorithm for the efficient evaluation of the Airy $Ai(x)$ function when x is moderately large [57]. Again, this work deals with a representative example, with the idea of trying to automate

the process as a future work. The Taylor series of the Airy Ai function (as many others such as, e.g., Bessel functions or erf) is ill-conditioned when x is not small. To overcome this difficulty, we extend a method by Gawronski, Müller and Reinhard, known to solve the issue in the case of the error function erf. We rewrite Ai(x) as $G(x)/F(x)$ where F and G are two functions with well-conditioned series. However, the coefficients of G turn out to obey a three-terms ill-conditioned recurrence. We evaluate this recurrence using Miller's backward algorithm with a rigorous error analysis.

- 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 [2], developed in collaboration with C. Lauter (Université Pierre et Marie Curie) and M. Joldeş (Uppsala University). During year 2012, a large effort has been made in view of the release of version 4.0 (to come in 2013). This effort (of about 400 commits in the svn repository of the project) is mainly intended to provide a library version of Sollya, as well as a robust test suite for the tool. As a matter of course, it allowed us to detect and fix a dozen of bugs.

---

[2]http://sollya.gforge.inria.fr/

<h1 style="text-align:center; color:red;">ASCLEPIOS Project-Team</h1>

5. New Results

5.1. Medical Image Analysis

5.1.1. Brain tumor cell density estimation from multi-modal MR images based on a synthetic tumor growth model

Participants: Ezequiel Geremia [Correspondant, Inria], Nicholas Ayache [Inria], Antonio Criminisi [MSRC], Bjoern Menze [Inria,ETHZ], Marcel Prastawa [University of Utah].

Published in the proceedings of the MCV Workshop at MICCAI 2012 [35]

biophysiological tumor growth simulator, multi-variate regression random forests, gliomas, MRI

- A generative-discriminative framework is presented to learn model-based estimations of the tumor cell density
- The ground truth for tumor cell density is very hard to obtain
- A biophysiological tumor growth simulator is used to generate the ground truth tumor cell densities and associated MRIs
- A multi-variate regression random forests is trained to estimate the voxel-wise distribution of tumor cell density from input MR images
- The training data contains 500 synthetic cases and their associated ground truth generated by the brain tumor simulator
- The method was tested on 200 synthetic cases with excellent results
- The method also provided very promising results for estimating the tumor cell density on 16 clinical cases showing low grade gliomas from the DKFZ (German Cancer Research Center)

5.1.2. Automatic indexation of cardiac MR images

Participants: Jan Margeta [Correspondant], Nicholas Ayache, Antonio Criminisi [MSRC].

This work has been partly supported by Microsoft Research through its PhD Scholarship Programme and the European Research Council through the ERC Advanced Grant MedYMA (on Biophysical Modeling and Analysis of Dynamic Medical Images).

Machine learning, Cardiac MR, MR preprocessing

- A generic random forest framework has been implemented and its recent modifications have been applied to a fully automatic and a semisupervised image segmentation methods, and manifold learning in cardiac MRI.
- We have performed image based cardiac function quantification from preprocessed cardiac cine MRI sequences.
- An image intensity standardization in magnetic resonance images method has been proposed.

### 5.1.3. Multimodal brain tumor segmentation

Participant: Bjoern Menze [Correspondant].

MICCAI 2012

- Further developed the generative brain tumor segmentation model
- Developed a generative-discriminative model for multimodal brain tumor segmentation
- Developed a new regularization approach for longitudinal tumor segmentation (with Guillaume Charpiat & Yuliya Tarabalka, Inria Sophia-Antipolis)
- Initiated and co-organized an international benchmark on multimodal brain tumor segmentation as a challenge workshop during MICCAI 2012 in Nice (http://www2.imm.dtu.dk/projects/BRATS2012)
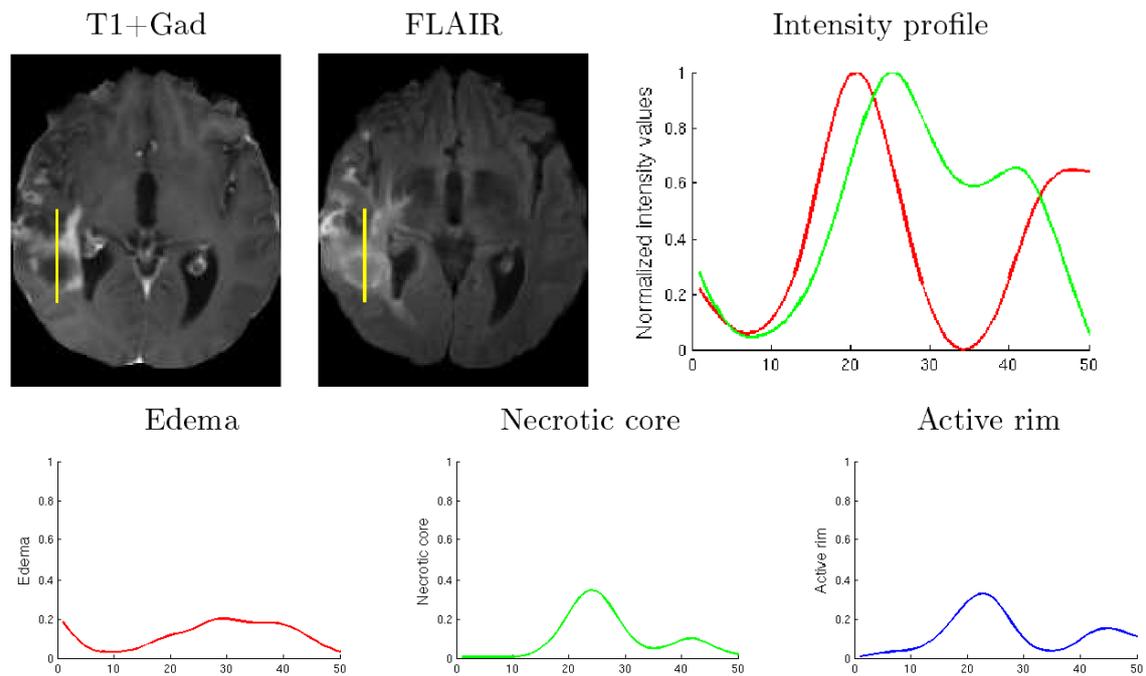
Figure 1. **Prediction of the cell densities along a section of the tumor.** *Top, from left to right: T1+Gadolinium, FLAIR image, the intensity profile along the section (yellow). Bottom, from left to right: prediction of the cell density for the edema, necrotic core and active rim, respectively.*
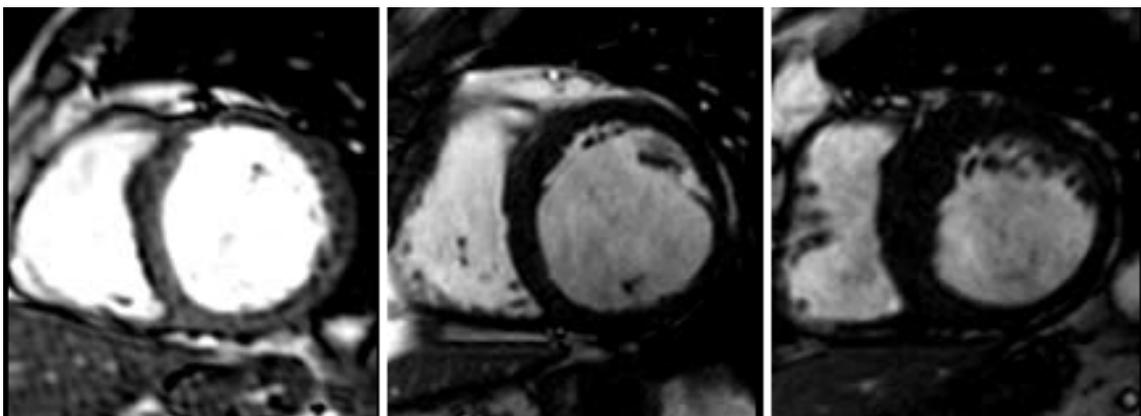
Figure 2. Three different patients with increasing (from left to right) functional deficiency of the heart.

5.1.4. Statistical Analysis of Diffusion Tensor Images of Brain
Participants: Vikash Gupta [Correspondant], Xavier Pennec, Nicholas Ayache.

Diffusion Tensor Imaging of Brain, Tractography, Super-resolution, Statistical analysis

Diffusion tensor imaging (DTI) is gaining interest as a clinical tool for studying a number of brain diseases pertaining to white matter tracts and also as an aid in neuro-surgical planning. Unfortunately, in a clinical environment, the diffusion imaging is hampered by the long acquisiton times, low signal to noise ratio and a prominent partial volume effect due to thick slices. The present work aims at robustifying the analysis of clinical images by developing a super-resolution algorithm for DTI and quantifying its improvements with respect to the existing tensor estimation methods. Part of the work was presented at the 1st International Symposium on Deep Brain Connectomics [68].

5.1.5. 3D/2D coronary arteries registration
Participants: Thomas Benseghir [Correspondant], Grégoire Malandain, Régis Vaillant [GE-Healthcare], Nicholas Ayache.

This work is done in collaboration with GE-Healthcare (Buc).

3D/2D registration ; coronary arteries ; Chronic Total Occlusion ; X-ray fluoroscopy / CT image fusion

The context of this work is to provide the cardiologist with an advanced guidance application, where a pre-operative 3D CT segmented image will be superimposed on the per-operative 2D live fluoroscopy. Since the relative positions of the 3D image and the 2D projective images are unknown, we are currently investigating robust pose estimation methods before using an upcoming registration algorithm.

5.2. Biological Image Analysis

5.2.1. Pre-clinical molecular imaging: breath-hold reconstruction in micro-SPECT and segmentation of IHC stomach slices
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.

SPECT/CT, small animal, respiratory motion, respiratory gating, 4D images, stomach, segmentation, immuno-histochemistry

Using the coupled CT and SPECT device, both the anatomy (with the CT) and physiology information targeted by a dedicated radio-pharmaceutical tracer (here the tumors, with the SPECT) can be imaged. However, tumor quantification is impaired by the respiratory motion that induces an artifical enlargement of the moving structures. Thus, the characterization of respiratory motion in dynamic images was studied.

- An ad hoc method for motion detection in dynamic image was developed and tested on two different modalities (4D-SPECT and 4D-CT).
- Image-based motion detection results were compared to the pressure signal and to lung volume variation. A temporal shift between the peak of motion in images and the ones in the pressure signal was observed (see Figure 3).
- The temporal shift suggested to carefully select data from the non moving phase for a motionless 3D-SPECT image reconstruction. This step was incorporated in a breath-hold like reconstruction method [65], [67], [66].

5.3. Computational Anatomy

5.3.1. Statistical Analysis of Transformations on Lie groups and longitudinal studies
Participants: Xavier Pennec [Correspondant], Marco Lorenzi, Nicolas Duchateau [Hospital Clinic, Univ. Barcelona].
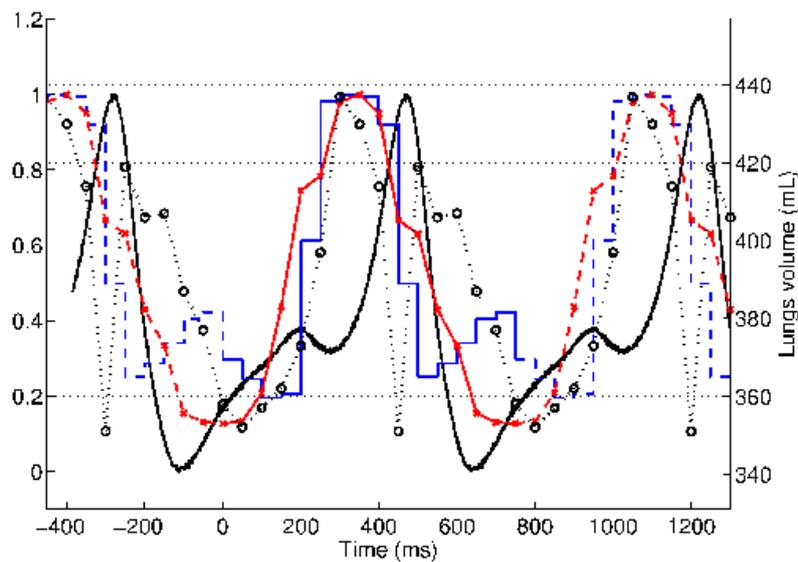
*Figure 3. **Comparison between pressure signal and image-based measures computed on 4D-SPECT and 4D-CT.**
Normalized average pressure signal (continuous black curve); normalised histogram $h_{4D-SPECT}(n)$ computed
on the 4D-SPECT image (blue staircase); normalised histogram $h_{4D-CT}(n)$ of motion phase computed on the
4D-CT image (red curve with cross-shaped markers); and lung volume (mL) measured on the 4D-CT image (black
dotted curve with circle-shaped markers).*

Lie groups, transformations, mean value, non-linear registration

In order to perform statistics on transformations for computational anatomy purposes, we investigate alternative theoretical structures to the right- (or left-) invariant Riemannian setting usually used.

- In order to define a notion of a mean which is consistent with Lie group operations we propose in [57] to replace the Riemannian metric by an affine connection structure on the group. We show that the canonical Cartan connections of a connected Lie group provides group geodesics including one-parameter subgroups which are completely consistent with the composition and inversion. To extend statistical operations to such a non-metric structure, we propose an implicit definition as an exponential barycenter (there is no Fréchet mean like in Riemannian Manifolds) and a linearly convergent iterative fixed point algorithm to reach it. This results into naturally bi-invariant means which are unique when the dispersion of the data is small enough. In some cases including rotations and rigid-body transformations, there is even a global existence and uniqueness theorem which is similar to the Riemannian case.

- In [14], we investigate the canonical Cartan connections and their associated parallel transport for diffeomorphisms, which justifies the use of one-parameter subgroups (the flow of stationary velocity fields or SVF) for diffeomorphic image registration. In particular, we derive closed-forms for different parallel transports and we compare SVF and LDDMM approaches with experiments on longitudinal and inter-subject registration.

- In [33], we analyses with practical expriments what kind of parallel transport is needed to reorient the deformation characteristics along the time sequences of the cardiac motion. Contrarily to the case of the brain, inter-subject transformations to normalize the heart between different subjects are of the same order than deformations along the sequence.

5.3.2. *Statistical Analysis of Longitudinal Transformations in the LDDMM framework*

Participants: Stanley Durrleman [Correspondant], Xavier Pennec, Alain Trouvé [CMLA, ENS Cachan], Nicholas Ayache, José Braga [UMR 5288 CNRS-Université Toulouse Paul Sabatier].

Lie groups, transformations, mean value, non-linear registration

The work initiated the previous years with the PhD of S. Durrleman on the spatio-temporal modeling of shapes was applied with J. Braga to quantify ontogenetic differences between bonobo (Pan paniscus) and chimpanzee (Pan troglodytes) endocrania, using dental development as a timeline. We perform a temporal surface regression that estimates typical endocranial ontogenetic trajectories separately for bonobos and chimpanzees which highlights non-linear patterns of endocranial ontogenetic change and significant differences between species at local anatomical levels rather than considering the endocranium as a uniform entity. The decomposition of the spatio-temporal inter-species difference into a morphological deformation (accounting for size and shape differences independently of age) and a time warp (accounting for changes in the dynamics of development) indicates that juvenile bonobos develop much slower than juvenile chimpanzees, suggesting that inter-specific ontogenetic shifts do not only concern endocranial volume increase, but also the rate of shape changes over time. Our method provides, for the first time, a quantitative estimation of inter-specific ontogenetic shifts that appear to differentiate non-linearly. This work was pusblished in the journal of human evolution [10].

5.3.3. *The Kernel Bundle Framework for Diffeomorphic Image Registration*

Participants: Xavier Pennec [Correspondant], Stefan Sommer [Computer Science Dpt, University of Copenhagen, DK], François Lauze [Computer Science Dpt, University of Copenhagen, DK], Mads Nielsen [Computer Science Dpt, University of Copenhagen, DK].

This work in collaboration with the Computer Science Department of the University of Copenhagen (DK) was initiated during the 6 month visit of S. Sommer at Asclepios in 2010-2011 and was continued remotely since then.

non-rigid registration algorithm, statistics, deformations, shapes, locally affine deformations, sparsity

In order to detect small-scale deformations during longitudinal registration while allowing large-scale deformation needed for inter-subject normalization, we wish to model deformation at multiple scales and represent the deformation at the relevant scales only. We combined in [48], [26] a sparsity prior with the multi-scale Kernel Bundle framework, resulting in an algorithm allowing compact representation of deformation across scales.

In [27], we further extend the framework by introducing higher-order momentum distributions in the LDDMM registration framework. While the zeroth order moments previously used in LDDMM only describe local displacement, the first-order momenta that are proposed here represent a basis that allows local description of affine transformations. Beyond the careful mathematical construction, we show the implications for sparse image registration and we provide examples of how the parametrization enables registration with a very low number of parameters.

5.3.4. *Spectral Correspondances in Non-linear Image Registration*

Participants: Xavier Pennec [Correspondant], Hervé Lombaert, Nicholas Ayache, Leo Grady [SCR, Princeton, US], Farida Cheriet [Saint-Justine Hospital, Montreal, CA].

This work was performed in collaboration with Saint-Justine Hospital in Montreal (CA) and Siemens Corporate Research in Princeton (US).

non-rigid registration algorithm

The demons algorithm was enhanced to include spectral feature correspondences between the images [37]. This feature proves to drastically enhance the robustness of the registration algorithm, which turns out to have a major impact on the construction of atlases. This work was awarded the best paper award at the Medical Computer Vision Workshop [38] and was protected by a patent filling in the US [62]

5.3.5. *Longitudinal Analysis of Brain Atrophy in Alzheimer's Disease*

Participants: Marco Lorenzi [Correspondant], Xavier Pennec, Nicholas Ayache, Giovanni B. Frisoni [IRCCS San Giovanni di Dio Fatebenefratelli, Brescia, Italy].

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

Alzheimer's Disease, non-rigid registration algorithm, longitudinal analysis.

The accurate analysis of the longitudinal structural changes in the brain plays a central role in the study of Alzheimer's disease (AD), for diagnostic purposes and for the assessment of the drugs efficacy in clinical trials. The goal of this project is to provide robust and effective instruments based on non-rigid registration of serial MR images for the modeling and the quantification of the brain atrophy evolution in AD. In 2012, our main scientific developments were the following:

- We developed a framework for the consistent definition of anatomical regions of longitudinal brain atrophy, and for the robust quantification of longitudinal regional percentage volume loss. The framework is based on the analysis of the flux associated to longitudinal deformations (see Figure 4), and was successfully applied to large public dataset of brain images (ADNI - http://adni.loni.ucla.edu/). The work was accepted for oral podium presentation at the MICCAI conference 2012 [40].

- We applied the flux analysis for the quantification of the longitudinal hippocampal and ventricular atrophy in AD. The proposed framework was presented at the NIBAD MICCAI Challenge 2012 [41] , and compared favorably with state-of-art methods in terms of accuracy and stability when applied on the challenge dataset.

- We proposed in [39] a model of the morphological changes in Alzheimer's based on the disentangling of the normal aging component from the pathological atrophy. The model was promoted and presented to the neuroscience community during international scientific conferences [71], [70].

These scientific advances were also included along with the previous ones in the PhD manuscript [1].
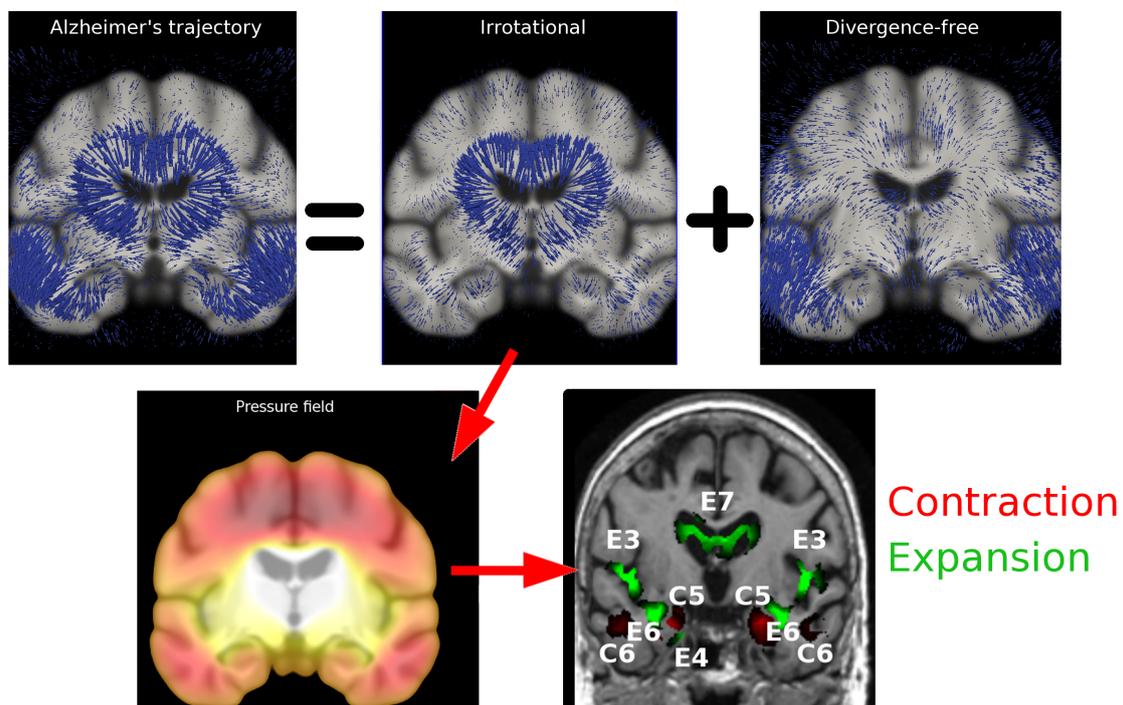
Figure 4. Flux analysis of the longitudinal morphological changes in Alzheimer's disease. The irrotational component associated to the longitudinal deformation encodes the volume loss and gain (top figure). This component is parameterized by a pressure potential (bottom left) which is used to determine the areas of significant flux of the deformation (matter expansion and contraction) in patients affected by Alzheimer's disease (bottom right).

5.3.6. *Statistical Modelling of Cardiac Growth, Deformation and Blood Flow from Medical Images*

Participants: Kristin McLeod [Correspondant], Adityo Prakosa, Christof Seiler, Maxime Sermesant, Xavier Pennec.

Image registration, Demons algorithm, LDDMM, reduced models, CFD, polyaffine, cardiac motion tracking

This work involves developing reduced models of cardiac motion, blood flow and growth.

- Extending the 2011 motion tracking challenge [43], the iLogDemons registration algorithm was applied this year to a data-set of synthetic echocardiography sequences with a training set (provided with ground truth) and testing set to quantitavely compare this algorithms with other cardiac motion tracking algorithms [45].

- A reduced order model of cardiac motion based on a polyaffine log-demons registration was developed to represent the motion along the cardiac cycle with a smaller number of parameters compared to previously proposed methods. The method was applied to a data-set of 10 volunteers and the results were presented at the 2012 STACOM workshop at the MICCAI conference [44].

- The analysis of a statistical model for reduced blood flow simulations in the pulmonary artery proposed in the 2010 STACOM workshop is currently being extended to a journal version with an improved method and a larger data-set.

- The statistical modeling of the right ventricle growth in a population of Tetralogy of Fallot patients was extended to a full bi-ventricular growth model on different data [55]. Results confirm the previous findings which were shown to be useful in providing insights for patient treatment [13] (see Figure 5).

5.3.7. *Trees on Geometrical Deformations to Model the Statistical Variability of Organs in Medical Images*

Participants: Christof Seiler [Correspondant], Xavier Pennec, Mauricio Reyes [Institute for Surgical Technology and Biomechanics, University of Bern, Switzerland].

Parametrization of diffeomorphisms, Shape statistics, Multiscale and hierarchical trees, Log-Euclidean polyaffine transformations, Polyaffine registration, Log-Demons registration, Generative statistical model, Bayesian registration, Mandibles, Femurs

Intersubject anatomical deformations between patients can be found on coarse and fine scales. Each level of granularity has specific regions of interest in clinical applications. The challenge is to connect geometrical deformations to clinical regions across scales.

- We presented this connection by introducing structured diffeomorphic registration [24]. At the core of our method is the parametrization of geometrical deformations with trees of locally affine transformations describing intersubject variability across scales (see Figure 6).

- The methodology of [24] was successfully applied to mandible implant design [31] and in a clinical journal paper on allograft selection [21].

- We statistically modeled the deformation parameters in a population by formulating a generative statistical model [47]. This model allowed us to incorporate deformation statistics as a prior in a Bayesian setting and it enabled us to extend the classical sequential coarse to fine registration to a simultaneous optimization of all scales.

- We explored cell shape statistics to classify stem cells [23].

- We investigated the benefits of considering patient metadata and morphometric measures to enhance bone surface shape prediction [6].
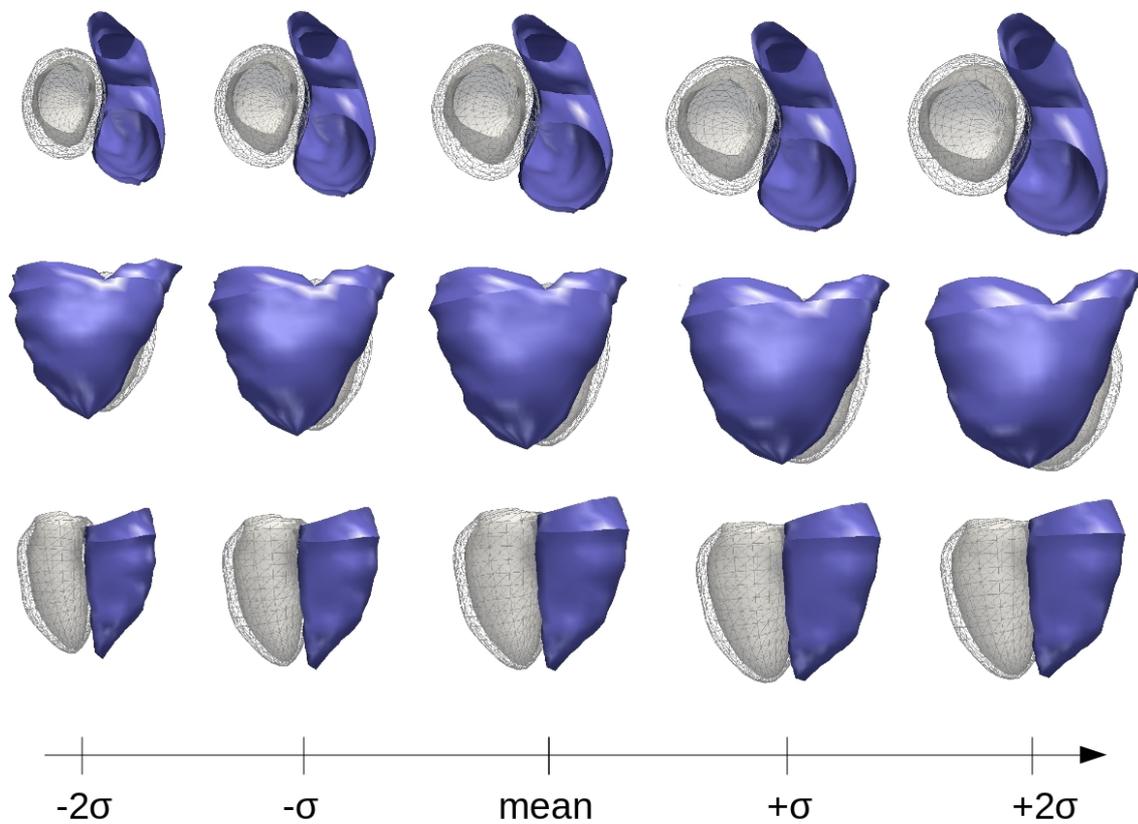
Figure 5. Mean growth model computed from a population of 13 repaired Tetralogy of Fallot patients. Both ventricles grow as body surface area (BSA) increases.
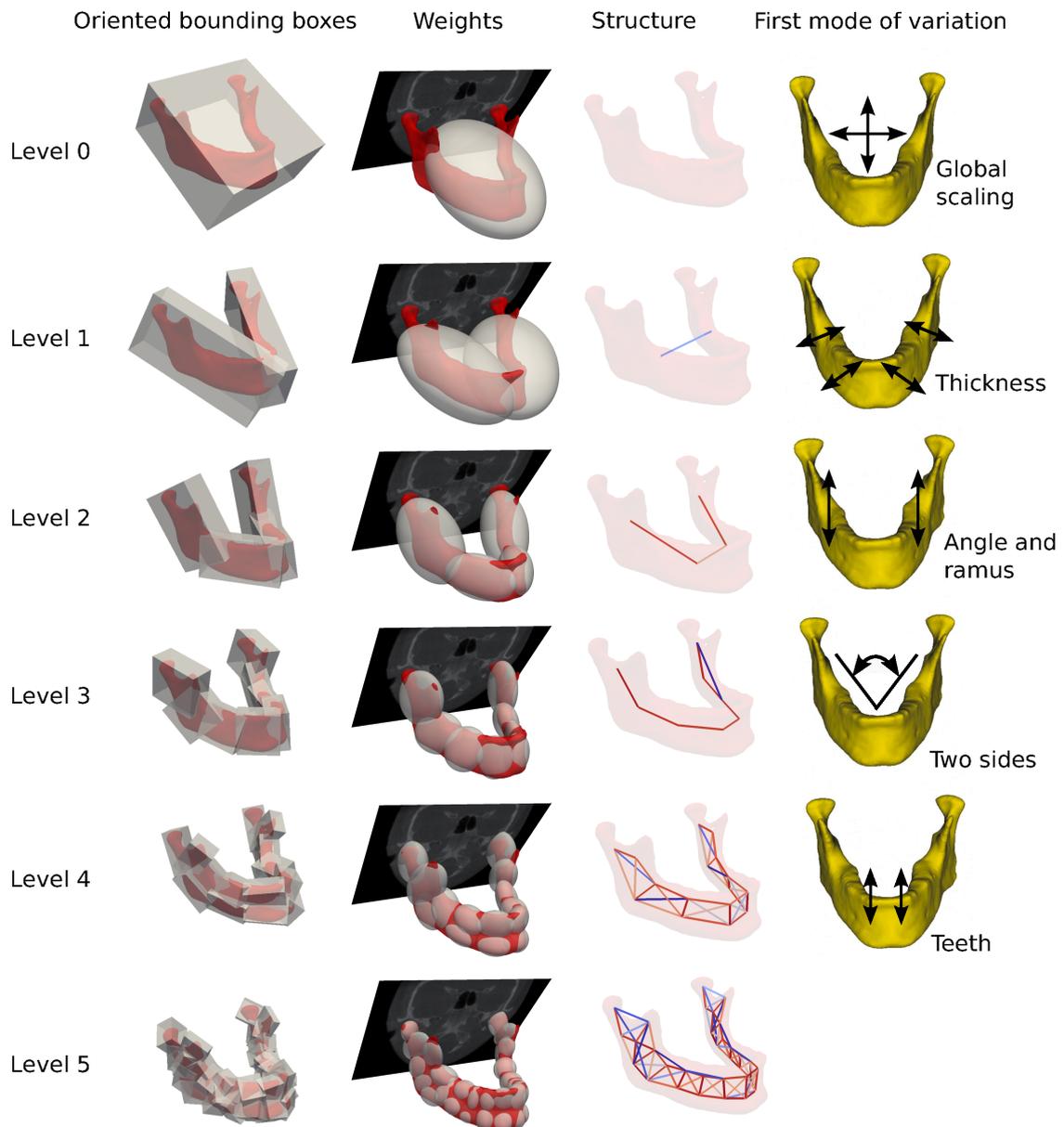
Figure 6. The red mandible is the surface extracted from the template image. Column 1: Oriented bounding boxes computed using the algorithm presented in [24]. Column 2: Ellipsoids representing Gaussian weights derived from the oriented bounding boxes. Column 3: Structure given by the weights with correlations between regions ranging from low (blue=0:4) to high (red=1). Column 4: First PCA mode at each level showing the residual variation.

5.3.8. *Evaluation of iLogDemons Algorithm for Cardiac Motion Tracking in Synthetic Ultrasound Sequence*

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

synthetic echocardiography, iLogDemons, cardiac motion tracking

- The LogDemons and iLogDemons non-linear registration algorithms were evaluated on a dataset of synthetic cardiac ultrasound sequences [32], [45]. With these synthetic sequences, it is possible to quantify the performance of these registration algorithms since the ground truth motion was given. Therefore the LogDemons/ iLogDemons can be evaluated objectively (see Figure 7).
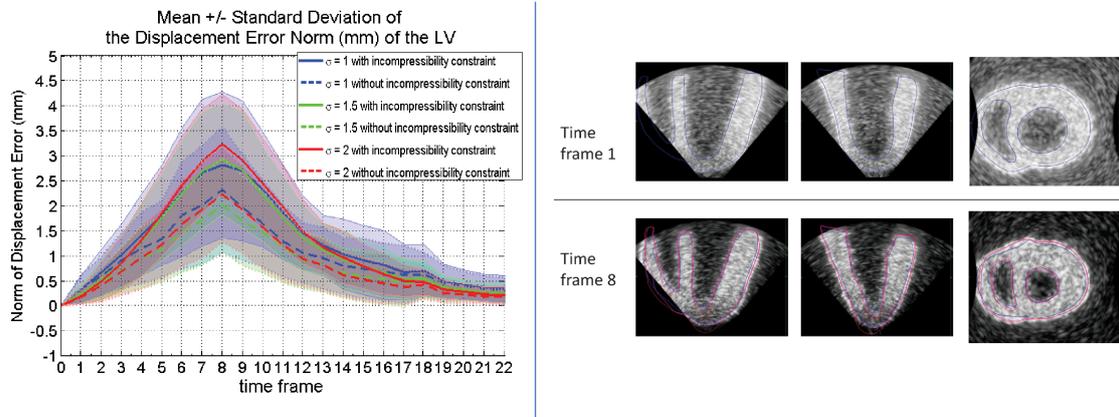
Figure 7. **Registration Error and Motion Tracking Result** *The error quantification for different registration parameter (left) and the myocardium tracking result during the maximum contraction (right): ground truth, LogDemons and iLogDemons estimation are shown in blue, purple and red respectively.*

5.3.9. *Simulation of Atrophy in Alzheimer's disease*

Participants: Arnaud Le Carvennec [Correspondant], Sebastien Ourselin [UCL], Nick Fox [UCL], Xavier Pennec [Inria], Nicholas Ayache [Inria].

Simulation, Alzheimer's disease, registration

- Evaluation of registration algorithm using multi-channel images.
- Simulation of atrophy based on registration.

5.4. Computational Physiology

5.4.1. *Tumor Growth Modeling*

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

DTI, MRI, simulation, clinical, tumor, brain, glioma

We aim at developing image analysis methods and biophysical models in order to guide the planning of therapies (surgical removal and radiotherapy) for brain cancer (glioma) patients. Our work is focused on those objectives :

- Predicting the location of glioma recurrence after a resection surgery [49].

- Determining the best description of tumor cell diffusion tensor in white matter (patient-based, atlas-based or isotropic) which leads to the most accurate results for predicting future tumor growth.

- Comparing tumor growth speeds on 3 patient cases. This is a work in progress and the objective is to reach 30 patients when the work is complete.

5.4.2. Generation of Synthetic but Visually Realistic Time Series of Cardiac Images Combining a Biophysical Model and Clinical Images

Participants: Adityo Prakosa [Correspondant], Maxime Sermesant, Hervé Delingette, Stéphanie Marchesseau, Eric Saloux [CHU Caen], Pascal Allain [Philips Healthcare], Nicolas Villain [Philips Healthcare], Nicholas Ayache.

synthetic 4D cardiac sequences, cardiac electromechanical model, non-rigid registration

- A pipeline to create visually realistic synthetic 4D cardiac sequence using the cardiac motion simulated by an electromechanical model is developed. This pipeline combines the simulated myocardium displacement field with the estimated myocardium displacement field from a registration method. This combined displacement field is then used to warp the original images in order to create the synthetic cardiac sequence.

- In [19],we proposed a new approach based on Stationary Velocity Fields to combine the two motions (see Figure 8). We also proposed a new method that diffuses displacement fields in order to maintain the continuity between the simulation and the real image with minimal texture distortion. Thanks to the detailed interplay between image processing and biophysical modeling, we can fully use a complete sequence in order to generate several new ones. This method also gives better realism compared to traditional methods based on the deformation of an end-diastolic image, since the generated synthetic sequence will also contain the motion of surrounding tissues such as the motion of the mitral valve.

- The new synthetic images are similar to the original ones except for the motion of the heart which is modified to follow the motion provided by a biophysical model. The parameters of the biophysical model can be modified to create variations around this motion. This pipeline has been applied to generate different synthetic sequences from different imaging modalities. It is generic and can be used with a different biophysical model or a different image registration algorithm, and it can be extended to other organs.

- As these synthetic 4D cardiac sequences have kinematic ground truth information, those sequences represent in themselves a valuable resource to benchmark motion tracking methods or to train machine-learning algorithm.

5.4.3. Real-Time Cardiac Electrophysiology Computing for Training Simulator

Participants: Hugo Talbot [Correspondant], Hervé Delingette, Stephane Cotin, Maxime Sermesant, Christian Duriez.
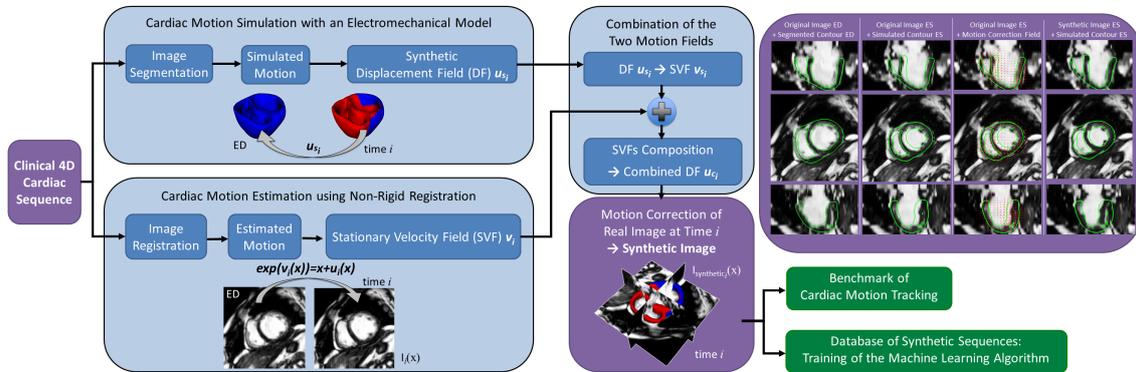
Figure 8. **Synthetic 4D Cardiac Sequence Generation Pipeline** *A clinical 4D sequence is used as an input to create a synthetic 4D sequence in which the myocardium motion follows a prescribed simulated displacement field. The combined simulated and registration motion are used to correct the motion of the real clinical images in order to create the synthetic cardiac sequence.*

Cardiac electrophysiology simulation, real-time, GPU computing, patient-specific study

Cardiac arrhythmia is a very frequent pathology related to an abnormal electrical activity in the myocardium. This work aims at developing a training simulator for interventional radiology and thermo-ablation of these arrhythmias.

- The latest improvements lead on electrophysiology simulation (see Figure 9) using GPU computing allowed us to reach real-time performance[51]. The issue of fast electrophysiology was a major bottleneck in the development of our simulator.
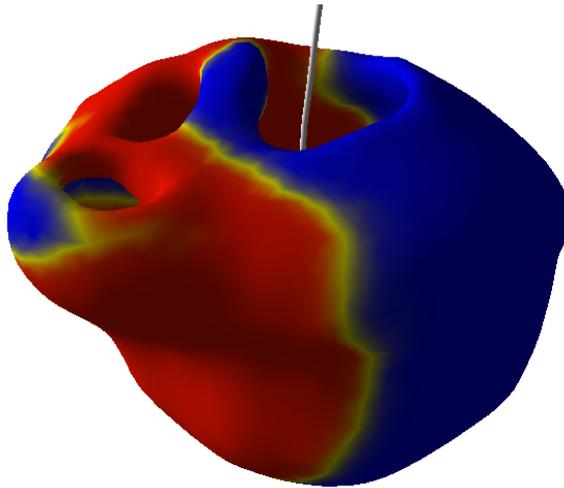
Figure 9. Cardiac electrophysiology computed on a patient-specific geometry

Coupling between the cardiac electrophysiology model with cardiac mechanical models has been achieved, thus leading to an interactive framework. Moreover, the electrophysiology simulation has been also coupled with a navigation simulation.

- In the context of his work on cardiac electrophysiology, we initiated two different collaborations. Joint work has been performed with the team CARMEN from Inria Bordeaux on bidomain modeling for cardiac electrophysiology. This exchange targeted the implementation in SOFA of these models. Secondly, a collaboration with the MACS team in Saclay has been initiated to personalize the cardiac electrophysiology model based on the Verdandi library.

5.4.4. *Personalized model of the heart for cardiac therapy planning*

**Participants:** Stéphanie Marchesseau [Correspondant], Hervé Delingette, Nicholas Ayache, Maxime Sermesant.

An award has been won for this work at the MICCAI 2012 Conference. It was partially funded by the European Community's euHeart project under grant agreement 224495 and by the ERC advanced Grant MedYMA 291080.

Cardiac simulation, sensitivity analysis, calibration algorithm, specificity study

- We implemented the full Bestel-Clement-Sorine electromechanical model of the heart in SOFA [54], [51].

- We ran a complete sensitivity analysis to check its behaviour for healthy and pathological cases [15].

- A new calibration algorithm was proposed [15] in order to initialize global mechanical parameters from the volume and pressure curves, before further personalization (see Figure 10).

- The application of this new method on 6 healthy and 2 pathological cases allowed to draw preliminary conclusions on specific parameters to a given pathology [42], [16].

- The model has also been used to create synthetic images in [19] and for the data of the STACOM 2012 challenge [32].

5.4.5. *Image-based glioma modeling for radiotherapy planning*

Participants: Bjoern Menze [Correspondant], Ender Konukoglu [MSR Cambridge], Jan Unkelbach [Harvard MGH].

- Implemented the generative tumor segmentation model together with the E. Konukoglu's tumor infiltration model for evaluation at the MGH Department of Radiation Oncology.

- Integrated tumor infiltration model with radiation therapy model.

5.4.6. *Cardiac Arrhythmia Radio-frequency Ablation Planning*

Participants: Rocio Cabrera Lozoya [Correspondant], Maxime Sermesant, Hervé Delingette, Nicholas Ayache.

This work is performed in the context of the the PhD of Rocio Cabrera Lozoya in collaboration with the IHU LIRYC Bordeaux and is funded by ERC MedYMA.

- Biophysical model development for the prediction of radio frequency ablation sites for ventricular tachycardias.

- Target site map generation for ablation therapy guidance

- Structural and functional characterization of target sites using 3D imaging and EP measurements through machine learning algorithms (see Figure 11).

- Prediction validation with acquired clinical data
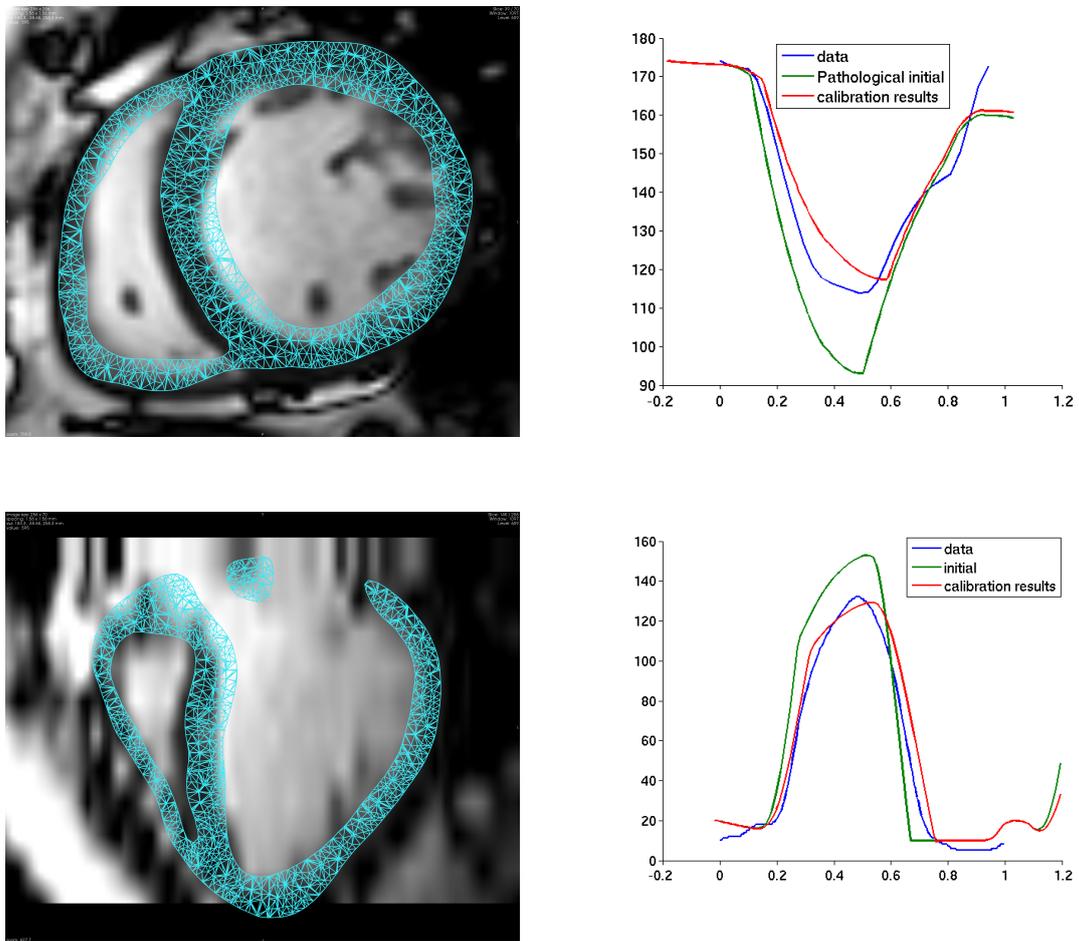
Figure 10. Results of the calibration algorithm fro one pathological case. (Left) Simulated mesh after calibration compared to the images (at end-diastole). (Right) Resulting volume and pressure curves.
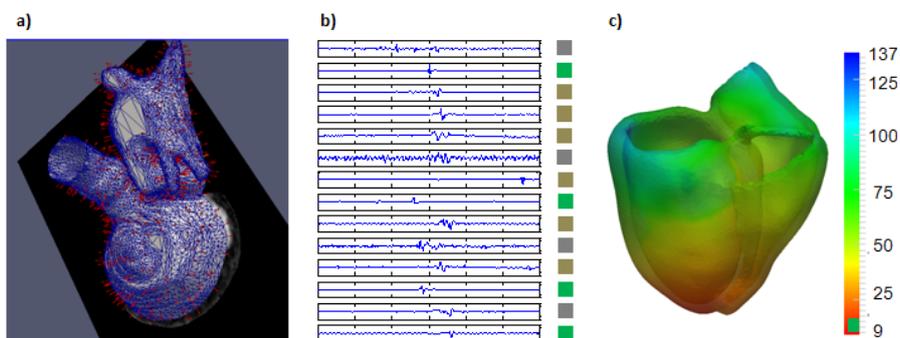
Figure 11. a) MRI Feature Extraction b) EP Signal Feature Extraction c) EP Model Personalization

5.4.7. Computational modeling of radiofrequency ablation for the planning and guidance of abdominal tumor treatment

Participants: Chloe Audigier [Correspondant], Herve Delingette, Tommaso Mansi [Siemens Corporate Research], Nicholas Ayache.

This PhD is carried out between Asclepios research group, Inria Sophia Antipolis, France and the Image Analytics and Informatics global field, Siemens Corporate Research, Princeton, USA.

Therapy planning, radio-frequecy ablation, Liver

The objective of this work is to develop a computational framework for patient-specific planning of radiofrequency ablation:

- A patient-specific detailed anatomical model of the liver is estimated from standard CT image and meshed to generate a tetrahedral volume mesh.
- A porous media model is used to compute the patient-specific blood flow in the hepatic circulatory system.
- Bio-heat equations have been implement in SOFA to model the heat propagation in biological tissues.
- A cell death model is included to account for the cellular necrosis.
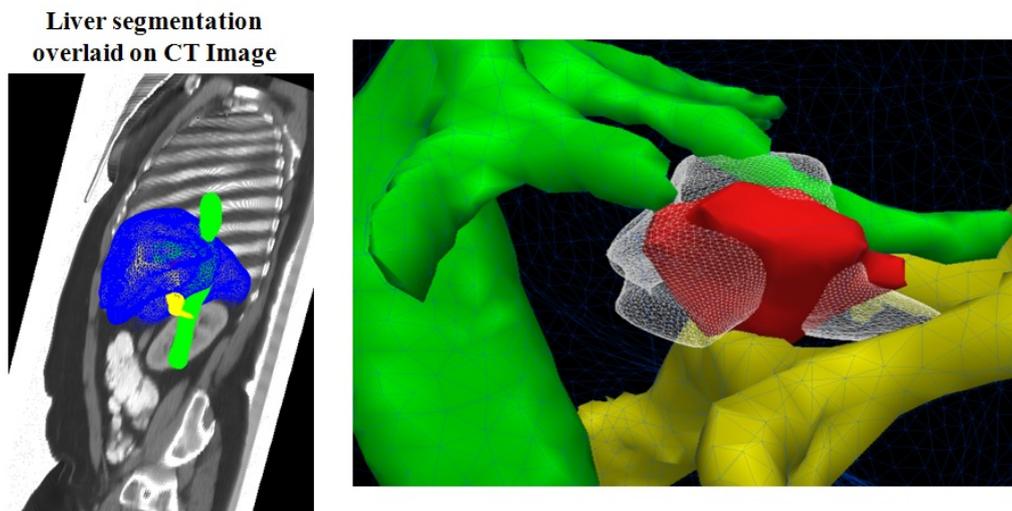
Figure 12. (Left) Anatomical model of the liver estimated from standard clinical CT image. (Right) The predicted necrosis computed with our model compares qualitatively well with the necrosis region observed on a post-operative MRI scan.

5.4.8. Tumor Growth Simulation for the creation of a database of virtual patients

Participants: Nicolas Cordier [Correspondant], Nicholas Ayache, Hervé Delingette, Bjoern Menze, Ezequiel Geremia.

This work was funded by the European Research Council through the ERC Advanced Grant MedYMA (on Biophysical Modeling and Analysis of Dynamic Medical Images).

Brain MRI, Tumor simulation.

- Synthesizing multi-channel MR images with healthy and glial tumors.
- Creating a database of synthetic images for training and validating of brain tumor segmentation algorithms.

5.4.9. Learning approach for the Mechanical personalization of cardiac models

Participants: Loic Le Folgoc [Correspondant], Hervé Delingette, Antonio Criminisi, Nicholas Ayache.

Inverse problem, machine learning, patient-specific, current, kinematics

- A machine-learning framework for the mechanical personalization of the Bestel-Clement-Sorine model of the heart from patient-specific kinematics
- The computational burden is moved to an offline stage, where the inter-subject variability in motion is captured via the statistical analysis of training samples
- Towards a probabilistic framework for the personalization and therapy planning problems, to better account for significant and diverse uncertainty sources
- Published at the MICCAI 2012 Workshop on Medical Computer Vision[36]
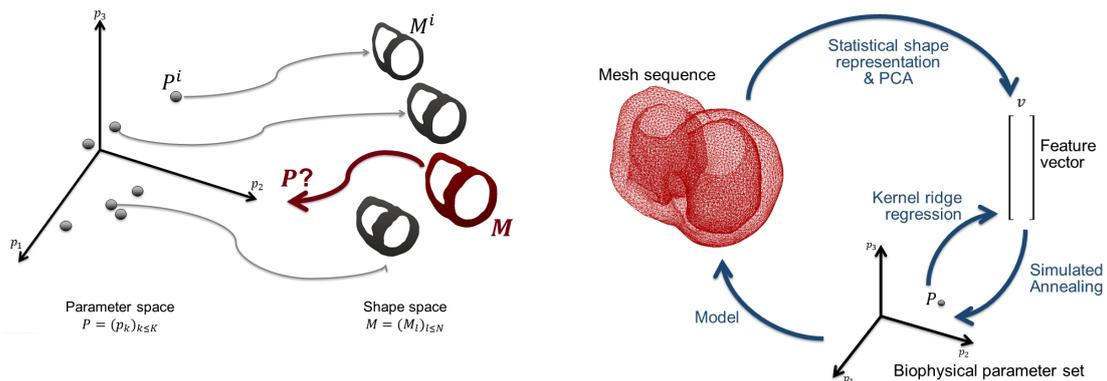
Figure 13.

5.4.10. Brain Tumor Growth Modeling

Participants: Matthieu Lê [Correspondant], Nicholas Ayache, Hervé Delingette.

Gliomas simulations, reaction-diffusion, brain tumors

- In collaboration with the MC2 research team in Bordeaux, we developed a tumor growth model based on different types of cell : necrotic, proliferative and quiescent cells (see Figure 14). It is also based on the underlying vascularization of the brain.
- We studied the impact of the vascularization angiogenesis factor and degradation factor.

5.4.11. Modeling of atrophy of the brain in Alzheimer's Disease

Participants: Bishesh Khanal [Correspondant], Xavier Pennec, Nicholas Ayache.
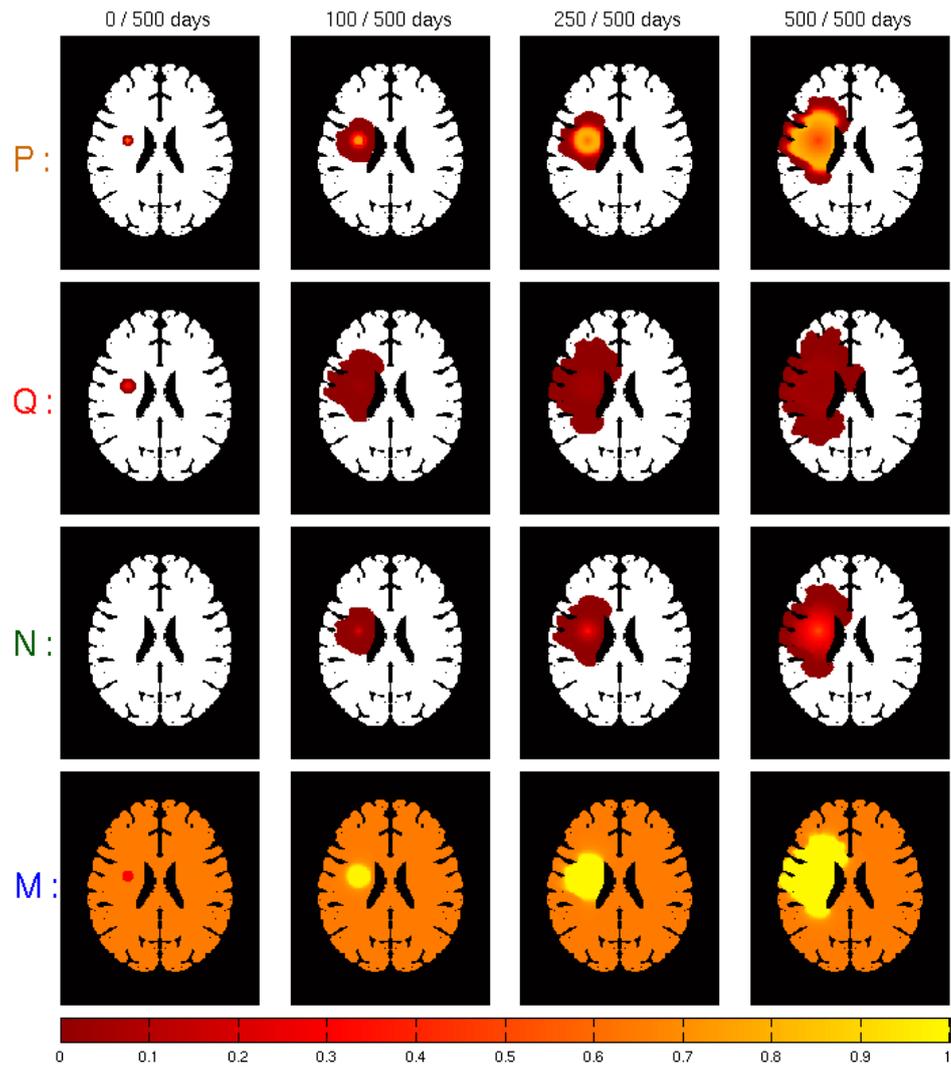
Figure 14. Results of a glioma simulation at day 0, 50, 250 and 500. The proliferative cells are on the first row, the quiescent cells on the second row, the necrotic cells on the third one and the vascularization is on the fourth row.

Alzheimer's Disease (AD), modeling atrophy, biomechanical model

- The idea is to have a model which produces deformation of the brain when a known distribution of local volume change (atrophy) is prescribed to the model. The study is to understand how brain deformation evolve in time with respect to temporal and spatial variation of atrophy.

- During the masters internship period a simple model was tested in 2D square and 3D cube where high atrophy regions acted as sinks for the displacement field [69] (see Figure 15).
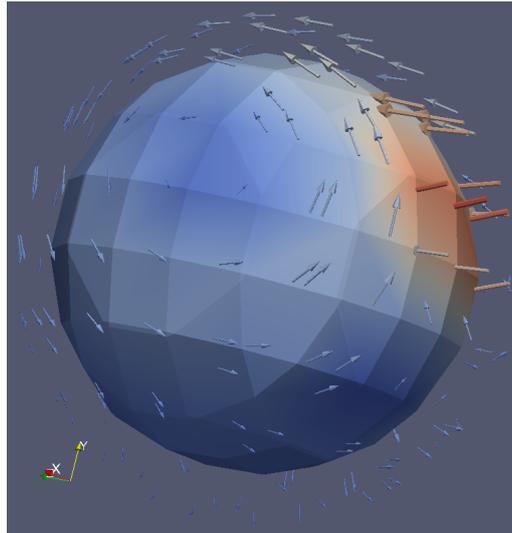
Figure 15. Displacement field in 3D when high atrophy is prescribed in the center of a cube.

ATHENA Project-Team

6. New Results

6.1. Computational Diffusion MRI

6.1.1. Improving dMRI Signal and Acquisitions

6.1.1.1. Diffusion MRI Signal Reconstruction with Continuity Constraint and Optimal Regularization
Participants: Emmanuel Caruyer, Rachid Deriche.

In diffusion MRI, the reconstruction of the full Ensemble Average Propagator (EAP) provides new insights in the diffusion process and the underlying microstructure. The reconstruction of the signal in the whole Q-space is still extremely challenging however. It requires very long acquisition protocols, and robust reconstruction to cope with the very low SNR at large b-values. Several reconstruction methods were proposed recently, among which the Spherical Polar Fourier (SPF) expansion, a promising basis for signal reconstruction. Yet the reconstruction in SPF is still subject to noise and discontinuity of the reconstruction. In this work, we present a method for the reconstruction of the diffusion attenuation in the whole Q-space, with a special focus on continuity and optimal regularization. We derive a modified Spherical Polar Fourier (mSPF) basis, orthonormal and compatible with SPF, for the reconstruction of a signal with continuity constraint. We also derive the expression of a Laplace regularization operator in the basis, together with a method based on generalized cross validation for the optimal choice of the parameter. Our method results in a noticeable dimension reduction as compared with SPF. Tested on synthetic and real data, the reconstruction with this method is more robust to noise and better preserves fiber directions and crossings.

This work has been published in [13]

6.1.1.2. A Computational Framework for Experimental Design in Diffusion MRI
Participants: Emmanuel Caruyer, Rachid Deriche.

In this work, we develop a computational framework for optimal design of experiment in parametric signal reconstruction. We apply this to the optimal design of one dimensional Q-space, Q-ball imaging and multiple Q-shell experimental design. We present how to construct sampling scheme leading to minimal condition number, and compare to state-of-the-art sampling methods. We show in particular a better noise performance of these scheme through Monte-Carlo simulations for the reconstruction of synthetic signal. This demonstrates the impact of this computational framework on acquisition in diffusion MRI.

This work has been published in [16]

6.1.1.3. Parametric Dictionary Learning in Diffusion MRI
Participants: Sylvain Merlet, Emmanuel Caruyer, Aurobrata Ghosh, Rachid Deriche.

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

In this work, we propose an approach to exploit the ability of compressive sensing to recover diffusion MRI signal and its characteristics from a limited number of samples. Our approach is threefold. First, we learn and design a parametric dictionary from a set of training diffusion data. This provides a highly sparse representation of the diffusion signal. The use of a parametric method presents several advantages: we design a continuous representation of the signal, from which we can analytically recover some features such as the ODF; besides, the dictionary we train is acquisition-independent. Next, we use this sparse representation to reconstruct the signal of interest, using cross-validation to assess the optimal regularization parameter for each signal reconstruction. The use of cross-validation is critical in the L1 minimization problem, as the choice of the parameter is sensitive to the noise level, the number of samples, and the data sparsity. Third, we use a polynomial approach to accurately extract ODF maxima. Finally, we motivate and describe the choice of experimental parameters for the HARDI contest.

This work has been published in [26].

6.1.1.4. Diffusion and Multiple Orientations from 1.5 MR Systems with Limited Gradient Tables
Participants: Sylvain Merlet, Rachid Deriche, Kevin Whittingstall [Radiology department, Université de Sherbrooke, Québec, Canada], Maxime Descoteaux [Sherbrooke Connectivity Imaging Laboratory, Computer Science Departement, Université de Sherbrooke, Québec, Canada].

This work has been performed within the framework of the Brain Connectivities Associate Team.

Diffusion MRI (dMRI) enables the quantification of water diffusion, influenced by the structure of biological tissues, from the acquisition of diffusion weighted magnetic resonance images (DW-MRI). While recent advances enable to recover complex fiber geometries using diffusion measurements along various sampling schemes of high order, some older MR systems work with limited gradient tables (ex: maximum of 6 or 12 directions). These systems are designed for Diffusion Tensor Imaging (DTI). Several hospitals and research institutes in the world are limited by these fixed DTI gradient sets. Therefore, groups that want to perform state-of-the-art tractography using high angular resolution diffusion imaging (HARDI) data are penalized and can only perform DTI tractography on their old system. The Gaussian assumption of the tensor model, in DTI, is an over simplification of the diffusion phenomenon of water molecules in the brain and thus cannot resolve crossing fibers. In this work, we show that new diffusion signal modeling and processing techniques enable to capture complex angular structure of the diffusion process even from a reduced gradient direction set arising from an older MR system.

This work has been published in [27].

6.1.1.5. A Robust variational approach for simultaneous smoothing and estimation of DTI
Participants: Rachid Deriche, Meizhu Liu [Department of CISE, University of Florida, Gainesville, USA], Baba C. Vemuri [Department of CISE, University of Florida, Gainesville, USA].

Estimating diffusion tensors is an essential step in many applications — such as diffusion tensor image (DTI) registration, segmentation and fiber tractography. Most of the methods proposed in the literature for this task are not simultaneously statistically robust and feature preserving techniques. In this work, we propose a novel and robust variational framework for simultaneous smoothing and estimation of diffusion tensors from diffusion MRI. Our variational principle makes use of a recently introduced total Kullback–Leibler (tKL) divergence for DTI regularization. tKL is a statistically robust dissimilarity measure for diffusion tensors, and regularization by using tKL ensures the symmetric positive definiteness of tensors automatically. Further, the regularization is weighted by a non-local factor adapted from the conventional non-local means filters. Finally, for the data fidelity, we use the nonlinear least-squares term derived from the Stejskal–Tanner model. We present experimental results depicting the positive performance of our method in comparison to competing methods on synthetic and real data examples.

This work has been accepted for publication in NeuroImage [63].

6.1.2. Modeling in Diffusion MRI

6.1.2.1. Fast and Analytical EAP Approximation from a 4th Order Tensor
Participants: Aurobrata Ghosh, Rachid Deriche.

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

Generalized Diffusion Tensor Imaging (GDTI) was developed to model complex Apparent Diffusivity Coefficient (ADC) using Higher Order Tensors (HOT) and to overcome the inherent single-peak shortcoming of DTI. However, the geometry of a complex ADC profile doesn't correspond to the underlying structure of fibers. This tissue geometry can be inferred from the shape of the Ensemble Average Propagator (EAP). Though interesting methods for estimating a positive ADC using 4th order diffusion tensors were developed, GDTI in general was overtaken by other approaches, e.g. the Orientation Distribution Function (ODF), since it is considerably difficult to recuperate the EAP from a HOT model of the ADC in GDTI. In this work, we present a novel closed-form approximation of the EAP using Hermite polynomials from a modified HOT model of the original GDTI-ADC. Since the solution is analytical, it is fast, differentiable, and the approximation converges

well to the true EAP. This method also makes the effort of computing a positive ADC worthwhile, since now both the ADC and the EAP can be used and have closed forms. We demonstrate our approach with 4th order tensors on synthetic data and in vivo human data.

This work has been accepted for publication in the International Journal of Biomedical Imaging [54].

6.1.2.2. A Polynomial Approach for Extracting the Extrema of a Spherical Function and its Application in Diffusion MRI
Participants: Aurobrata Ghosh, Elias Tsigaridas [PolSys Project-Team, Inria, Paris Rocquencourt, France], Bernard Mourrain [Galaad Project-Team, Inria, Sophia Antipolis, Méditerranée, France], Rachid Deriche.

Antipodally symmetric spherical functions play a pivotal role in diffusion MRI in representing sub-voxel-resolution microstructural information of the underlying tissue. This information is described by the geometry of the spherical function. In this work, we propose a method to automatically compute all the extrema of a spherical function. We then classify the extrema as maxima, minima and saddle-points to identify the maxima. We take advantage of the fact that a spherical function can be described equivalently in the spherical harmonic (SH) basis, in the symmetric tensor (ST) basis constrained to the sphere, and in the homogeneous polynomial (HP) basis constrained to the sphere. We extract the extrema of the spherical function by computing the stationary points of its constrained HP representation. Instead of using traditional optimization approaches, which are inherently local and require exhaustive search or re-initializations to locate multiple extrema, we use a novel polynomial system solver which analytically brackets all the extrema and refines them numerically, thus missing none and achieving high precision. To illustrate our approach we consider the Orientation Distribution Function (ODF). In diffusion MRI, the ODF is a spherical function which represents a state-of-the-art reconstruction algorithm whose maxima are aligned with the dominant fiber bundles. It is, therefore, vital to correctly compute these maxima to detect the fiber bundle directions. To demonstrate the po- tential of the proposed polynomial approach we compute the extrema of the ODF to extract all its maxima. This polynomial approach is, however, not dependent on the ODF and the framework presented in this work can be applied to any spherical function described in either the SH basis, ST basis or the HP basis.

This work has been submitted to Medical Image Analysis and has been accepted for a publication to appear early 2013 [57].

6.1.2.3. 4th Order Symmetric Tensors and Positive ADC Modelling
Participants: Aurobrata Ghosh, Rachid Deriche.

High Order Cartesian Tensors (HOTs) were introduced in Generalized DTI (GDTI) to overcome the limitations of DTI. HOTs can model the apparent diffusion coefficient (ADC) with greater accuracy than DTI in regions with fiber heterogeneity. Although GDTI HOTs were designed to model positive diffusion, the straightforward least square (LS) estimation of HOTs doesn't guarantee positivity. In this work, we address the problem of estimating 4th order tensors with positive diffusion profiles. Two known methods exist that broach this problem, namely a Riemannian approach based on the algebra of 4th order tensors, and a polynomial approach based on Hilbert's theorem on non-negative ternary quartics. In this work, we review the technicalities of these two approaches, compare them theoretically to show their pros and cons, and compare them against the Euclidean LS estimation on synthetic, phantom and real data to motivate the relevance of the positive diffusion profile constraint.

This work is under submission.

6.1.2.4. Higher-Order Tensors in Diffusion Imaging: A Survey
Participants: Thomas Schultz [MPI for Intelligent Systems, Tubingen, Germany], Andrea Fuster [Eindhoven University of Technology, The Netherlands], Aurobrata Ghosh, Luc Florack [Eindhoven University of Technology, The Netherlands], Rachid Deriche, Lek-Heng Lim [University of Chicago, USA].

Diffusion imaging is a noninvasive tool for probing the microstructure of fibrous nerve and muscle tissue. Higher-order tensors provide a powerful mathematical language to model and analyze the large and complex data that is generated by its modern variants such as High Angular Resolution Diffusion Imaging (HARDI) or Diffusional Kurtosis Imaging. This survey gives a careful introduction to the foundations of higher-order tensor algebra, and explains how some concepts from linear algebra generalize to the higher-order case. From the application side, it reviews a variety of distinct higher-order tensor models that arise in the context of diffusion imaging, such as higher-order diffusion tensors, q-ball or fiber Orientation Distribution Functions (ODFs), and fourth-order covariance and kurtosis tensors. By bridging the gap between mathematical foundations and application, it provides an introduction that is suitable for practitioners and applied mathematicians alike, and propels the field by stimulating further exchange between the two.

This work has been submitted and is under review.

6.1.2.5. *Nonnegative Definite EAP and ODF Estimation via a Unified Multi-Shell HARDI Reconstruction*
Participants: Rachid Deriche, Jian Cheng [ATHENA and LIAMA, China], Tianzi Jiang [LIAMA, China].

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

In High Angular Resolution Diffusion Imaging (HARDI), Orientation Distribution Function (ODF) and Ensemble Average Propagator (EAP) are two important Probability Density Functions (PDFs) which reflect the water diffusion and fiber orientations. Spherical Polar Fourier Imaging (SPFI) is a recent model-free multi-shell HARDI method which estimates both EAP and ODF from the diffusion signals with multiple b values. As physical PDFs, ODFs and EAPs are nonnegative definite respectively in their domains $S2$ and $R3$. However, existing ODF / EAP estimation methods like SPFI seldom consider this natural constraint. Although some works considered the nonnegative constraint on the given discrete samples of ODF / EAP, the estimated ODF/EAP is not guaranteed to be nonnegative definite in the whole continuous domain. The Riemannian framework for ODFs and EAPs has been proposed via the square root parameterization based on pre-estimated ODFs and EAPs by other methods like SPFI. However, there is no work on how to estimate the square root of ODF/EAP called as the wavefuntion directly from diffusion signals. In this work, based on the Riemannian framework for ODFs / EAPs and Spherical Polar Fourier (SPF) basis representation, we propose a unified model-free multi-shell HARDI method, named as Square Root Parameterized Estimation (SRPE), to simultaneously estimate both the wavefunction of EAPs and the nonnegative definite ODFs and EAPs from diffusion signals. The experiments on synthetic data and real data showed SRPE is more robust to noise and has better EAP reconstruction than SPFI, especially for EAP profiles at large radius.

This work has been published in [11] and [18].

6.1.2.6. *An Intrinsic Diffeomorphism Invariant Riemannian Framework for Probability Density Function Computing in diffusion MRI*
Participants: Rachid Deriche, Jian Cheng [ATHENA and LIAMA, China], Aurobrata Ghosh, Tianzi Jiang [LIAMA, China].

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

In High Angular Resolution Imaging (HARDI),Ensemble Average Propagator (EAP) and Orientation Distribution Function (ODF) are two important Probability Density Functions (PDFs), which describe the diffusion probability respectively in 3D space and along directions. Fisher information metric has been successfully applied in Diffusion Tensor Imaging (DTI) on tensor estimation, filtering, registration, statistical analysis, etc. However, to our knowledge, existing works in HARDI mainly focus on ODF/EAP estimation, not on ODF and EAP data processing. In this work, we propose a general state-of-the-art Riemannian framework as a mathematical tool to process such PDF data, by representing the square root of the PDF, called *wavefunction* based on quantum mechanics, as a linear combination of some orthonormal basis functions. The proposed Riemannian framework is showed to be a natural extension of previous Riemannian framework for tensors. We deduced the Riemannian metric for the PDF family via orthonormal basis representation, and proved the statistical manifold to be a convex subset of a high dimensional sphere. In this framework, the exponential map, logarithmic map and geodesic have closed forms, and the weighted Riemannian mean and median uniquely

exist. Moreover, we generalized the Log-Euclidean framework and the Geodesic Anisotropy (GA) form tensors to ODFs/EAPs. The theoretical results can be applied to any general PDF data under any orthonormal basis representation. Furthermore we analyzed theoretically the similarities and differences between the Riemannian frameworks for EAPs, ODFs and for tensors, and demonstrated the proposed Riemannian metric is diffeomorphism invariant, which is the natural extension of the previous affine-invariant metric for tensors. Some potential applications were proposed via the Riemannian operations on the ODF/EAP field, such as anisotropy description via GA, nonnegative definite ODF/EAP estimation, interpolation, filtering, Principal Geodesic Analysis (PGA) and atlas estimation. The Riemannian framework and its applications were validated in synthetic, phantom and real data. The experiments demonstrated that the Riemannian framework is very useful for ODF/EAP computing, although the results from Riemannian metric and Euclidean metric are similar for ODFs but much different for EAPs.

This work has been published in [11]. A longer version has been submitted and is under revision for the journal IEEE transaction on Medical Imaging.

6.1.2.7. Ensemble Average Propagator Reconstruction via Compressed Sensing: Discrete or Continuous Bases ?
Participants: Sylvain Merlet, Michael Paquette [Sherbrooke Connectivity Imaging Laboratory, Computer Science Departement, Université de Sherbrooke, Québec, Canada], Rachid Deriche, Maxime Descoteaux [Sherbrooke Connectivity Imaging Laboratory, Computer Science Departement, Université de Sherbrooke, Québec, Canada].

This work has been partly supported within the framework of the Brain Connectivities Associate Team.

In this work, we propose to compare the sparsity of two classes of representations for the EAP : The discrete representations, via the Haar, Daubechies-Cohen-Fauveau (DCF) 5-3, DCF 9-7 wavelets bases, and the continuous representations, via Spherical Polar Fourier (SPF) and 3D-SHORE bases.

This work has been published in [28].

6.1.2.8. Parametric dictionary learning for modeling EAP and ODF in diffusion MRI
Participants: Sylvain Merlet, Emmanuel Caruyer, Rachid Deriche.

In this work, we propose an original and efficient approach to exploit the ability of Compressed Sensing (CS) to recover Diffusion MRI (dMRI) signals from a limited number of samples while efficiently recovering important diffusion features such as the Ensemble Average Propagator (EAP) and the Orientation Distribution Function (ODF). Some attempts to sparsely represent the diffusion signal have already been performed. However and contrarly to what has been presented in CS dMRI, in this work we propose and advocate the use of a well adapted learned dictionary and show that it leads to a sparser signal estimation as well as to an efficient reconstruction of very important diffusion features. We first propose to learn and design a sparse and parametric dictionary from a set of training diffusion data. Then, we propose a framework to analytically estimate in closed form two important diffusion features : the EAP and the ODF. Various experiments on synthetic, phantom and human brain data have been carried out and promising results with reduced number of atoms have been obtained on diffusion signal reconstruction, thus illustrating the added value of our method over state-of-the-art SHORE and SPF based approaches.

This work has been published in [25].

6.1.2.9. Constrained Diffusion Kurtosis Imaging Using Ternary Quartics and MLE
Participants: Tristan Milne [Queen's University, Kingston, Ontario, Canada], Aurobrata Ghosh, Rachid Deriche.

This work has been partly supported by the Inria International Internship Program.

We present a ternary quartic based approach with an improved gradient based optimization scheme for diffusion kurtosis imaging to estimate constrained and physically realistic diffusion and kurtosis tensors. We account for the signal noise by considering a maximum likelihood estimation based on the Rician noise model. Diffusion kurtosis imaging (DKI) is a recent important improvement over the diffusion tensor imaging (DTI) model that quantifies the degree of non-Gaussian diffusion in a tissue using a 3D 4th order tensor. However, DKI estimation needs to consider three constraints to be physically relevant. By adopting the implicit ternary quartic parameterization which allows to elegantly impose a positivity constraint on the kurtosis tensor and by employing gradient based optimization schemes, we show dramatically improved performance in terms of estimation time and quality. We derive the mathematical framework and show results on extensive synthetic data experiments.

This work has been published in [30]. A longer version has been submitted and is under revision for the journal Magnetic Resonance in Medicine.

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

6.1.3.1. From Diffusion MRI to Brain Connectomics
Participants: Aurobrata Ghosh, Rachid Deriche.

Diffusion MRI (dMRI) is a unique modality of MRI which allows one to indirectly examine the microstructure and integrity of the cerebral white matter in vivo and non-invasively. Its success lies in its capacity to reconstruct the axonal connectivity of the neurons, albeit at a coarser resolution, without having to operate on the patient, which can cause radical alterations to the patient's cognition. Thus dMRI is beginning to assume a central role in studying and diagnosing important pathologies of the cerebral white matter, such as Alzheimer's and Parkinson's diseases, as well as in studying its physical structure in vivo. In this work, we present an overview of the mathematical tools that form the framework of dMRI – from modelling the MRI signal and measuring diffusion properties, to reconstructing the axonal connectivity of the cerebral white matter, i.e., from Diffusion Weighted Images (DWIs) to the human connectome.

This work will be published in [55].

6.1.3.2. Biomarkers for HARDI : 2nd & 4th Order Tensor Invariants
Participants: Rachid Deriche, Aurobrata Ghosh, Théodore Papadopoulo.

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

In this paper, we explore the theory of tensor invariants as a mathematical framework for computing new biomarkers for HARDI. We present and explain the integrity basis, basic invariants and principal invariants of 2nd & 4th order tensors to expand on a recently proposed paper on 4th order tensor invariants. We present the mathematical results and compute the basic and principal invariants on a controlled synthetic dataset and an in vivo human dataset. We show how the integrity bases of these two sets of invariants can form a promising framework for developing new biomarkers for HARDI.

This work has been published in [22].

6.1.3.3. Generalized Invariants of a 4th order tensor: Building blocks for new biomarkers in dMRI
Participants: Aurobrata Ghosh, Théodore Papadopoulo, Rachid Deriche.

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

This paper presents a general and complete (up to degree 4) set of invariants of 3D 4th order tensors with respect to SO3. The invariants to SO3 for the 2nd order diffusion tensor are well known and play a crucial role in deriving important biomarkers for DTI, e.g. MD, FA, RA, etc. But DTI is limited in regions with fiber heterogeneity and DTI biomarkers severely lack specificity. 4th order tensors are both a natural extension to DTI and also form an alternate basis to spherical harmonics for spherical functions. This paper presents a systematic method for computing the SO3 invariants of 3D 4th order tensors, derives relationships between the new (generalized) invariants and existing invariants and shows results on synthetic and real data. It also presents, hitherto unknown, new invariants for 4th order tensors. Analogously to DTI, these new invariants can perhaps form building blocks for new biomarkers.

This work has been published in [23].

6.1.3.4. Tractography via the Ensemble Average Propagator in diffusion MRI

Participants: Sylvain Merlet, Anne-Charlotte Philippe, Rachid Deriche, Maxime Descoteaux [Sherbrooke Connectivity Imaging Laboratory, Computer Science Departement, Université de Sherbrooke, Québec, Canada].

This work has been partly supported within the framework of the Brain Connectivities Associate Team.

It's well known that in diffusion MRI (dMRI), fibre crossing is an important problem for most existing diffusion tensor imaging (DTI) based tractography algorithms. To overcome these limitations, High Angular Resolution Diffusion Imaging (HARDI) based tractography has been proposed with a particular emphasis on the the Orientation Distribution Function (ODF). In this work, we advocate the use of the Ensemble Average Propagator (EAP) instead of the ODF for tractography in dMRI and propose an original and efficient EAP-based tractography algorithm that outperforms the classical ODF-based tractography, in particular, in the regions that contain complex fibre crossing configurations. Various experimental results including synthetic, phantom and real data illustrate the potential of the approach and clearly show that our method is especially efficient to handle regions where fiber bundles are crossing, and still well handle other fiber bundle configurations such as U-shape and kissing fibers.

This work has been published in [29].

6.1.3.5. Using Radial NMR Profiles to Characterize Pore Size Distributions

Participants: Rachid Deriche, John Treilhard [Queen's University, Ontario, Canada].

This work has been partly supported by the Inria International Internship Program.

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 [78], 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 has been published [20].

6.1.3.6. Elliptic Fourier Features of Brain White Matter Pathways

Participants: Rachid Deriche, Ali Demir [Sabancy University, TU], Gozde Unal [Sabancy University, TU].

Magnetic resonance imaging provides diffusion weighted images (DMRI), which non-invasively reconstruct the brain white matter pathways. DMRI is used to study brain white matter diseases as well as aid surgical planning. As localization of different white matter pathways surrounding a pathology is crucial for surgical planning, automatic extraction and classification of different anatomical white matter pathways pre-operatively becomes an important computational tool. In this work, we propose a method for classification of brain white matter pathways based on 3D elliptic Fourier descriptors, which are extended from the 2D elliptic Fourier descriptors. We performed experiments and validation of the proposed method on a white matter atlas space and on real pathological cases.

This work has been published [41].

6.2. Multi-Imaging Modalities

6.2.1. Coupling functional and structural models

6.2.1.1. A nested cortex parcellation combining analysis of MEG forward problem and diffusion MRI tractography
Participants: Anne-Charlotte Philippe, Maureen Clerc, Théodore Papadopoulo, Rachid Deriche.

Understanding the relationship between structure and function is a major challenge in neuroscience. Diffusion MRI (dMRI) in the only non-invasive modality allowing to have access to the neural structure. Magnetoen-cephalography (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- subjects 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 has been published in [31].

6.2.1.2. dMRI tractography of WM fibers to recover the anatomical connectivity supporting a MEG epileptic network
Participants: Anne-Charlotte Philippe, Maureen Clerc, Théodore Papadopoulo, Rachid Deriche.

Cerebral organization is determined by segregated and integrated regions both functionally and anatomically. These cerebral networks are the foundations of the execution of major part of cognitive processes. Information about the structure of the white matter (WM) and the functionality of networks are both needed to understand these cerebral networks.

This work proposes an efficient method to inform a given functional network on its anatomical support: how many anatomical connections exist between functionally connected regions and what are their geometries . Diffusion MRI being the only non invasive method allowing to have access to the micro-structure of the WM, we used diffusion information to underline the degree of connectivity between functionally connected regions while taking advantage of WM fibers reconstruction to determine the way taken by the anatomical network supporting the functional network.

Due to the complex dynamical alteration of epilepsy, the study of large-scale functional connectivity is difficult. But diffusion imaging studies have shown alterations of the WM between epileptic zones and connected areas. This methodology allows to add qualitative (degree of connectivity) and geometrical (WM fibers reconstruction) information on the anatomical network supporting an epileptic network mostly determined by magneto-encephalography (MEG).

This work has been published in [35].

6.2.1.3. Whole cortex parcellation combining analysis of MEG forward problem, structural connectivity and Brodmann's atlas
Participants: Anne-Charlotte Philippe, Maureen Clerc, Théodore Papadopoulo, Rachid Deriche.

Functional cortex parcellation is one of the most important ways to understand the link between structure and function in the brain. Brodmann's atlas remains a fundamental pillar to understand this relationship because its areas are defined by similar cytoarchitecture and functional imaging notably had revealed that they correspond, entirely or in part, to functional areas. So, its integration to diffusion MRI (dMRI) data is pertinent, dMRI being the only non invasive and in-vivo imaging modality able to have access to a detailed geometric description of the anatomical connectivity between brain areas. In this work, we propose to define a new connectivity profile of cortical sources based on the Brodmann's atlas. After its registration to T1 and diffusion weighted images of the same subject, we reconstructed the brain surfaces and considered the cortical sources to be the vertices of the white matter/ grey matter boundary mesh. We performed a probabilistic tractography taking each cortical sources as seeds and the L Brodmann's areas as L targets. Thus, we obtained the connectivity profile of a cortical source: a vector v of size L where v (l) is the degree of connectivity of the source to the l th Brodmann's

area. Then, we developed a cortical parcellation method jointly analyzing the MEG forward problem and the connectivity profiles based on Brodmann's atlas of cortical sources. We computed the leadfield matrix that relates the sources to the MEG sensors. We applied a k-means algorithm to the leadfield matrix to cluster sources having a close magnetic field to the MEG sensors. Then, in each leadfield-based cluster, we clustered sources via their connectivity profile based on Brodmann's atlas. This automatic parcellation is an efficient preprocessing to compute a MEG inverse problem on functional data informed by its structural connectivity.

This work has been published in [32].

6.2.1.4. Study of the brain connectivity in an Immersive Space
Participants: Anne-Charlotte Philippe, Jean-Christophe Lombardo [Dream Project-Team, Inria, Sophia Antipolis, Méditerranée, France].

Virtual reality is a powerful tool for scientific visualization. When the amount and complexity of the visualized data grows, standard visualization applications on desktop computers become inefficient. In this work, we present the use of a CAVE like VR facility in a neuroscientific context. The aim is to have a better understanding of the brain connectivity. Both anatomical and functional data are attached to a mesh representing the brain surface.

Specific tools developed for this study and the way we used them are presented in [36] emphasizing drawbacks and advantages of virtual reality in a scientific visualization context.

This work has been published in [36].

6.2.1.5. Cortex parcellation via diffusion data as prior knowledge for the MEG inverse problem
Participants: Anne-Charlotte Philippe, Maureen Clerc, Théodore Papadopoulo, Rachid Deriche.

In this work, we present a new approach to the recovery of dipole magnitudes in a distributed source mo-del for magnetoencephalographic (MEG) imaging. This method consists in introducing prior knowledge regarding the anatomical connectivity in the brain to this ill-posed inverse problem. Thus, we perform cortex parcellation via structural information coming from diffusion MRI (dMRI), the only non-invasive modality allowing to have access to the structure of the WM tissues. Then, we constrain, in the MEG inverse problem, sources in the same diffusion parcel to have close magnitude values. Results of our method on MEG simulations are presented and favorably compared with classical source reconstruction methods.

This work is currently under submission.

6.2.1.6. Fractality in the neuron axonal topography of the human brain based on 3-D diffusion MRI
Participants: Panayotis Katsaloulis [Institute of Physical Chemistry "Demokritos" (IPC),National Center for Scientific Research "Demokritos", Greece], Aurobrata Ghosh, Anne-Charlotte Philippe, Astero Provata [Institute of Physical Chemistry "Demokritos" (IPC),National Center for Scientific Research "Demokritos", Greece], Rachid Deriche.

In this work, we conduct a group study, with 18 subjects, to validate the computational robustness of the fractal dimension of the neuron axonal topography in the human brain that is derived from diffusion MRI (dMRI) acquisitions. We extend the work done in a previous paper by some of the current authors where the fractal dimension of the neuron axonal topography from dMRI data was computed from 2-D regions of interest. The fractal dimensions Df of the entire 3-D volume of the brain is here estimated via the Box Counting, the Correlation DImension and the Fractal Mass Dimension methods. 3-D neuron axon data are obtained using tractography algorithms on Diffusion Tensor Imaging of the brain. We find that all three calculations of Df give consistent results across subjects, namely, they demonstrate fractal characteristics in the short and medium length scales: different fractal exponents prevail at different length scales, an indication of multifractality. We surmise that this complexity stems as a collective property emerging when many local brain units performing different functional tasks and having different local topologies are recorded together.

This work has been published in [15].

6.3. Forward and Inverse Problems

6.3.1. *Source localization using rational approximation on plane sections*

Participants: Maureen Clerc, Théodore Papadopoulo, Juliette Leblond [Apics Project-Team, Inria, Sophia Antipolis, Méditerranée, France], Jean-Paul Marmorat [CMA, Ecole des Mines Paristech, Sophia Antipolis, France].

In functional neuroimaging, a crucial problem is to localize active sources within the brain non-invasively, from knowledge of 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. This work presents a new method for 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 published in the journal Inverse Problems [14].

6.3.2. *The adjoint method of OpenMEEG for EEG and MEG with large source space*

Participants: Maureen Clerc, Théodore Papadopoulo, Alexandre Gramfort [Telecom Paristech], Emmanuel Olivi [Forner member of the Athena Project-Team].

In EEG or MEG, a lead field is the linear operator which associates unitary dipolar sources to the resulting set of sensor measurements. In practise, the source space often includes over 10 000 dipoles, which sometimes causes memory problems. The adjoint approach considers the forward problem from the viewpoint of sensors instead of sources: this drops down the number of linear systems to solve by two orders of magnitude. The adjoint approach is here proposed in the context of the Boundary Element Method, and its implementation is provided by the OpenMEEG library.

This work was presented at the BIOMAG conference [38].

6.3.3. *Comparison of Boundary Element and Finite Element Approaches to the EEG Forward Problem*

Participants: Maureen Clerc, Carsten Wolters [Institute for Biomagnetism and Biosignal Analysis, University of Münster], Johannes Vorwerk [Institute for Biomagnetism and Biosignal Analysis, University of Münster], Martin Burger [Institut für Numerische und Angewandte Mathematik, Fachbereich Mathematik und Informatik, Westfälische Wilhelms Universität (WWU) Münster], Jan de Munck [Vrije Universiteit Medical Centre (VUMC), The Netherlands].

The accurate simulation of the electric fields evoked by neural activity is crucial for solving the inverse problem of EEG. Nowadays, boundary element methods (BEM) are frequently applied to achieve this goal, usually relying on the simpli-fication of approximating the human head by three nested compartments with isotropic conductivities (skin, skull, brain). Here, including the highly-conducting cerebrospinal fluid (CSF) is a difficult task due to the complex geometrical structure of the CSF, demanding a high number of additional nodes for an accurate modeling and thus a strongly increased computational effort. Though, CSF conductivity is well-known and nearly not varying inter-individually and its significant influence on EEG forward simulation has been shown. The CSF can be included at negligible computational costs when applying finite element (FE) forward approaches. In this study we compare the accuracy and performance of state-of-the-art BE and FE approaches in both artificial and realistic three layer head models, showing that all approaches lead to high numerical accuracies. Furthermore, we demonstrate the significant influence of modeling the CSF compartment as disregarding this compartment leads to model errors that lie clearly above the observed numerical errors.

A book chapter on BEM and FEM models has been published in the Handbook for Neural Activity Measurement [40]. The comparison was presented at the BIOMAG conference [19].

6.3.4. Domain Decomposition to handle versatile conductivity models

Participants: Maureen Clerc, Théodore Papadopoulo, Emmanuel Olivi [Forner member of the Athena Project-Team].

Source localization from external data such EEG or MEG, requires a good understanding of the electromagnetic behavior of the patient head. Several models can been 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, it is easier with BEM than with FEM to accurately represent sources in isotropic media. Thanks to domain decomposition, we propose to solve a EEG forward problem using BEM where the sources are (the brain) and FEM for other tissues (with notably inhomogeneities in the skull).

This work was presented at the BIOMAG conference [37].

6.4. Brain Computer Interfaces

6.4.1. Combining ERD and ERS features to create a system-paced BCI

Participants: Maureen Clerc, Théodore Papadopoulo, Joan Fruitet, Eoin Thomas.

An important factor in the usability of a brain computer interface (BCI) is the setup and calibration time required for the interface to function accurately. Recently, brain-switches based on the rebound following motor imagery of a single limb effector have been investigated as basic BCIs due to their good performance with limited electrodes, and brief training session requirements. Here, a BCI is proposed which expands the methodology of brain-switches to design an interface composed of multiple brain-buttons. The algorithm is designed as a system paced interface which can recognise 2 intentional-control tasks and a no-control state based on the activity during and following motor imagery in only 3 electroencephalogram channels. An online experiment was performed over 6 subjects to validate the algorithm, and the results show that a working BCI can be trained from a single calibration session and that the post motor imagery features are both informative and robust over multiple sessions.

This work, which was partially presented at the EMBS conference [33], is currently under revision for the Journal of Neuroscience Methods.

6.4.2. Bandit algorithms for faster task selection in BCI

Participants: Maureen Clerc, Joan Fruitet, Alexandra Carpentier [Sequel Project-Team, Inria Lille, France], Rémi Munos [Sequel Project-Team, Inria Lille, France].

BCIs based on sensorimotor rhythms use a variety of motor tasks, such as imagining moving the right or left hand, the feet or the tongue. Finding the tasks that yield best performance, specifically to each user, is a time consuming preliminary phase to a BCI experiment. This study presents a new adaptive procedure to automatically select, online, the most promising motor task for an asynchronous brain-controlled button.
We develop for this purpose an adaptive *Upper Confidence Bound* algorithm based on the stochastic bandit theory, and design an EEG experiment to test our method. We compare (offline) the adaptive algorithm to a naive selection strategy which uses uniformly distributed samples from each task. We also run the adaptive algorithm online to fully validate the approach.
By not wasting time on inefficient tasks, and focusing on the most promising ones, this algorithm results in a faster task selection and a more efficient use of the BCI training session. More precisely, the offline analysis reveals that the use of this algorithm can reduce the time needed to select the most appropriate task by almost half without loss in precision, or alternatively, allow to investigate twice the number of tasks within a similar time span. Online tests confirm that the method leads to an optimal task selection.
This study is the first one to optimize the task selection phase by an adaptive procedure. By increasing the number of tasks that can be tested in a given time span, the proposed method could contribute to reducing "BCI illiteracy".

This work is the result of the collaboration between Sequel and Athena within the ANR CoAdapt. It has been published in NIPS [21] and is accepted in the Journal of Neural Engineering [52].

6.4.3. *An analysis of performance evaluation for motor-imagery based BCI*

Participants: Maureen Clerc, Matthew Dyson [Laboratoire de Neurosciences Cognitives, Aix-Marseille Université, France], Eoin Thomas.

In recent years, numerous brain computer interfaces (BCIs) have been proposed which incorporate features such as adaptive classification, error detection and correction, fusion with auxiliary signals and shared control capabilities. Due to the added complexity of such algorithms, the evaluation strategy and metrics used for analysis must be carefully chosen to accurately represent the performance of the BCI. In this article, metrics are reviewed and contrasted using both simulated examples and experimental data. Furthermore, a review of the recent literature is presented to determine how BCIs are evaluated, in particular focusing on the correlation between how the data are used relative to the BCI subcomponent under investigation. From the analysis performed in this study, valuable guidelines are presented regarding the choice of metrics and evaluation strategy dependent upon any chosen BCI paradigm.

This work was supported by the ANR Co-Adapt and is currently under revision for the Journal of Neural Engineering.

AXIS Project-Team

5. New Results

5.1. Introduction

As planned, our new results are splitted into our three sub-objectives as introduced below:

5.1.1. *Mining for Knowledge Discovery in Information Systems*

This year we get six main results: one related to how to integrate domain knowledge in a multi-view KDD process (cf. section 5.2.4), two on new KDD methods involving clustering (cf. sections 5.2.3 and 5.2.2), one on the construction of hierarchical structures of concepts in the field of e-tourism (cf. section 5.2.6), one on partitioning objects taking into account simultaneously their relational descriptions given by multiple dissimilarity matrices (cf. section 5.2.1), and finally improvment of our work on critical edition of Sanskrit texts (cf. section 5.2.5).

- Zhang based on his thesis (2010) has published this year his work on modeling and clustering users with evolving profiles in usage streams [32]. This paper will propose three models to summarize bi-streaming data, which are the batch model, the Evolving Objects (EO) model and the Dynamic Data Stream (DDS) model. Through creating, updating and deleting user profiles, the models summarize the behaviours of each user as an object. Based on these models, clustering algorithms are employed to identify the user groups. The proposed models are tested on a real-world data set showing that the DDS model can summarize the bi-streaming data efficiently and effectively, providing better basis for clustering user profiles than the other two models.
- The work described in 2011(see our AxIS annual report) on critical edition of Sanskrit texts and submitted as a paper at the Cicling 2012 conference has been accepted [21].
- A past work accepted in an international journal with A. Ciampi and colleagues [16].
- One article in an international conference on functional data analysis issued from a collaboration with F. Rossi [40].
- Two articles have been deposit in the Computing Research Repository (CoRR): one on clustering Dynamic Web Usage data [65] from A. Da Silva's thesis and one on functional data analysis [66].

5.1.2. *Information and Social Networks Mining for Supporting Information Retrieval*

This year, we pursued two main works on clustering methods:

- the detection of communities in a social network (graph extracted from relationnal data) (cf. section 5.3.1),
- the improvment of our dynamic hard clustering method for relational data (cf. section 5.3.2).

5.1.3. *Multidisciplinary Research For Supporting User Oriented Innovation*

With the expansion of the innovation community beyond the firm's boundaries (the so-called "open innovation") a lot of changes have been introduced in design and evaluation processes : the users can become co-designers, HCI design and evaluation focus is no longer placed on usability only but also on the whole user experience, experimentations take place out of lab with large number of heterogeneous people instead of carefully controlled panels of users ... All these deep changes require improvements of existing practices, methods and tools for the design / evaluation of information systems as well as for usage analysis. This evolution calls also for a structured user centered methodology (methods and ICT tools) to deal with open innovation. Various different disciplines and trends are dedicated in understanding user behavior on Internet and with Digital Technologies, notably Human Computer Interaction community (HCI), CSCW, Workplace Studies, Distributed Cognition and Data Mining. Our contribution to open innovation research keeps its focus on usage analysis for design, evaluation and maintenance of information systems and our activities this year, as indicated in our roadmap presented at the Inria theme evaluation (2011) have been conducted both breadth wise and in depth with two main objectives :

- Improving, designing and evaluation support tools for innovation,
- Development of the FocusLab platform.

The research was conducted along three focus:

- Extension of usability methods and models (cf. section 5.4),
- Designing and evaluating user experience in the context of a living lab process (cf. section 5.5),
- FocusLab Platform (cf. section 5.6).

Let us note one research work related to Living labs done in 2011 and published in 2012 [26].

5.2. Mining for Knowledge Discovery in Information Systems

5.2.1. *Clustering on Multiple Dissimilarity Matrices*
Participants: Yves Lechevallier, F.A.T. de Carvalho, Guillaume Pilot, Brigitte Trousse.

In [17], we introduce hard clustering algorithms that are able to partitioning objects taking into account simultaneously their relational descriptions (relations + values) given by multiple dissimilarity matrices. 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.

During 2012 we show interest and disadvantages of these approaches to classifying curves with a Urso and Vichi distance based on the mathematical properties of curves (first derivative and second). The curves are issued from temperature sensors placed in 40 offices during one year (See section 6.1.3). This period was divided into the periods before and after challenge and the challenge period. During the challenge period the occupants had information by bonus / malus messages on energy consumption [34].

5.2.2. *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. We pursued our previous work [102] and defined a new approach of knowledge extraction using graph theory. which is described in [29].

This work is done in collaboration with the laboratory LRIA At the Ferhat Abbas University, Sétif, Algérie.

5.2.3. *Multi-criteria Clustering with Weighted Tchebycheff Distances for Relational Data*
Participants: F.A.T. de Carvalho, Yves Lechevallier.

The method described in [27] uses a nonlinear aggregation criterion, weighted Tchebycheff distances, more appropriate than linear combinations (such as weighted averages) for the construction of compromise solutions. We obtain a partition of the set of objects, the prototype of each cluster and a weight vector that indicates the relevance of each criterion in each cluster. Since this is a clustering algorithm for relational data, it is compatible with any distance function used to measure the dissimilarity between objects.

5.2.4. *Knowledge management in Multi-View KDD Process*
Participant: Brigitte Trousse.

E.L. Moukhtar Zemmouri, in the context of his PhD thesis supervised by Hicham Behja, A. Marzark and B. Trousse pursued his work based on a Viewpoint Model in the context of a KDD process [30], [19].

Knowledge Discovery in Databases (KDD) is a highly complex, iterative and interactive process aimed at the extraction of previously unknown, potentially useful, and ultimately understandable patterns from data. In practice, a KDD process (data mining project according to CRISP-DM vocabulary) involves several actors (domain experts, data analysts, KDD experts, etc.) each with a particular viewpoint. We define a multi-view analysis as a KDD process held by several experts who analyze the same data with different viewpoints.

We propose to support users of multi-view analysis through the development of a set of semantic models to manage knowledge involved during such an analysis. Our objective is to enhance both the reusability of the process and coordination between users.

To do so, we propose first a formalization of viewpoint in KDD and a Knowledge Model that is "a specification of the information and knowledge structures and functions involved during a multi-view analysis". Our formalization, using OWL ontologies, of viewpoint notion is based on CRISP-DM standard through the identification of a set of generic criteria that characterize a viewpoint in KDD. Once instantiated, these criteria define an analyst viewpoint. This viewpoint will guide the execution of the KDD process, and then keep trace of reasoning and major decisions made by the analyst.

Then, to formalize interaction and interdependence between various analyses according to different viewpoints, we propose a set of semantic relations between viewpoints based on goal-driven analysis. We have defined equivalence, inclusion, conflict, and requirement relations. These relations allow us to enhance coordination, knowledge sharing and mutual understanding between different actors of a multi-view analysis, and reusability in terms of viewpoint of successful data mining experiences within an organization.

5.2.5. *Critical Edition of Sanskrit Texts*

Participants: Yves Lechevallier [correspondant], Marc Csernel, Ehab Assan.

With the help of Ehab Assan we improved the prototype made last year by Nicolas Bèchet (cf. 2011 AxIS activity report,[21]). It is now included in the construction process of critical editions of Sanskrit texts. Ehab also added LaTeX output to the process, we now have paper as well as Web output. It was possible to present these new features [33], [36] at the 13th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing) in Delhi.

5.2.6. *Construction and Settlement of hierarchical Structures of Concepts in E-tourism*

Participant: Yves Lechevallier.

The work of Nicolas Bechet (AxIS member in 2011) and Yves Lechevallier in collaboration with Marie-Aude Aufaure (Ecole Centrale de Paris), was published in 2012 [20] related to a method for the construction and the automatic settlement of hierarchical structures of concepts. We were particularly interested in the construction of a hierarchical structure of services offered in Hotels from a data set of an application in the field of e-tourism motivated by our contacts with the SME Addictrip. The goal is to associate to each service a concept that provides a common representation of all services. Our experiments are carried out using resources from partners specialized in online hotel booking, in particular from Addictrip. The establishment of a structure of concepts is essential to these partners that use their own terminologies description of hotel services. Indeed it provides a common representation space allowing the comparison of service coming from different resources. Our approach is based on proximity of literal terms in the service having a nearby measure based on n-grams of characters. The results during our experiments show the quality of this approach and its limitations.

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, Bruno Almeida Pimentel.

The automatic detection of communities in a social network can provide a kind of graph aggregation. The objective of graph aggregations is to produce small and understandable summaries and it 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.

During this year many updates of the program developed by Louati Amine in 2011 were performed by Bruno Almeida Pimentel. A book chapter, included the new aggregation criteria proposed ans evaluted by Bruno Almeida Pimentel, was written and will be published in 2013.

This work is done in collaboration with Marie-Aude Aufaure, head of the Business Intelligence Team, Ecole Centrale Paris, MAS Laboratory.

5.3.2. *Multi-View Clustering on Relational Data*

Participants: Thierry Despeyroux, Yves Lechevallier.

In the work reported in [23] in collaboration with Francisco de A.T. de Carvalho, we introduce an improvement of a clustering algorithm described in [17] that is able to partition objects taking into account simultaneously their relational descriptions given by multiple dissimilarity matrices. In this version of the prototype clusters depend on the variables of the representation space. These matrices could have been generated using different sets of variables and dissimilarity functions. This method, which is based on the dynamic clustering algorithm for relational data, is designed to provided a partition and a vector of prototypes for each cluster as well as to learn a relevance weight for each dissimilarity matrix by optimizing an adequacy criterion that measures the fit between clusters and their representatives. These relevance weights change at each algorithm iteration and are different from one cluster to another. Moreover, various tools for the partition and cluster interpretation furnished by this new algorithm are also presented.

Two experiments demonstrate the usefulness of this clustering method and the merit of the partition and cluster interpretation tools. The first one uses a data set from UCI machine learning repository concerning handwritten numbers (digitalized pictures). The second uses a set of reports for which we have an expert classification given a priori. which we have an expert classification given a priori.

5.4. Extension of Usability Methods and Tools

5.4.1. *User Evaluation and Tailoring of Personal Information*

Participants: Claudia Detraux, Dominique Scapin.

In the context of the ANR project PIMI (Personal Information Management through Internet see section 6.2.1), ergonomic inspections have been carried out to evaluate the usability of the PIMI V0.1. prototype, in its PC and mobile versions [49], [48]. Also, an experiment [24], [35] was conducted on a mockup of a Personal Information Space. Users were asked to perform scenario-based data entry and retrieval tasks, then to modify the mockup according to their wishes and needs. The results allowed to validate the item content and structure for the future personal space, as well as to assess the role of user modifications as evaluation cues, and for the development of further ergonomic recommendations. Detailed information was obtained on how users enter and retrieve data, by modifying the interface settings and shows that the adaptable nature of a Personal Information Space can indeed influence its acceptance, and provides useful cues for ergonomic evaluation

5.4.2. *Usability Methods for Information Visualization*

Participant: Dominique Scapin.

A collaboration between UFRGS (Federal University of Rio Grande do Sul, Institute of Informatics), Brazil and Inria-AxIS led to a book chapter [37] dealing with potential methodologies for including a user-centered approach into information visualization techniques. It starts by presenting the evolution of visualization techniques evaluation, briefly summarizing the main contributions in this area since its humble beginning as a collateral activity until the recent growth of interest. Then, the focus is on current issues related to such evaluations, particularly concerning the way they are designed and conducted, taking into account a background of well-known usability evaluation methods from HCI to help understanding why there are still open problems. A set of guidelines for a (more) user-centered usability evaluation of information visualization techniques is proposed and discussed.

5.4.3. *Usability Recommendations for MIS (Mixed Interactive Systems)*

Participant: Dominique Scapin.

A collaboration between University of Toulouse – IUT Tarbes, IRIT and Inria-AxIS led to a book chapter [38] dealing with Mixed Interactive Systems (MIS) which denote an advanced form of interaction that aims at combining physical and digital worlds, such as mixed and augmented reality, tangible user interfaces, ubiquitous computing, etc. Their main interest relates to the use of physical artifacts from the user's activity customary context. The book chapter first reports on a systematic review of the literature on MIS evaluation. From that review, usability recommendations were selected and deciphered before reformulating them under a common format. Finally, three different classification schemes of the usability recommendations obtained are proposed to facilitate search and retrieval, but also to better integrate them into the MIS development process.

5.5. Designing and Evaluating User Experience and Methods for Open Innovation

5.5.1. *From Usability to User Experience: an HCI Review*

Participants: Dominique Scapin, Bernard Senach, Brigitte Trousse, Marc Pallot.

Through an extensive review of the literature, a paper [28] attempted to characterize a rather novel and popular view on human-computer interaction: User Experience (UX). After introducing its polysemous nature, this paper describes the origins of UX, its scope, underlying concepts and components, as well as its various definitions. Then, UX methods are surveyed and classified, distinguishing processes, frameworks, and specific methods. The paper identifies a set of issues about the needs for increased UX maturity. Even though UX can still be viewed as an extension of usability, its future may correspond to a paradigm evolution rather than simply a buzz word. The evolution is not drastic, but it adds complexity (including new measurements) by considering a few more user areas than traditional usability.

5.5.2. *Evaluation of our Methods for Idea Generation Process*

Participants: Anne-Laure Negri, Caroline Tiffon, Brigitte Trousse, Bernard Senach.

In 2011 we proposed a methodology coupling two methods [25] (GenIoT a generative method based on probes (fake sensors and/or actuators) and ALoHa! a bodystorming method for designing service concepts in the specific paradigm of the Internet of Things (IoT). In the frame of the European project ELLIOT - Experiential Living Lab for Internet Of Things -, ICT Usage Lab (cf. section 6.1.8) aims at co-creating "green" services, i.e. services based on air quality and noise measurement.

Both IoT ideation methods Aloha! and GenIoT were used for the co-creation of health related services (cf. section 6.3.1.1   ). The participants of the methods were Environment and Health professional. Results were very different than the workshops run with citizen in the frame of the mobility scenario in ELLIOT (see 2011 AxIS activity report). Comparison of these workshops shows that hybrid approaches –i.e. co-creation approaches mixing both real and virtual meetings are not working as well as pure face to face or pure online approaches. Moreover, GenIoT method seems to be more effective with citizen than with professional. Aloha! is effective in both groups but more efficient with professionals. However the participant experience of Aloha! is higher in the case of citizen (mainly because professionals are not used to practice creative thinking methods and do not appreciate to go out of their comfort zone).

5.5.3. *Leading People Behavior Changes: Mining Evolutive Data*

Participants: Brigitte Trousse, Yves Lechevallier, Guillaume Pilot, Carole Goffart, Bernard Senach.

The ECOFFICES project (cf. [62], [22] and section 6.1.3) was for AxIS project team our first step towards eco-behavior study. It provided us a very rich context to study how to analyse the evolution of the energy consuming of employees during an energy challenge. A qualitative analysis from questionnaires (before and after the challenge) has been done as well quantitative analysis. The data set for quantitative data is composed of heterogeneous data issued from around 400 sensors (temperature, presence, behavior in terms of opening doors, windows, bonus, malus, etc.). We made different studies related to data preprocessing and data analysis. In our first study [64], we cleaned the data set and selected reliable data for data analysis (only temperature of various equipments, user presence and bonus/malus points). We decided not to work with aggregated variables such as the initial ponderation (defined by partners) for the various bonus-malus rules and the energy consuming at the office level. We decided to use (office, day) as statistical unit (i.e. 9995 units) with a vectoral representation. Finally we realized that the three initial periods (before the challenge, during and after the challenge) on 379 days (2011-2012) should be in fact decomposed in five periods, due to the fact the first and the last periods were split into two subperiods (with and without heat). For the analysis, we apply for each (office,day) a first analysis on a vectorial representation of temperature with the MND method (cf. section focuslab) in order to identify the best partition of these. The MND method uses euclidean distance between each value of the vectorial representation and the prototypes are defined by the means. Second we did a clustering of these units based on bonus and malus and finally we made the correspondence between these two partitions. Three classes for (office, day) are obtained. The interpretation in terms of team relied difficult but we proposed various conclusions for a winner for managing a specific bonus, or in managing ambient temperature or in behavioral change.

In our second study [34] in collaboration with Francisco de A.T. De Carvalho, our goal was to improve the interpretation task at the office and team level by applying AxIS advanced methods. To do this, we applied our hard clustering method presented in [34] on this dataset where each office was characterized by two different representations:

- Interval representation: each office is characterized by a vector of intervals corresponding to the average, minimum and maximum of daily temperatures on the three temperature sensors during these five periods. Then the office is represented by a vector of 15 intervals and the distance used is Hausdorff distance. This classification is consistent with the partition into three classes obtained during the ECOFFICES project. The class obtained with nine eco-responsible ecoffices is the same. However, other offices are divided into two classes according to the type of heating used during the winter period. The classical method divided these offices into two clusters, one of which contains the offices using the radiators during the winter period.

- Sequential representation: Each office is characterized by a vector of 9 measures, the min, max and average of daily temperatures of the three sensors in these five periods. The values are ordinate versus time and the distance used is Urso and Vichi distance (adequate for curves). The results of this approach are quite different from the classical approach results. These results required more effort for their interpretation in collaboration by specialists.

5.5.4. *Leading People Behavior Changes: a Literature Review*

Participants: Bernard Senach, Anne-Laure Negri.

Our research towards eco-behavior study started with the ECOFFICES project (cf. sections 5.5.3 and 6.1.3 for more details) was recently complemented with a litterature review aiming at a deeper understanding of breaks and levers to eco behavior adoption. A first work was focused on the so-called "modal change problem", compiling methods and tools aimed at supporting people to use public transportation system rather than their personal car. A second work was initiated to get a better understanding of the role that users interface could play in encouraging people to adopt a specific behavior. This work is still in progress.

Eco mobility : prompting people to adopt public transportation mode rather than their personal car.

The first review of work conducted in the fields of Persuasive Communication, Commitment, Nudges and Persuasive Technology showed that behavioral change is a process with many steps requiring to support each step with specific means. For instance, if mass communication can support the public awareness of a problem, information is not sufficient to convince people to really change their behavior. It is necessary to push them to act and numerous well-known influence techniques are nowadays available. All recent technological development (geo localization, mobile devices, social networks) can provide very effective support for behavioral changes as far as they rely on design principles identified by research in Persuasive Technology. A presentation was done on this topic for GreenCode Forum [67] (see the video on youtube).

5.5.5. *Future of Internet and User-Open Innovation for Smart Cities*

Participants: Marc Pallot, Brigitte Trousse, Bernard Senach.

We pursued our work on this topic and contributed to a white paper [59] which is one of the main outcomes of the FIREBALL project [cf. section 6.3.1.2), a Coordination Action within the 7th Framework Programme for ICT, running in the period 2010-2012. The aim of this project was to bring together communities and stakeholders who are active in three areas, namely: research and experimentation on the Future Internet (FIRE); open and user-driven innovation in Living Labs; and urban development for smarter cities. The goal was to develop a common vision on how the different methodologies and concepts in these areas can be aligned for cities as playgrounds of open and user driven innovation related to the Future Internet.

The white paper addresses several aspects that are critical for understanding the 'smart city' concept and the current progress in this area. Based on cases studies and foresight reports we aim to shed light on how the concept of smart city is currently adopted by European Cities and what the ambitions and expectations are in using this concept. It investigates the drivers and bottlenecks that influence the transformation towards a "smart city". Underlying approaches to smart cities are discussed, both in terms of the strategies and planning approaches. From this point of view, this paper explores the conditions that must be established to stimulate the transformation towards smart cities, and the resources that are available or should be made available such as investments in broadband networks and in smart applications, as well as in the capabilities to innovate. This also points to the changing structures and processes of innovation and city development. Interestingly, we see a tendency towards more decentralized and bottom-up approaches to planning and innovation. Innovation ecosystems are characterized by a combination of top down and bottom up initiatives, leading to networking and collaboration among stakeholders, which eventually extend to real innovation communities. Increasingly, citizens, advanced businesses and local governments act as proactive catalysers of innovation, shaping cities as "agents of change".

5.6. FocusLab Platform

5.6.1. *FocusLab platform: software part*

Participants: Brigitte Trousse, Yves Lechevallier, Semi Gaieb, Xavier Augros, Guillaume Pilot, Florian Bonacina.

FocusLab v1.3 (software component) done inside the ELLIOT project (cf. section 6.3.1.1) and for the purposes of the CPER Telius (cf. section 6.1.5) corresponds to the design and the implementation of a set of web-services providing basic and advanced functionalities for data analysis and some other tools supporting the living lab process.

In this version, five data analysis web services are proposed including three generic web services: a classical linear regression and two AxIs methods:

- SMDS/SCDS [91]: SCDS (Sequence Clustering in Data Stream) is a clustering algorithm for mining sequential patterns (Java) in data streams developed by A. Marascu during her thesis. This software takes batches of data in the format "Client-Date-Item" and provides clusters of sequences and their centroids in the form of an approximate sequential pattern calculated with an alignment technique. We propose in this version to return the apparition frequency (min, max, average, slope) of a sequential pattern from data streams (SCDS algorithm) (see references

- GEAR for data streams compression [93], [91], [92], [94]: GEAR (REGLO in french) is an implementation of the history management strategy proposed in Marascu's thesis [1]. It takes a set of time series and provides a memory representation of these series based on a new principle, where salient events are important (in contrast to the recent events of decaying models) .

Other data analysis services and tools have been added for Living Labs needs. We propose also two clustering methods which must be downloaded as standalone software and used for mining data from living labs:

- ATWUEDA (Axis Tool for Web Usage Evolving Data Analysis) for Analysing Evolving Web Usage Data (Da Silva 'thesis 2009 [79], [83], [81], [82]) was developed in Java and uses the JRI library (http://www.r-project.org/). The ATWUEDA tool is able to read data from a cross table in a MySQL database, split the data according to the user specifications (in logical or temporal windows) and then apply the proposed approach in order to detect changes in dynamic environment. Such an approach characterizes the changes undergone by the usage groups (e.g. appearance, disappearance, fusion and split) at each timestamp. Graphics are generated for each analysed window, exhibiting statistics that characterizes changing points over time. This application for the next experiment of Green services use case is under study.

- MND method (Dynamic Clustering Method for Multi-Nominal Data) [90]: The proposed MND method (developed in C++ language) determines iteratively a series of partitions which improves at each step the underlying clustering criterion. The algorithm is based on: a) Prototypes for representing the classes; b) Representation space; c) Proximities (distances or similarities) between two individuals; d) Context-dependent proximity functions for assigning the individuals to the classes at each step. The clustering criterion to be optimized is based on the sum of proximities between individuals and the prototype of the assigning clusters.

  This method has been also successfully applied on Web logs in 2003. This year we improved our code and tested it on IoT data (temperature) issued from the ECOFFICES project (cf. sections 5.5.3 and 6.1.3).

The application of the services provided by FocusLab 1.3 and other AxIS data mining methods for the purposes of ELLIOT use cases and other experimental projects are under study.

AYIN Team

6. New Results

6.1. Markov random fields and Marked point processes

6.1.1. *Stochastic modeling for very high resolution SAR image processing*

Participants: Aurélie Voisin, Vladimir Krylov, Josiane Zerubia [contact].

This work was done in collaboration with DITEN, University of Genoa, with Dr. Gabriele Moser and Pr. Sebastiano B. Serpico with partial financial support of the French Defense Agency, DGA (http://www.defense. gouv.fr/dga/). The data were provided by the Italian Space Agency, ISA (http://www.asi.it/en) in the framework of the project "Development and validation of multitemporal image analysis methodologies for multirisk monitoring of critical structures and infrastructures (2010-2012)".

high resolution, synthetic aperture radar data, multi-sensor data, urban areas, supervised classification, hierarchical Markov random fields, statistical modeling, wavelets, textural features

The classification of remote sensing images including urban areas is relevant in the context of the management of natural disasters (earthquakes, floodings...), and allows to determine land-use and establish land cover maps, or to localise damaged areas. Given the huge amount and variety of data available nowadays, the main difficulty is to find a classifier that takes into account multi-band, multi-resolution, and possibly multi-sensor data. A minor part of our work was also dedicated to the change detection [14], still in the frame of the management of natural disasters.

We developed a supervised Bayesian classifier that combines a joint class-conditional statistical modeling and a hierarchical Markov random field. The first classification step deals with the statistical modeling for each target class (e.g. vegetation, urban, etc.) by using a finite mixture model, estimated by resorting to a modified stochastic expectation maximization (SEM) algorithm. Such a model is well-adapted to deal with heterogeneous classes, and each mixture component reflects the contribution of the different materials contained in a given class. When considering optical images, the statistics are modeled by using finite mixtures of Gaussian distributions. In the case of SAR amplitude imagery, we favor a finite mixture of generalized Gamma distributions. Then, at each considered resolution, the different input bands are statistically combined by using multivariate copulas. The second classification step relies on the integration of this statistical modeling in a hierarchical Markov random field integrated in a quad-tree structure. Such contextual classifier helps improving the robustness of the method with respect to noise, or to SAR speckle. 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.

The first experiments were run on single-polarized, mono-resolution synthetic aperture radar (SAR) amplitude images. The challenge of the problem considered here is that the given input is at a single resolution and should be integrated in a multi-scale tree. Thus, we extract an extra information in the form of a multi-scale wavelet decomposition from the initial image. Then, at each level, a textural feature map (e.g. Haralick's variance) is obtained from each image in the decomposition stack, and integrated as an additional information that aims at discriminating the urban areas. Finally, at each level, the wavelet image is combined with the textural image by using copulas, as described previously. These results were presented in [10], [23].

Such a classifier is sufficiently flexible to take into account different types of data [21], [22]. Thus, we also tested coregistered data of a given area acquired at different resolutions (e.g., multiresolution SAR images), directly integrated at the different levels of the hierarchical tree. The classification of multisensor (optical/SAR) data is illustrated in Fig. 1 . In this specific example, we consider a GeoEye acquisition (resolution: 65 centimeters) and a coregistered COSMO-SkyMed SAR acquisition (resolution: 2.5 meters) of the Port-au-Prince quay (Haiti). Spatially disjoint training and test areas were manually annotated. The classification is done following 5 classes: urban areas, natural landscape, sand, containers and wet areas.
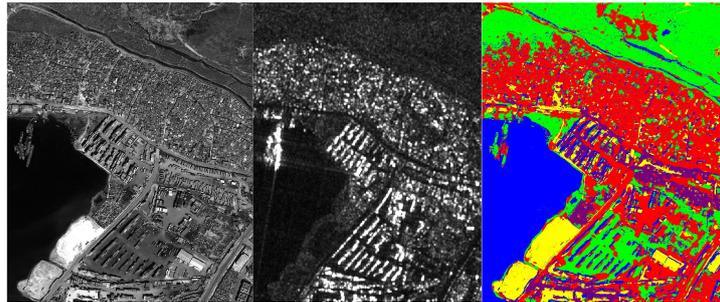

Figure 1. Left: Initial optical image of Port-au-Prince (Haiti) (©GeoEye, 2010). Middle: Initial SAR image of Port-au-Prince (Haiti) (©ISA, 2010). Right: Classification map obtained with the proposed hierarchical method for the 5 classes (Blue: wet areas; Green: vegetation; Red: urban areas; Yellow: sand; Purple: containers).

We have also run experiments on other types of acquisitions, such as histological images [22], to prove the robustness of the proposed algorithm with respect to different image sources.

6.1.2. *Satellite image classification using Bootstrap EM*

Participants: Siddharth Buddhiraju, Josiane Zerubia [contact].

This program has been partially funded by the Direction of International Relations of Inria (DRI).

Bootstrapping, Expectation-Maximization Algorithm, Iterated Conditional Expectation, Markov Random Fields, Simulated Annealing.

We implemented both Bootstrap EM and Iterated Conditional Expectation algorithms for parameter estimation of first order Markov Random Field models followed by Simulated Annealing, for optimal segmentation of gray-scale images. The objective was to perform a quantitative comparison of the two methods. Apart from successful implementation of these algorithms, an extension of these to multispectral images was performed, and the obtained results were found to be of superior quality compared against the original gray-scale ones (see Fig. 2).

First, Bootstrapped EM or Iterated Conditional Expectation were performed based on the user's requirement. The estimated parameters were then used to obtain the optimal segmentation of the image via simulated annealing. The algorithm was extended using multivariate Gaussian models to perform the same for multispectral images.

6.1.3. *Boat detection and counting in Mediterranean Harbors using Marked Point Processes*

Participants: Paula Craciun, Josiane Zerubia [contact].

This work has been conducted in collaboration with ACRI-ST (http://www.acri-st.fr/) and the French Space Agency (http://www.cnes.fr/), which provided the high resolution satellite images.

stochastic modeling, marked point process, object extraction, SEM, simulated annealing
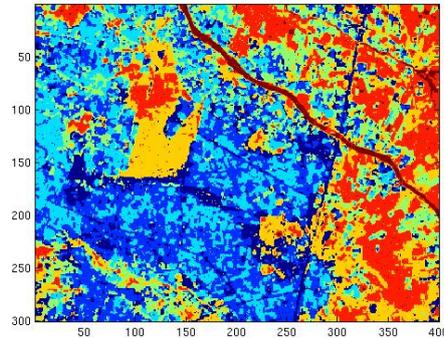
Figure 2. A sample result for a 4-band IRS-1A satellite image of a village in India. 7 clusters were considered to segment the image using Bootstrapped EM. The image was provided by Prof. Krishna Mohan Buddhiraju of CSRE, Indian Institute of Technology Bombay.

Marked point processes have been successfully applied in image processing analysis, when dealing with high resolution images in the purpose of feature extraction. The objective of this research was to improve the already existing marked point process model of ellipses to better fit the detection of boats in a harbor. The model involved two types of energy terms: a data term, used to determine the fidelity to the existing data (i.e. image) and a prior energy term, used to describe relationships between the objects. We proposed new energy components to model boat detection. The proposed model relied on a high number of parameters. While most of these parameters had an intuitive meaning and could be, thus, set manually, others were difficult to determine. We therefore used a parameter estimation method, based on the Stochastic Expectation - Maximization (SEM) algorithm, which proved to provide good results when combined with marked point processes. Furthermore, we proposed additional automatic procedures based on mathematical morphology to determine critical parameters of this model. Experimental results of boat detection are shown on Fig. 3 .

Figure 3. Result of boat detection using Marked Point Processes.

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, Josiane Zerubia [contact].

This work was done in collaboration with Xavier Descombes (Morpheme team, Inria-SAM), Dr. Pierre Coureron and Dr. Christophe Proisy at IRD, UMR AMAP (http://amap.cirad.fr/), Montpellier.

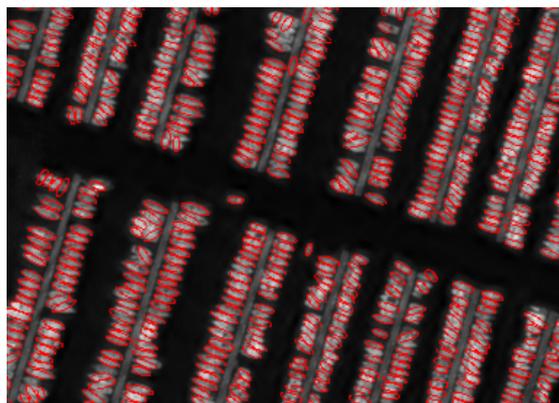object detection, marked point processes, tree crowns, forest structure

This work aimed at providing information on the forest structure through the analysis of canopy properties as described by the spatial distribution and the crown size of dominant trees. Our approach was based on the Marked Point Processes (MPP) theory, which allows modeling tree crowns observed in remote sensing images by discs belonging to a two dimensional space. The potential of MPP to detect the trees crowns automatically was evaluated by using very high spatial resolution optical satellite images of both *Eucalyptus* plantations in Brazil and mangrove forest in French Guyana. LIDAR and simulated reflectance images were also analyzed for the mangrove application. Different adaptations (parameter settings, energy models) of the MPP method were tested and compared through the development of quantitative indices that allowed comparison between detection results and tree references derived from the field, photo-interpretation or the forest mockups.

In the case of mangroves, the estimated crown sizes from detections were consistent with the outputs from the available allometric models (Fig. 4 (Left and Middle)). Other results indicated that tree detection by MPP allowed mapping the local density of trees of young *Eucalyptus* plantations (Fig. 4 (Right), [11]) even if crown size was close to the image spatial resolution (0.5 m). However, the quality of detection by MPP decreased with canopy closeness. To improve the results, further work may involve MPP detection using objects with finer shapes and forest data measurements collected at the tree plant scale.
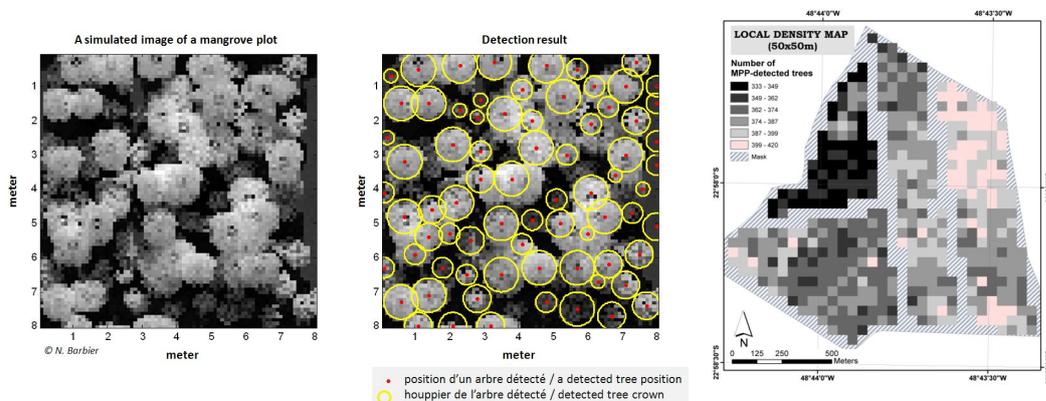
Figure 4. (Left and Middle) Result of detection on an image of mangrove. (Right) Example of a local tree density map computed over the entire plantation at 50 m resolution of the marked point process modeling. Masked areas are centered along the road network.

6.1.5. Detection of the hyperpigmentation of the skin on color images using Marked Point Process and Mathematical Morphology

Participants: Adrien Lacage, Josiane Zerubia [contact].

The source images were provided by the AYIN team itself for the study of folliculitis, and by an industrial leader in skin care for acne.

skin care, hyperpigmentation, acne, folliculitis, marked point process, mathematical morphology

Automatic detection of the skin hyperpigmentation helps in estimating the severity of some skin diseases like *acne vulgaris* and *folliculitis*. We compared two methods for studying acne and folliculitis lesions and hyperpigmentation of the skin. We adapted a model based on Marked Point Processes and initially developed for flamingo's population counting to dermatological images of acne and folliculitis. Then, we developed an algorithm which uses mathematical morphology together with volume and shadows compensation. Finally, we compared results in term of detection accuracy.
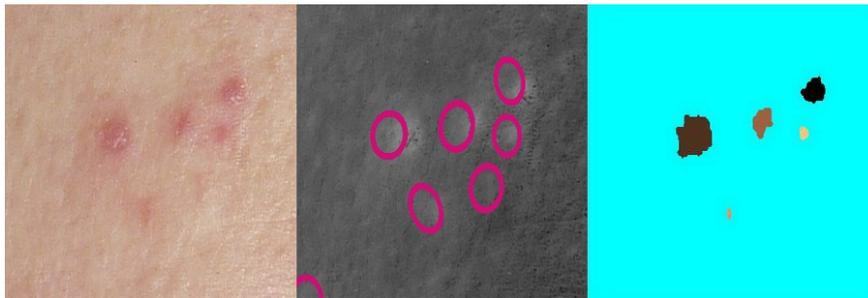
Figure 5. Source image of folliculitis (Left) and results obtained with Marked Points Processes (Middle), then with Mathematical Morphology (Right)

6.1.6. Efficient Monte Carlo sampler for detecting parametric objects in large scenes
Participants: Yannick Verdie, Florent Lafarge [contact].

This work is supervised by Florent Lafarge (Geometrica team, Inria-SAM) in collaboration with the AYIN team.

Point processes have demonstrated both efficiency and relevance when addressing object recognition problems in vision. However, simulating these mathematical models is a difficult task, especially on large scenes. Existing samplers suffer from average performances in terms of computation time and stability. We propose a new sampling procedure based on a Monte Carlo formalism. Our algorithm exploits Markovian properties of point processes to perform the sampling in parallel. This procedure is embedded into a data-driven mechanism such that the points are non-uniformly distributed in the scene. The performances of the sampler are analyzed through a set of experiments on various object recognition problems from large scenes, and through comparisons to the existing algorithms.

6.2. Statistical methods

6.2.1. Change detection on synthetic aperture radar images based on hypothesis testing
Participants: Vladimir Krylov, Josiane Zerubia [contact].

This work was conducted in collaboration with DITEN, 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).

Change detection, synthetic aperture radar, hypothesis test, likelihood ratio test, high resolution
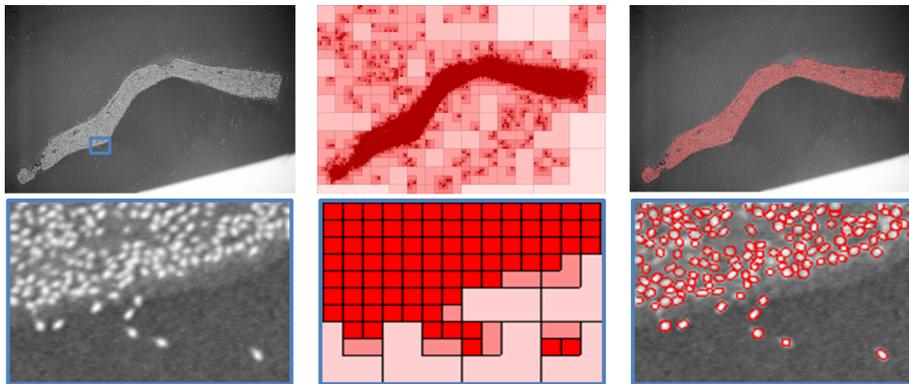
Figure 6. Bird counting by a point process of ellipses. (right) More than ten thousand birds are extracted by our algorithm in a few minutes from (left) a large scale aerial image. (middle) A quad-tree structure is used to create a non-uniform point distribution. Note, on the cropped parts, how the birds are accurately captured by ellipses in spite of the low quality of the image and the partial overlapping of birds.

Modern synthetic aperture radar (SAR) sensors represent an essential source of all-weather and 24-hour imagery with a fixed re-visit cycle at competitive high resolution. Two-date change detection from SAR images is a process that employs two SAR images acquired over the same geographical area with possibly the same (or close) acquisition characteristics at two different times to map the areas where changes occur between the two acquisition dates. The central disadvantage of the SAR imagery is given by an inherent multiplicative speckle noise, which restricts the direct application of optical-based change detection methods to SAR imagery.

We have developed a non-parametric statistical change detection approach. We avoided the ambiguity of choosing a restrictive clutter model by assuming no specific probability distribution function model [25] for the statistics of SAR. We developed a modified hypothesis test which is based on the classical Wilcoxon two-sample test that verifies whether one of two samples of independent observations tends to have larger values than the other. The choice of the Wilcoxon statistic as compared to the other available goodness-of-fit test statistics, such as, e.g., that of Cramér-von Mises' test, is a compromise solution to have simultaneously an analytically tractable asymptotic distribution (which is needed to formulate the likelihood ratio test) and a non-parametric testing procedure. Furthermore, the experimental validation demonstrated the adequacy of this statistic to the considered problem. To be able to take a decision at each pixel of the coregistered image pair we considered samples originating from the local windows centered in each pixel. Finally, we constructed a likelihood ratio test on the image with Wilcoxon statistic values. This formulation allowed to overcome the limitation of a classical independency assumption for the Wilcoxon test which is violated (at least, locally) with the local window samples. The resulting technique is related to the statistical false discovery rate approach developed for "large-scale simultaneous hypothesis testing" problems, however the derivation and interpretation are different.

Encouraging detection results were obtained on XSAR and very high resolution COSMO-SkyMed images [14].

6.2.2. *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/RampD.aspx). It was a collaboration between AYIN (Josiane Zerubia) and Morpheme (Xavier Descombes) teams.

multispectral imaging, skin, hyperpigmentation, hypothesis tests, statistical inferences
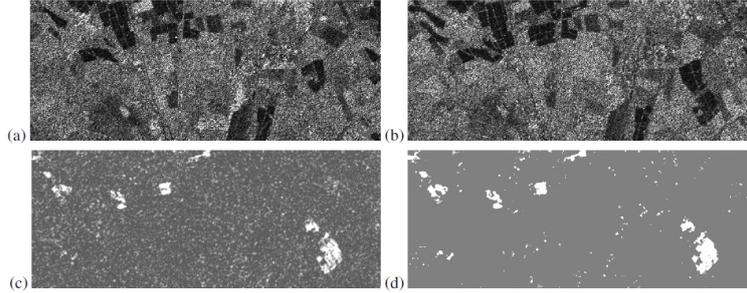
Figure 7. Coregistered XSAR images (©Univ. of Pavia) acquired on (a) April 16, 1994 and (b) April 18, 1994 and the detection results: (c) 5×5 window-based image ratio, (d) proposed method with 5×5 window.

One of the steps to evaluate the efficacy of a therapeutic solution is to test it on a clinical trial involving several populations of patients. Each population receives a studied treatment and a reference treatment for the disease. For facial hyper-pigmentation, a group of N_e patients receives the treatment on one cheek and a comparator on the other. The comparator can be a reference treatment or a placebo. 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 time t along the treatment period.

We proposed a methodology to estimate the efficacy a treatment by calculating three differential criteria: the darkness, the area and the homogeneity. The darkness measures the average intensity of the disease on a gray scaled image I obtained by a linear combination of the spectral bands of the original multi-spectral image. A differential darkness is then obtained by measuring the deviation between the initial measurement at time t_0, and the measurement at time t_k. The differential area criterion is calculated by analyzing the histogram of $I_{diff} = I_{t_0} - I_{t_k}$ a difference gray scale image between two measurements in a time series. The differential homogeneity criterion is obtained with a multi-scale analysis of I_{diff} adapted from the Statistical Parametric Mapping (SPM) methodology. Indeed, statistical inferences allow to assign a probability of change to each region of I_{diff} above a set of thresholds. These probabilities are calculated with respect to the maximum intensity and the spatial extend of each region. An integration of the obtained statistical map denoted SM, allows to get a homogeneity criterion.

The Fig. 8 illustrates the differential score calculated on a patient whose pathology decreases during the clinical trial. The proposed differential score has been tested in a full clinical study and provided results that agreed with the clinical analysis. This work have been patented, submitted to ISBI'13 conference and to the IEEE TMI journal, and published in Inria research reports [26], [27].

6.3. Hierarchical models

6.3.1. Hierarchical and graph cut-based models for multiyear sea ice floe segmentation
Participant: Yuliya Tarabalka [contact].

This work has been done in collaboration with Dr. Guillaume Charpiat (STARS team, Inria-SAM), Dr. Ludovic Brucker (NASA GSFC, USA) and Dr. James Tilton (NASA GSFC, USA).

hierarchical model, graph cut, segmentation, multiyear sea ice floes, shape analysis

(a) $I_{diff}^{t_1}$ equalised   (b) $I_{diff}^{t_2}$ equalised   (c) $I_{diff}^{t_3}$ equalised

(d) SM_{t_1} (e) SM_{t_2} (f) SM_{t_3}

(g) score t_1 (h) score t_2 (i) score t_3
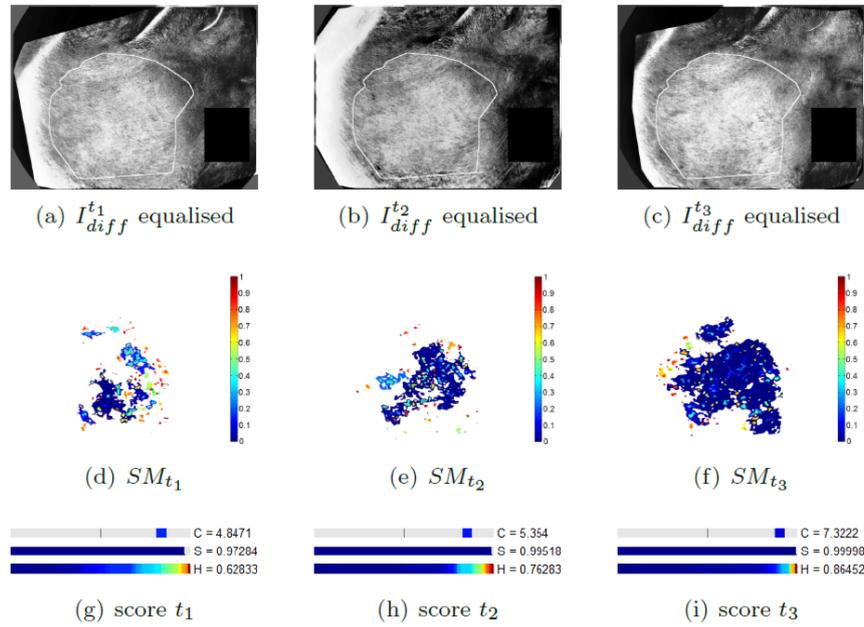
Figure 8. I_{diff}, SM and differential score for the three measurements t_1, t_2, t_3 calculated for a patient whose disease decrease.

The melting of sea ice is correlated to increases in sea surface temperature and associated climatic changes. Therefore, it is important to investigate how rapidly sea ice floes melt. We proposed two methods for segmentation of a time series of a melting sea ice floe. The first method employs hierarchical model for ice floe segmentation. Image features are extracted using morphological operators, and the floe of interest is marked based on AMSR-E satellite measurements. Then, hierarchical step-wise optimization segmentation is performed, by iteratively merging adjacent regions with the smallest dissimilarity criterion. We proposed to use area and shape parameters of the floe at two previous time moments as priors for computing a segmentation map at the next time moment.

Fig. 9 (a) depicts a graph of the multiyear ice floe area as a function of time, computed by applying the proposed hierarchical model to the summer series of Moderate-Resolution Imaging Spectroradiometer (MODIS) images. While a multiyear ice floe can only melt in the summer period, peaks on the graph correspond to segmentation errors, which are a consequence of either a cloud cover or weakness of contrast between the multiyear ice and the surrounding young ice floes. These segmentation imprecisions can be avoided by simultaneously optimizing all segmentation maps in a time series. For this purpose, we developed a new method based on graph cuts for joint segmentation of monotonously shrinking (or growing) shapes. We impose shape shrinkage (or growth, respectively) constraint in graph cuts, and minimization of energy computed on the resulting graph of the image sequence yields globally optimal segmentation. Fig. 9 (c-d) show examples of floe segmentations using the new approach. Fig. 9 (b) presents a graph of the floe area as a function of time computed by performing the proposed graph cut-based method. The results are compared to those obtained by applying graph cut segmentation to each single image in the considered time series. It can be seen that the new approach yields results with continuous shrinkage of the shape size.

6.3.2. Hierarchical model for spectral-spatial classification of hyperspectral images

Participant: Yuliya Tarabalka [contact].

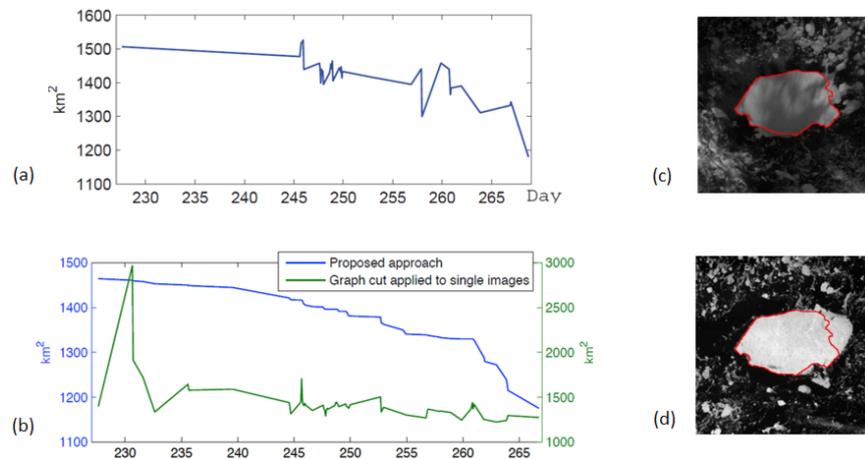*Figure 9. (a) Area of a multiyear ice floe as a function of time, computed by applying the proposed hierarchical model. (b) Area of the floe as a function of time (days), computed by using the proposed graph cut-based model (blue) and by applying graph cut segmentation to each single image in a time series. (c-d) Examples of floe segmentations using the proposed graph cut-based model.*

This work has been done in collaboration with Dr. James Tilton (NASA GSFC, USA).

hyperspectral images, classification, segmentation, geometrical features, rectangularity.

The recent advances in hyperspectral remote sensor technology makes it possible to acquire data with a very high spectral (hundreds of spectral channels) and spatial (order of a meter) resolution. The rich spectral information of the hyperspectral data leads to the potential of a more accurate classification, but also presents challenges of high-dimensional data analysis.

We developed a new method for spectral-spatial classification of hyperspectral images. The method is based on the integration of probabilistic classification and shape analysis within the hierarchical step-wise optimization algorithm. First, probabilistic support vector machines classification is applied. Then, at each iteration two neighboring regions with the smallest dissimilarity criterion are merged, and classification probabilities are recomputed. We proposed to estimate a dissimilarity criterion between regions as a function of statistical, classification and geometrical (area and rectangularity) features. Fig. 10 shows the obtained classification results on a 102-band ROSIS image of the Center of Pavia, Italy, which are compared with Support Vector Machines (SVM) classification results. These results did show that the proposed method succeeded in taking advantage of both spatial and spectral information for accurate hyperspectral image classification.

6.3.3. Classification of combined hyperspectral and panchromatic data using spectral-spatial approaches

Participants: Yuliya Tarabalka [contact], Josiane Zerubia.

This work has been conducted in collaboration with the French Space Agency CNES (http://www.cnes.fr/), with Dr. Marie-José Lefèvre, Dr. Hélène DeBoissezon and Mr. Manuel Grizonnet.

hyperspectral data, HYPXIM, data fusion, panchromatic image, segmentation

Hyperspectral imaging records a detailed spectrum for each pixel, opening new perspectives in classification. Currently, several hyperspectral satellite missions such as EnMAP (210 bands, GSR 30m) are under development. The future hyperspectral satellite missions PRISMA and HYPXIM also include a panchromatic channel

Figure 10. Center of Pavia image. (Left) Three-band color composite. (Middle) SVM classification map, overall classification accuracy = 94.9%. (Right) Classification map obtained using the proposed hierarchical approach, overall classification accuracy = 97.1%.

with better spatial resolution. We explored if a panchromatic channel at a higher spatial resolution (factor 4) contributes for more accurate classification of hyperspectral images in space conditions.

We adapted and compared several classification methods for combined hyperspectral and panchromatic images, and conducted experiments on the simulated HYPXIM data provided by CNES. We fused both data sources using principal component and Gram-Schmidt fusion methods, as well as the vector stacking approach. We then applied Support Vector Machines (SVM) classification on the resulting feature sets. Furthermore, we considered spatial information for more accurate classification by: (1) including Haralick's texture features in the feature set; (2) segmenting an image into homogeneous regions using a Hierarchical Step-Wise Optimization (HSWO) technique, and assigning each segmented region to the dominant class within this region.

Classification results are illustrated in Fig. 11 . We concluded that classification accuracies of the HYPXIM simulated data have been improved when including a panchromatic channel at a higher spatial resolution into a classification system. These results are close to hyperspectral aerial data classifications. For the image containing one-pixel regions and mixed pixels, standard spectral-spatial classification methods are not well adapted and thus do not improve accuracies when compared to pixelwise classification. In the future, we plan to develop methods which would use both spatial information and a spectral unmixing concept for efficient fusion of hyperspectral and panchromatic data.

6.4. Other detection approaches

6.4.1. *Multiple-instance object detection using a higher-order active shape prior*

Participants: Ikhlef Bechar, Josiane Zerubia [contact].

This work is done in collaboration with Dr. Ian Jermyn of Durham University (United Kingdom, https:// www.dur.ac.uk/mathematical.sciences/) and was funded by a contract with the EADS foundation (http://www. fondation.eads.com/).

object detection, shape prior, transformation invariance, higher-order active contours, energy minimization, non-convex energy, exact convex relaxation.
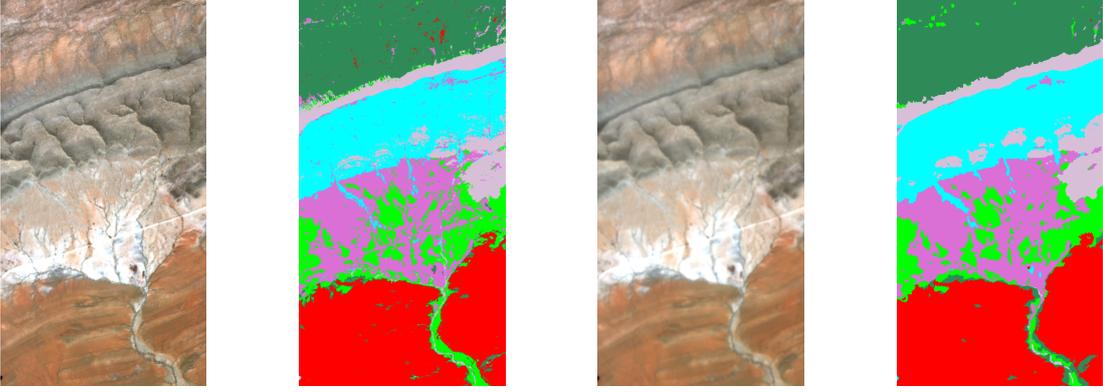
Figure 11. From left to right (a-d): (a) The false-color HYMAP aerial image (126 bands, GSR 4.8m). (b) SVM classification map for the HYMAP image, overall classification accuracy = 83%. (c) Simulated HYPXIM image (126 bands, GSR 14.4m). (d) Classification map of the fused by vector stacking hyperspectral image (c) and panchromatic image at GSR 4.8m, using HSWO-based spatial regularization, overall classification accuracy = 80.7%.

The problem under consideration is the multiple-instance object detection from imagery using prior shape knowledge. As mathematical and algorithmic framework, we have used the higher-order active contour (HOAC) model framework in order to incoporate prior shape knowledge about a class of objects of interest. On top of its robustness and its computational attractiveness (due to its parameter-estimation free method), the HOAC object-detection framework allows to incorporate shape knowledge about multiple occurrences of an object of interest in an image and to carry out object detection in a single algorithmic framework via the minimization of energy of the form:

$$\min_{\text{over all shapes} \gamma} E(\gamma) = E_{image}(\gamma) + E_{prior}(\gamma) \tag{1}$$

where γ stands for the contour an image object, $E_{image}(\gamma)$ stands for its image-based energy and $E_{prior}(\gamma)$ stands for a prior energy which is only a function of an objet's shape (and not of image data). The goal of this project is thus to model $E_{prior}(\gamma)$ using the HOAC methodology.

In this work, we have developed a fourth-order active contour (FOAC) framework for incorporating prior shape knowledge about target shapes. Typically, we express a FOAC energy model as

$$E_{foac}(\gamma) = \lambda_C L(\gamma) + \alpha_C A(\gamma) + \beta_C \int \int \int \int \langle \dot{\gamma}_p, \dot{\gamma}_q \rangle \mathbf{K}_C \left(|\gamma_p - \gamma_q|, |\gamma_s - \gamma_t| \right) \langle \dot{\gamma}_s, \dot{\gamma}_t \rangle \, dp \, dq \, ds \, dt \tag{2}$$

where $L(\gamma)$ and $A(\gamma)$ stand respectively for the length and the area of a contour γ and $\int \int \int \int \langle \dot{\gamma}_p, \dot{\gamma}_q \rangle \mathbf{K}_C \left(|\gamma_p - \gamma_q|, |\gamma_s - \gamma_t| \right) \langle \dot{\gamma}_s, \dot{\gamma}_t \rangle \, dp \, dq \, ds \, dt$ models fourth-order interactions between quadruples of contour points, and λ_C, α_C and β_C stand for some tradeoff parameters that control the contribution of each term of the FOAC energy. Note that the parameters of the method include both the real coefficients λ_C, α_C and β_C and the bivariate kernel $\mathbf{K}_C(u, v)$; $\forall\, u, v \in \mathbf{R}^+$. These parameters need to be tuned optimally for a given target shape γ^*. Thus we have developed a direct method for the optimal estimation of the FOAC parameters.

We have then shown that shapes with arbitrary geometric complexity can be modeled using such the FOAC framework 2 , and we have developed a direct method for the estimation of the parameters for a given class of shapes. In order to be able to detect multiple occurrences of a target object in an image, one needs to re-express such an originally contour-based energy 2 by replacing appropriately in formula 2 the one-dimensional contour quantity γ with an equivalent two-dimensional quantity (ie. with respect to the image domain) such as the characteristic function of γ and to minimize with respect to it the resulting energy functional. This allows topological changes of an evolving contour and hence the detection of possible multiple instances of a target object in an image. We have shown that such a new formalism is a third-order Markov Random Field (MRF) which practical optimization was a challenging question. Therefore, we have also developed a method for the exact minimization of the energy of the resulting MRF model (using a equivalent convex-relaxation approach, see Fig. 12).
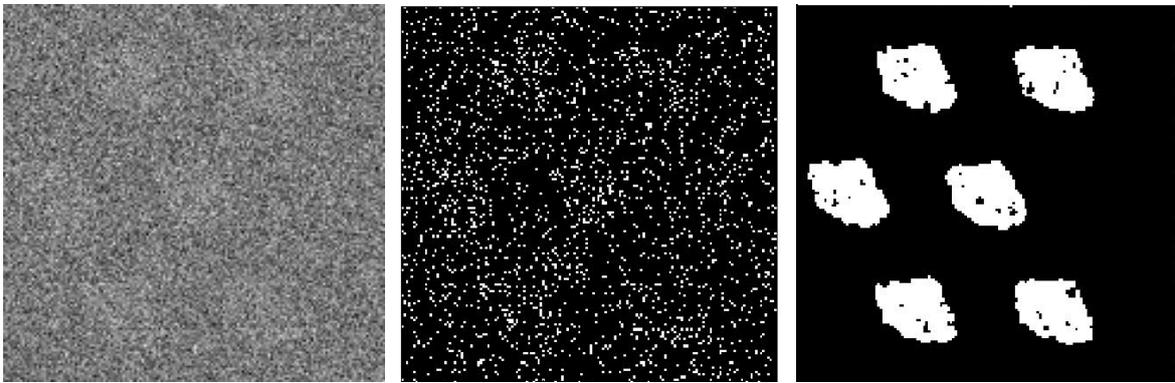
Figure 12. (Left) A very noisy input image; Multiple-instance object detection using: (Middle) a traditional segmentation method (Mumford-Shah model with a TV-based regularization) alone; (Right) with a FOAC shape prior.

6.4.2. *Image analysis for automatic facial acne detection and evaluation*

Participants: Zhao Liu, Josiane Zerubia [contact].

This work is part of LIRA Skin Care Project, which includes four key partners: Philips R&D (Netherlands, http://www.philips.nl), CWI (Netherlands, http://www.cwi.nl/), Fraunhofer Institutes (Germany, http://www. fraunhofer.de/en.html) and Inria (France).

image processing, feature extraction, pigmentation distributions, acne, cosmetology

Acne vulgaris is a highly prevalent skin disease, which has a significant life quality impact on sufferers. Although acne severity is readily observed by the human eye, it is an extremely challenging task to relate this visual inspection to measurable quantities of various skin tones and types. So far there is no golden standard for acne diagnosis in clinics, and it entirely depends on dermatologists' experience for evaluation of acne severity. But significant inter-rater variability among individual assessment may lead to less trustworthy diagnosis when several clinicians get involved in the study. In addition, less reproducibility of human evaluation makes comparison of acne changes over time difficult. Therefore, the long-term objective of this study is to construct an automatic acne grading system through applying spectroscopy imaging techniques and image processing methods, to objectively evaluate severity of skin disorder. Such a computer-based tool would also significantly benefit the development of better skin care products, if it can reliably characterize treatment effects of products in individual skin layers in agreement with physiological understanding.

Acne segmentation is normally considered as the first significant step in an automatic acne grading system, because segmentation accuracy directly influences the definition of acne pigmentation level, what has an impact on the goodness of acne severity evaluation. An initial unsupervised segmentation method is proposed for conventional RGB images, whose process is demonstrated in Fig. 13 (a). After several pre-processing steps (background and skin hair removal, illumination corrections), nine pigmentation descriptors were extracted from three RGB channels based on colorimetric transformations and absorption spectroscopy of major chromophores. It has been proved that the derived hemoglobin, normalized red, and normalized green descriptors can properly characterize pigmentation distributions of acne, and they are used as segmentation features. Finally, an iterative unsupervised segmentation was performed to maximize pigmentation distributions between acne and normal skin. Fig. 13 (b) shows an example of acne image on human face captured by a conventional RGB camera, while experimental result in Fig. 13 (f) illustrates that suspicious acne areas and healthy human skin can be automatically discriminated by applying the proposed method. Moreover, it only takes 90.8 seconds to segment the example image with the size of 640×428 pixels, which demonstrates the computation efficiency of the algorithm.

It should be noted that the segmentation method stated above is an initial approach. Shadows around non-flatten areas on human face (e.g. areas around nose) have a large influence on accuracy of automatic acne detection. However, based on the initial experimental results, it is difficult to entirely get rid of these effects using RGB channels only. Our finding is actually consistent with the existing studies, where researchers divided human face into several sub-regions and worked on these sub-regions individually to avoid shadow influence. Therefore, the next step study will compare acne segmentation results derived from RGB images and multi- or hyperspectral images, to investigate the most effective bands for describing acne pigmentation, as well as whether the introduction of multi- or hyperspectral analysis to the automatic acne detection and evaluation is necessary.

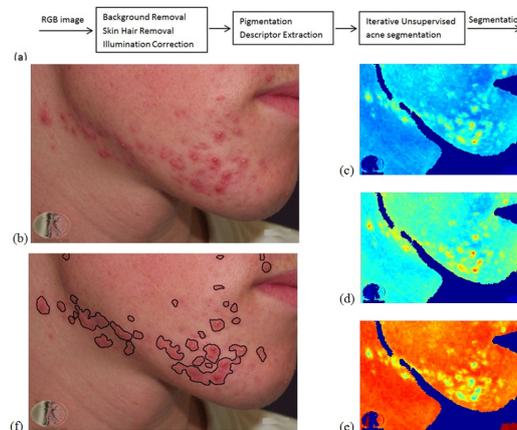*Figure 13. An initial method and result of automatic acne detection. (a) Specific steps in the acne segmentation method. (b) An example of acne disorder on human face (640×428 pixels) from DermnetNZ (http://www.dermnetnz.org/). (c)-(e) pigmentation descriptors: hemoglobin, normalized red, and normalized green, respectively. (f) Segmentation result outlined with black borders on original image.*

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

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, Francis Mairet, Rafael Muñoz-Tamayo, Elsa Rousseau.

6.1.1.1. Mathematical study of semi-discrete models

Semi-discrete models have shown their relevance in the modeling of biological phenomena whose nature presents abrupt changes over the course of their evolution [95]. We used such models and analysed their properties in several situations that are developed in 6.2.3 , most of them requiring such a modeling 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 [59] or in the co-existence analysis of epidemiological strains [19], [50]. Seasonality also plays a big role in the semi-discrete modeling required for the analysis of consumers' adaptive behavior in seasonal consumer-resource dynamics, where only dormant offspring survives the 'winter' [52].

6.1.1.2. 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 [30], 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.

The competition of several microalgal species was also studied in order to determine conditions that may give a competitive advantage to a species of interest. We study the competition for two species subject to photoinhibition at high light. This leads to a closed loop control strategy based on the regulation of the light intensity at the bottom of the reactor. The winning species is the one with the highest growth rate at high light. Then we show that the proposed controller allows the selection of a species of interest among n species [102].

6.1.2. *Model design, identification and validation*

Participants: Olivier Bernard, Francis Mairet.

One of the main families of biological systems that we have studied involves mass transfer between compartments, whether these compartments are microorganisms or chemical species in a bioreactor, or species populations in an ecosystem. We have developed methods to estimate the models of such systems [2]. These systems can be represented by models having the general structure popularized by [69], [74], and based on an underlying reaction network:

$$\frac{d\xi}{dt} = K\,r(\xi, \psi) + D(\xi_{in} - \xi) - Q(\xi)$$

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, among which activated sludge processes [68], anaerobic digestion [87], [114] and anaerobic digestion of microalgae [20]. Recently it was also used to reduce the ADM1 model in the case of winery effluent wastewater [88].

6.1.3. *Nonlinear observers*

Participants: Jean-Luc Gouzé, Olivier Bernard, Francis Mairet.

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 [7]. 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 [105]. This approach was then extended to n-dimensional linear systems, leading to the design of interval observers in high dimensions [106]. Interval observers for non linear triangular systems satisfying Input to State Stability has been proposed [22]. Extension to time-delay systems have also been proposed [23]. The efficiency of the interval observer design, even with chaotic systems has been developed and applied considering parameters uncertainties of the system and biased output [108], [105].

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" [73]. These algorithms have been improved by the estimation of the observer gain providing the best estimate [40], [21]. The approach has been applied to estimation of the microalgae growth and lipid production [101].

These works are done in collaboration with Frédéric Mazenc (DISCO, Inria) and Marcelo Moisan (EMEL S.A., Chile).

6.1.4. *Metabolic and genomic models*

Participants: Jean-Luc Gouzé, Madalena Chaves, Alfonso Carta, Ismail Belgacem, Xiao Dong Li, Olivier Bernard, Wassim Abou-Jaoudé, Luis Casaccia, Caroline Baroukh, Rafael Muñoz-Tamayo, Jean-Philippe Steyer.

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. 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].

Global stability for metabolic models and unreduced Michaelis-Menten equations

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 types of metabolic models 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 behavior in a more general genetic-metabolic loop [26]. Moreover, with the same techniques, we studied full (i.e. not reduced by any time-scale argument) Michelis-Menten reactions or chains of reactions: we prove global stability when the equilibrium

exists, and show that it may not exist. This fact has important consequences for reduction of metabolic systems in a coupled genetic-metabolic system [34], [45], [70].

Interconnections of Boolean modules: asymptotic and transient behavior

A biological network can be schematically described as an input/output Boolean module: that is, both the states, the outputs, and the inputs are Boolean. The dynamics of a Boolean network can be represented by an asynchronous transition graph, whose attractors describe the system's asymptotic behavior. 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 semi-attractors (states of the attractor with same output) of each module. In [82], the *asymptotic graph* was proposed, which is quite fast to compute and identifies all attractors of the interconnected system, but may also generate some spurious attractors. In [31] the *cross graph* is proposed, which exactly identifies the attractors of the interconnected system but is slower to compute. The asymptotic dynamics of high-dimensional biological networks can thus be predicted through the computation of the dynamics of two isolated smaller subnetworks. An application is, for instance, to interconnect four individual "cells" to obtain all the attractors of the segment polarity genes model in *Drosophila*.

Probabilistic approach for predicting periodic orbits in piecewise affine differential models

The state space of a piecewise affine system is partitioned into hyperrectangles which can be represented as nodes in a directed graph, so that the system's trajectories follow a path in a transition graph. Using this property we defined a *transition probability* between two nodes A and B of the graph, based on the volume of the initial conditions on the hyperrectangle A whose trajectories cross to B [15]. The parameters of the system can thus be compared to the observed or experimental transitions between two hyperrectangles. This definition is useful to identify sets of parameters for which the system yields a desired periodic orbit with a high probability, or to predict the most likely periodic orbit given a set of parameters, as illustrated by a gene regulatory system composed of two intertwined negative loops.

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 [35].

E. coli modeling and control

In the framework of ANR project Gemco, we developed and analyzed a model of a minimal synthetic gene circuit, that describes part of the gene expression machinery in *Escherichia coli*, and enables the control of the growth rate of the cells during the exponential phase.

This model is a piecewise non-linear system with two variables (the concentrations of two gene products) and an input (an inducer). We studied the qualitative dynamics of the model and the bifurcation diagram with respect to the input. Moreover, an analytic expression of the growth rate during the exponential phase as function of the input has been derived. A relevant problem was that of parameters identifiability of this expression supposing noisy measurements of exponential growth rate. We presented such an identifiability study that we validated in silico with synthetic measurements [36].

We also studied a model of the global cellular machinery designed by D. Ropers and collaborators (IBIS team, Grenoble). This model has 11 variables and many parameters ; we explored different techniques for reduction and simplification [56], [57].

Transition graph and dynamical behavior of piecewise affine systems

We investigated the links between the topology of the transition graph and the number and stability of limit cycles in a class of two-dimensional piecewise affine biological models. To derive these structure-to-dynamics principles, we use the properties of continuity, monotonicity and concavity of Poincare maps associated with transition cycles of the transition graph [64].

Robust estimation for a hybrid model of genetic networks

State estimation problems with Boolean measurements for a classical negative loop genetic network governed by a piecewise affine (PWA) model have been studied in [39]. Observers are proposed for the cases where either full state or only partial state Boolean measurements are available. In the first case, sliding modes may occur, which leads to finite time convergence for the observer. In the second case, an algebraic computation is proposed to solve the initial condition inverse problem. The robustness of the observer for a parametric uncertain model is investigated, and we show that the error bound is proportional to the magnitude of the uncertainty.

Modeling the metabolic network in non balanced growth conditions

We have developed a new approach to represent the metabolic network of organisms for which the hypothesis of balanced growth is not satisfied [67]. This is especially true for microalgae which store carbon during the day and nitrogen during the night [44]. The proposed formalism is based on the assumption that some parts of the metabolic network satisfy the balance growth conditions, *i.e.* there is no accumulation of intermediate compounds. This hypothesis specifically applies to the main functions in the cell (respiration, photophosphorylation,...). Between two functions, some compounds can accumulate with storage/reuse kinetics. The resulting system is thus a slow-fast system.

6.2. Fields of application

6.2.1. Bioenergy

6.2.1.1. Modeling of microalgae production

Participants: Olivier Bernard, Antoine Sciandra, Frédéric Grognard, Philipp Hartmann, Rafael Muñoz-Tamayo, Ghjuvan Grimaud, Charlotte Combe, Hubert Bonnefond, Jean-Philippe Steyer, Francis Mairet.

Experimental developments

Experiments have been carried out to study the effects of nitrogen limitation on the lipid production in microalgae [18], [17], [48] and support model development. These experiments have been carried out in the Lagrangian simulator (SEMPO), under constant or periodic light and temperature, varying the total amount of light dose in the day. The response in terms of storage carbon (triglycerides and carbohydrates) has been observed.

Other experiments were carried out to reproduce the light percept by a cell in a raceway [58]. An electronic platform was developed to reproduce the flashing light which, from the hydrodynamical studies, is likely to happen in a raceway at the cell scale. The experiments show that the microalgae adapt their pigments to the average light that they have received.

The effect in the cell cycle of both the light periodic signal and a nitrogen limitation were studied. The strong interactions of the interactions between the different phases of the cell cycle through checkpoints was highlighted [24].

This work is done in collaboration with Amélie Talec, Thomas Lacour, and Christophe Mocquet (CNRS-Océanographic Laboratory of Villefranche-sur-Mer).

Modeling the effect of temperature

The effect of temperature on microalgae has been represented by adapting the CTMI model developed for bacteria [115]. The proposed model [14], [28], was able to correctly represent the growth response to temperature for 15 different species. A procedure for model calibration and estimation of the parameter uncertainties was specially developed, allowing to gather experimental data from various sources. It was shown that different strains of the same species have a very similar response to temperature fluctuations. Moreover, for low light intensities, a simple model can represent both effects of light and temperature [14].

Modeling light distribution within a photobioreactor

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.

Modeling lipid accumulation

We have proposed a new model for lipid production by microalgae which describes the fate of the CO_2 incorporated during photosynthesis [10]. 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. 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 [96].

Modeling a microalgae production process

The integration of different models developed in the group [72], [96], [10],[14] was performed to represent the dynamics of microalgae growth and lipid production in raceway systems, on the basis of the dynamical model developed to describe microalgal growth in a photobioreactor under light and nitrogen limitations [72]. 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 ...), temperature effect and the radiative transfer in the culture (light attenuation due to the microalgae).

Using these approaches, we have developed a model which predicts lipid production in raceway systems under varying light, nutrients and temperature [109], [110]. This model is used to predict lipid production in the perspective of large scale biofuel production.

6.2.1.2. Coupling growth of microalgae with hydrodynamics
Participants: Olivier Bernard, Antoine Sciandra, Philipp Hartmann, Charlotte Combe.

Modeling the coupling between hydrodynamics and biology

In collaboration with the Inria ANGE team, a model coupling the hydrodynamics of the raceway (based on multilayer Saint Venant system) with microalgae growth was developed [13]. This model is supported by the work of ANGE aiming at reproducing the hydrodynamics of the raceway, with a specific attention to the effect of the paddle wheel on the fluid.

Modeling the photosynthesis response to fast fluctuating light

The impact of the hydrodynamics on the light percept by a single cell was studied thanks to fluid dynamics simulations of a raceway pond [37] [92]. 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. A new model accounting for photoacclimation was also proposed [93].

6.2.1.3. Optimization of microalgae production
Participants: Olivier Bernard, Antoine Sciandra, Frédéric Grognard, Rafael Muñoz-Tamayo.

Numerical optimization

Using the detailed model for raceway systems, we assessed strategies for optimal operation in continuous mode [109]. Two strategies were developed. The first one resides in solving numerically an optimal control problem in which the input flow rate of the raceway is calculated such that the productivity in microalgae biomass is maximized on a finite time horizon. In the second strategy, we aimed at translating the optimization problem into a regulation problem. We proposed a simple operational criterion that when integrated in a strategy of closed-loop control allows to reach biomass productivities very near to the productivities obtained with the optimal control. We demonstrated that the practical advantages for real implementation makes our proposed controller a suitable control strategy for optimizing microalgae production in raceways.

Analytical optimization

Optimization strategies were based on simple microalgae models : first, biomass production has been optimized in a constant light environment [104], 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 Pontryiagin's maximum principle (assuming a periodic working mode) [61] [90].

6.2.2. CO_2 fixation by microalgae

Participants: Olivier Bernard, Antoine Sciandra, Ghjuvan Grimaud.

Experimental work

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 [24].

Modeling cell cycle

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 [107]. 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 [107].

Calcification of coccolithophorids

The effect of CO_2 partial pressure increase on photosynthesis and calcification of the calcareous microalgae *Emiliania huxleyi* have been experimentally observed. It results in an increase of the coccolith size together with a decrease in the calcification rate [25].

Three models accounting for the possible coupling between photosynthesis and calcification [75] were included in an ocean model; they account for settling and predation by grazers, and a bloom of coccolithophorids was simulated [76], [77].

Nitrogen fixation by nitrogenotrophs

The fixation of nitrogen by *Croccosphera watsonii* was represented with a macro metabolic model [60] [89]. The main fluxes of carbon and nitrogen are represented in the cell. The accumulation of starch during the day to fuel the nitrogenase working in the absence of oxygen during the night was the key process to explain the nitrogen fixation. The strong influence of the cell cycle was also included in the model. Finally, the model was calibrated and validated with the data of 3 experiments carried out with different duration of the light period and daily dose. The model succeeded to efficiently reproduce the experimental data.

This work is done in collaboration with Sophie Rabouille (CNRS-Océanographic Laboratory of Villefranche-sur-Mer).

Including phytoplankton photoadaptation into biogeochemical models

The complexity of the marine ecosystem models and the representation of biological processes, such as photoadaptation, remain an open question. We compared several marine ecosystem models with increasing complexity in the phytoplankton physiology representation in order to assess the consequences of the complexity of photoadaptation models 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 (Joint Global Ocean Flux Study) data. It turns out that the more complex models are trickier to calibrate and that intermediate complexity models, with an adapted calibration procedure, have a better prediction capability [12], [43], [42].

This work is done in collaboration with Sakina Ayata (UPMC-Océanographic Laboratory of Villefranche-sur-Mer).

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, Nicolas Bajeux.

Optimization of biological control agent introductions

The question of how many and how frequently natural enemies should be introduced into crops to most efficiently fight a pest species is an important issue of integrated pest management. The topic of natural enemies introductions optimization has been investigated for several years [9] [111]. It had allowed to unveil the crucial influence of within-predator density dependent processes, and especially negative density dependence. In particular, we concluded that pest control is more efficiently achieved through the frequent introduction of small populations of natural enemies as compared to larger and rarer ones. Because contrarily to predatory biocontrol agents, parasitoids may be more prone to exhibit positive density dependent dynamics rather than negative ones, the current modeling effort concentrates on studying the impact of positive predator-predator interactions on the optimal introduction strategies [55].

Connected experimental research is also being pursued in the laboratory on *trichogramma spp.* which tends to show positive density dependence because of demographic stochasticity [32], and the PhD thesis of Thibaut Morel Journel (UMR ISA) has just started on this topic.

Food source diversity and classical biological control efficiency using generalist natural enemies

Because generalist biocontrol agents can feed on different food sources like, e.g. a given pest and pollen, they are capable of surviving pest absence within crops [118]. From the biological control point of view, this makes it possible to sustain natural enemies populations able to fight pests at the onset of pest attacks. Moreover, when supplied with different food types, generalists organisms are expected to thrive. Alternative prey, banker plants or more generally habitat enhancement based biological control strategies are thus becoming popular IPM (Integrated Pest Management) methods [112]. Although it has clear advantages, the simultaneous presence of various food sources also has important drawbacks: feeding on different food sources means that a given individual cannot feed on each food source at the same moment. This distraction effect thus potentially reduces the overall predation pressure imposed by the natural enemy population, and the interaction between the demographic response of the predator population and individual behavior is complex. To investigate such questions, we developed and analyzed behavioral-demographic population models taking into account the negative density dependent character of most generalist biocontrol agents. We found out that predator distraction effects can dominate the demographic response of the predator populations, potentially disrupting pest control [120]. An additional conclusion of our study, is that higher predator densities can actually bring about lesser pest suppression. Such results question current biological control practices, and show that, counter-intuitively, recording a lot of predators within fields does not ensure efficient pest control.

Plant compensation, pest control and plant-pest dynamics

Plant compensation is the process by which plants respond positively to recover from the effects of pest injury on plant growth. It is a common phenomenon, which has been repeatedly reported in various plant taxa during the last thirty years. Of special interest is the overcompensation phenomenon: consecutively to a pest attack, a plant may reach a higher biomass or have a better fitness compared to the no-pest-attack situation [65]. Although this phenomenon has mainly been documented in wild plants [65] it has also been observed on agricultural plants [121], [113]. To understand better this plant-herbivore interaction and to assess the efficacy of different pest control strategies we built a plant-pest model of plant compensatory growth. We have shown that depending on plants and pests characteristics, plant overcompensation may or may not happen. Moreover, because the model undergoes a backward bifurcation, it is shown that plant overcompensation is also dependent on the level of pest attacks and does not necessarily show up even when the plant-pest couple do have the potential to produce overcompensation [38].

This work is part of the PhD thesis of Audrey Lebon (Cirad), and done in collaboration with Yves Dumont (Cirad).

6.2.3.2. Controlling plant pathogens
Participants: Frédéric Grognard, Ludovic Mailleret, Elsa Rousseau.

Sustainable management of plant resistance

The introduction of plant 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. 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. We showed that pathogen's fitness cost associated with resistance breakdown was one of the main factors governing damage reduction at the landscape scale, although the optimal resistance deployment also strongly relied on epidemic characteristics and landscape coonnectivity [16], [51]. The capacity for a virulent virus to establish itself in such an environment, as well as the evolution of the virus characteristics have been studied [59].

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

Eco-evolutionary dynamics of plant pathogens in seasonal environments

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. Most agricultural systems in temperate environments are characterized by the cyclical presence and absence of the crop, due to cropping practices such as harvest and planting. The seasonal character of agrosystems can induce complex plant-pathogens dynamics [19] and is an important force promoting evolutionary diversification of plant pathogens [91]. Plant parasites reproduction mode may strongly interact with seasonality. In this context, we investigated the influence of cyclical parthenogenesis, i.e. the alternation of sexual and asexual reproduction phases, on the eco-evolutionary dynamics of plant parasites [80]. By means of a theoretical approach, we show that an obligate sexual event prior to overseasoning promotes evolutionary divergence in terms of investment into asexual reproduction in plant parasites. Yet, polymorphism may be transient; namely, morphs mostly investing into sexual reproduction may eventually exclude morphs mostly investing into asexual reproduction. Our findings nicely echo with recent population genetics results on *Leptosphaeria maculans*, the causal agent of the blackleg disease of canola, reporting differential investments into sexual and asexual reproduction both at the global and continental scales.

This work is part of the PhD thesis of Magda Castel (Agrocampus Ouest) and is done in collaboration with Frédéric Hamelin (Agrocampus Ouest).

6.2.4. Biological depollution - Anaerobic digestion

6.2.4.1. Coupling microalgae to anaerobic digestion
Participants: Olivier Bernard, Antoine Sciandra, 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 [117], [116].

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 optimization [97], [79], [20]. 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 [119] (a reference model which considers 19 biochemical reactions) to represent the same dataset. This model, after modification of the hydrolysis step [99], [100], [98] has then been used to evaluate process performances (methane yield, productivity...) and stability though numerical simulations.

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 [83].

An analysis of the potential environmental impacts of biodiesel production from microalgae has been carried out using the life cycle assessment (LCA) methodology [94]. 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 electricty is produced on the plant [86], [85]. 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 [84] 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. Optimality in consumer-resource dynamics
Participants: Frédéric Grognard, Ludovic Mailleret, Pierre Bernhard.

Adaptive behavior in seasonal consumer-resource dynamics

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 prey (choose their behavior) according to the solution of an optimal control problem [66]. We then examined how (resident) predators adopting this optimal behavior 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 behavior. We studied the resulting optimal control problem where the mutants maximize their own number of offspring using the knowledge of the resident's behavior, 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 behavior [52].

Optimal foraging and residence times variations

Charnov's marginal value theorem (MVT) [81] is a central tenet of ecological theory. In fragmented environments, the MVT connects the quality and distribution of patches to the optimal time an individual should spend on any patch, and thus the rate of movement in the habitat. Unfortunately, it does not offer explicit predictions regarding how changing habitat quality would affect residence times. In this work, we answer that question in a very general setting, for habitats with homogeneous or heterogeneous patches and with general fitness functions. We then particularize it to the resource consumption framework and indicate how the residence times variations relate to the curvatures of the functional responses [49], [78].

This last work is done in collaboration with Vincent Calcagno and Eric Wajnberg (INRA Sophia Antipolis)

6.2.5.2. Growth models of zooplankton
Participants: Jean-Luc Gouzé, Jonathan Rault, Eric Benoît.

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

5. New Results

5.1. Simulations in plasma Physics

5.1.1. *Fourier-spectral element approximation of a two fluid model of edge plasma*

Participants: Richard Pasquetti, Sebastian Minjeaud.

We especially work on a two fluid physical model developed in close connection with Ph. Ghendrih (IRFM). It is based on the electrostatic assumption, i.e. the magnetic field is given (the magnetic field induced by the plasma itself is negligible), and on the hypothesis of electroneutrality (the density of ions and electrons are proportional). On the basis of the conservation equations of density, electron and ion velocities, electron and ion temperatures and electrical charges, a set of 10 non-linear coupled partial differential equations (PDE) can be set up. A high order Fourier-SEM (Spectral Element Method) code is currently developed. This Fourier-SEM code is now operational for the full set of PDEs in a 3D toroidal geometry. The torus section is discretized with quadrangular elements, within which the polynomial approximation degree is an input to the code. In time one uses an RK3 (third order Runge-Kutta) IMEX (Implicit-Explicit), so that the Lorentz terms are handled implicitly. The capability of this code to handle a strongly anisropic diffusion in a 3D toroidal geometry has already been tested. The Braginskii closure has been implemented. The Bohm boundary conditions at the plates are also considered. A parallel version of this code is currently developed. It remains to improve the robustness of our algorithms, i.e. to implement an efficient stabilization strategy. This could be based on the so-called spectral vanishing viscosity or entropy viscosity techniques. Up to our knowledge, this will be the first code that fully implements a two fluid ion-electron approximation (i.e. without using the drift velocity approximation), and the Braginskii closure of the governing equations.

5.1.2. *Hydrodynamic model with strong Lorentz force*

Participants: Audrey Bonnement, Hervé Guillard, Boniface Nkonga, Richard Pasquetti.

The thesis of A. Bonnement [1] was devoted to the development of a code based on the FluidBox/plaTo software of B. Nkonga and co-workers. It is based on a Finite volume / Finite element approach. This code is now operational in an axisymmetric geometry for a simplified PDE system in which the Lorentz force is approximated by a constant forcing field. Thus, the FluidBox/PlaTo code essentially solves the 3D axisymmetric Euler, Navier-Stokes or Braginskii PDEs to compute the ion density, momentum and energy. In the Braginskii system, the thermal diffusion and the kinematic viscosity are both non-linear and strongly anisotropic. A. Bonnement, who was co-directed by H. Guillard and R. Pasquetti, defended her thesis "Modélisation numérique bi-fluide du plasma de bord des tokamaks: application à ITER" in July 2012. A. Bonnement has provided a detailed description of the works carried out with the FluidBox/PlaTo code in her thesis manuscript. She has specially addressed one of the main difficulties related to simulations of tokamak plasmas, which is that the dynamic of the flows occurs in the vicinity of an equilibrium where the plasma pressure balances the Lorentz force. There are two ways to deal with this difficulty. The most common one in tokamak studies is to work with a set of governing equations such that this equilibrium is already contained in the formulation. This can be done by using formulations where the variables are indeed fluctuating departures from the equilibrium or by using special approximations as done in reduced MHD. The other way is purely numerical and consists to design a numerical method such that the equilibrium is an exact solution of the discrete equations. This has been the subject of the thesis of Audrey Bonnement in the framework of a finite volume method on non-structured meshes and where special Riemann solvers have been designed incorporating plasma equilibrium in the definition of the numerical fluxes. Combined with mesh refinement, this approach has been applied to some preliminary numerical experiments studying the effect of density perturbations (as a crude model of pellet injections) on the dynamics of the flow. At present, this approach is under evaluation to qualify its interest with respect to reduced MHD or formulations using a potential representation of the velocity field.

5.1.3. *Finite volume methods in curvilinear system of coordinates*

Participants: Hervé Guillard, Boniface Nkonga, Afeintou Sangam, Marco Bilanceri.

Finite volume methods are specialized techniques to approximate systems of conservation laws. The application of these methods to curvilinear systems of coordinate is however problematic because the space variation of the metric coefficients introduces artificial source terms. However it can be shown that whatever the curvilinear system used, a strong conservation form of the equations exists at the level of vector variables (but not at the level of the scalar components of the vectors in the curvilinear system due to the aforementioned space dependence of the metric coefficients). Based on this result, we have developed an original technique that uses an approximation of the vector form of the equation followed by local projection on the curvilinear system (here parallel to the poloidal magnetic field).
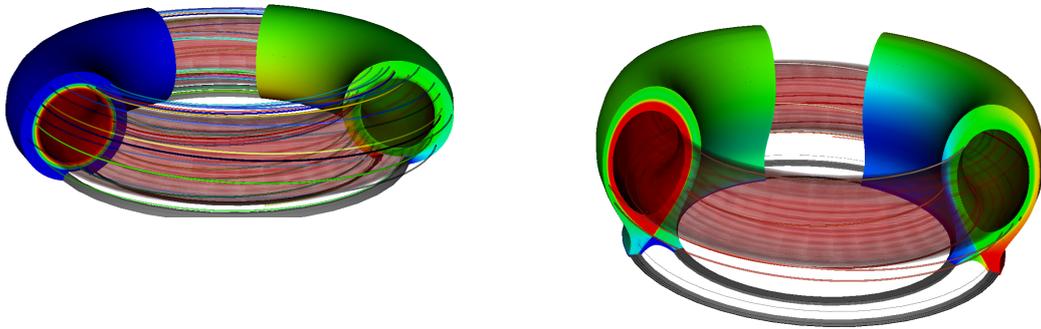
Figure 1. Density (left side) and parallel velocity (right side) color plots of the edge region of a tokamak with limiter (left plot) and tokamak with X point and divertor(right plot). Due to Bohm's boundary conditions, the parallel flux of out-flowing ions is supersonic on the limiter and divertor plates

This method has been applied to the approximation of a reduced MHD model using a decomposition of the velocity field into a parallel component and a perpendicular one given by the electric drift. The method is general and can be applied to any type of geometry. Figure 2 shows for instance the steady state density and parallel velocity fields in the edge region of a limiter tokamak (left) and of a divertor tokamak (right). Bohm's boundary conditions have been applied to the limiter and divertor plates producing a supersonic outflow velocity field.

5.1.4. *Mesh singularities and triangular elements*

Participants: Boniface Nkonga, Marie Martin, José Costa.

C1-finite elements as used for instance in the Jorek code are associated to isoparametric cubic-Bezier representation over quadrangles in the poloidal plane and sine-cosine Fourier expansion in the toroidal direction. Mesh singularities are associated to the structure imposed by the cubic-Bezier representation over quadrangles. In the context of the ANR-ANEMOS and in collaborations with IRFM and the Galaad team (Inria Sophia Antipolis), a geometrical toolbox is under development to manage these singularities and improve the alignment with equilibrium flux surfaces. As an alternative, we are also developing a more flexible C1-element over triangles using either Reduced-quintic (Bell) or quadratic Powell-Sabin polynomials. Optimal order of accuracy is achieved with simple boundary conditions. Many cycles of the "current hole" instability has been accurately reproduced. Additional improvement, with isoparametric formulation, will hopefully achieve this optimality for curved boundaries and improve mesh alignment to flux surfaces. We have investigated the possibility to use cubic splines representation in the toroidal direction. Indeed, for pellet injection the

local resolution needed in the toroidal direction requires a large number of Fourier modes. This resolution need is very local, adapted splines representation can be more efficient. This solution is under analysis and structuration. First application is expected at end of 2013 with a possible update of Jorek in 2014.

5.1.5. *Mesh adaptation Methods*

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

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

5.1.6. *Stabilization for finite / spectral element*

Participants: Boniface Nkonga, Marie Martin, Richard Pasquetti, Sebastian Minjeaud.

Formulation of Reduced MHD eliminates fast acoustic waves but material, slow acoustic and Alfven waves are included in this model. On the other hand, finite element approximation, when applied to hyperbolic systems (with finite speed waves) needs additional control of the effect of unresolved scales. We have developed and validated a Taylor Galerkin Stabilizations of order 2 and 3 (TG2-TG3) for reduced MHD. This global approach has been implemented in a simplified form, validated and updated in the latest versions of the Jorek code. Even if significant improvements have been observed with this stabilization where only material and Alfven waves subscales are stabilized, more robustness is expected by taking into account slow acoustic waves. Stabilization techniques well adapted to high order approximations, like the spectral vanishing viscosity method or the entropy viscosity technique, remain to be implemented in the Fourier-SEM code.

5.1.7. *Validity of the Reduced MHD and extensions*

Participants: Hervé Guillard, Boniface Nkonga, Afeintou Sangam.

The available reduced MHD model in Jorek uses a set of assumption that can be reasonable close to the equilibrium and during the linear grow of instabilities. In order to obtain accurate and robust simulations of the nonlinear instabilities saturations, careful analysis and derivation of the reduced MHD has been performed, more mathematical derivations are under progress under the asymptotic analysis framework. It turns out that some of the neglected terms can be of relative importance for the saturation process when MHD instabilities move the plasma far from equilibrium.

5.1.8. *High performance parallel computing*

Participants: Hervé Guillard, Boniface Nkonga, Sebastian Minjeaud.

Applications under concern in this project needs to manage large meshes (10^7 to 10^9 nodes) and solve many huge sparse nonlinear systems. This makes the use of domain partitioning techniques unavoidable. In addition, since different numerical methodologies are under studies and evaluations in this project, we need to develop a quite general setting allowing the use of different data structures (element-oriented for FE vs edge-oriented for FV) and the possibility to consider different domain overlapping to efficiently communicate between processors. For this we develop the PaMPA (Parallel Mesh Partitioning and Adaptation) software in collaboration with the Bacchus team (Inria, Bordeaux). PaMPA is based on the PT-Scotch graph partitioning tool and allows on the fly mesh redistribution. Up to now, PaMPA has been tested on 10000 processors with a mesh of 20M tetrahedrons. Integration of PaMPA as an external library to the codes developed in this project is under progress and early results are promising. Similarly, the Fourier-SEM code is currently parallelized.

5.2. Optimisation and control for magnetic fusion plasmas

5.2.1. *Evolutive equilibrium and transport coupling and optimization of scenarii*

Participants: Jacques Blum, Cédric Boulbe, Afeintou Sangam, Gael Selig, Blaise Faugeras, Holger Heumann.

5.2.1.1. Research of optimal trajectories for the monitoring of Tokamak discharges

The direct equilibrium code CEDRES++ in its static version (resp.dynamic) computes for externally applied PF currents (resp. voltages) and given plasma current density profile the (resp. evolution of the) poloidal flux and the plasma free boundary. The research of optimal trajectories is the corresponding inverse problem : find externally applied currents (resp. voltages), such that the plasma reaches a certain desired state. This desired state is mainly (resp. the evolution of) a prescribed plasma boundary. We formulate these inverse problems as so-called optimal control problems, where the PF currents (voltages) are the so-called control variables and the poloidal flux the so-called state variable. Optimal control problems are optimization problems with PDE (partial differential equations) constraints. In our case, the Grad-Shafranov equation is the constraint and the functional to be minimized is a cost-function that measures the mismatch between the computed plasma boundary and the desired plasma boundary. The Sequential Quadratic Programming (SQP) method is known to be a very efficient algorithm for solving non-linear constrained optimization problems. We implemented in CEDRES++ the SQP method for the two cases of finding either currents or voltages that corresponds to a desired boundary or a desired evolution of the boundary. These implementations are built on the orignal Newton methods for the direct non-linear problems. For optimization problems it is of great importance that the Newton methods are 'real' Newton methods in the sense that the Newton matrices are real derivatives. In the original implementation of CEDRES++ these matrices were the discretization of analytic derivatives of the non-linear operators, hence not derivatives of the discrete problem. We had to rewrite large parts of the code to eliminate this problem. Further, we added an interface to the linear solver library UMFPACK. For the current mesh resolution level, the performance of this linear solver for the stationary problems, both in the direct and in the inverse versions, is superior to iterative linear solvers. In the case of the inverse non-stationary problem, the problem of finding voltages that correspond to a desired evolution, the memory requirements forbid the use of UMFPACK. There, we used Conjugate Gradient-type iterations. In the future, we will have to investigate if other types of iterative solvers are suitable and allow a certain parallelism that will speed up the simulation time.

5.2.1.2. A new method of coupling equilibrium and resistive diffusion equations

In the framework of Gael Selig's PhD thesis, the resistive diffusion equation has been incorporated in the evolutive equilibrium system of CEDRES++. This equation has as unknown variable the derivative of the poloidal flux with respect to the averaged minor radius of the magnetic surface. This choice was made instead of the poloidal flux itself because this is the quantity directly involved in the averaged Grad-Shafranov equation used to compute the FF' term and thus this allows us not to perform a supplementary numerical differentiation which might introduce some numerical instability. An algorithm based on a successive prediction and correction method is proposed in order to ensure the consistency between the evolution of the 2D poloidal flux in the equilibrium equation and the evolution of the poloidal flux in the 1D resistive diffusion equation. The algorithm guarantees that at the end of each time step the total plasma current Ip and the mean radius of the plasma have the same values in both systems (see fig.2). The convergence of this new code (called CEDRES-DIF) has been numerically validated and the method has been successfully compared by G. Selig to the CEDRES-CRONOS coupled code which uses another coupling algorithm.

5.2.1.3. Introduction of halo currents in the equilibrium resolution

When VDE (Vertical Displacement events) instabilities occur in a Tokamak, currents flow from the plasma to the machine vessel structures, and then return to the plasma. These currents are called halo currents . In turn, these currents induce forces on the wall when crossing with Tokamak poloidal and toroidal magnetic fields. Moreover, when VDE instabilities take place, the plasma hits the wall with all its energy. Therefore, it is worth understanding the contribution of halo currents to total plasma current and other related plasma parameters, particularly the distribution, magnitude, and temporal evolution of halo currents for large scale machine such as ITER. Even if halo currents are actually 3D phenomena, it is important to take into account their effects in 2D free boundary equilibrium codes. In halo region, the pressure can be considered as negligeable so that the current follows the magnetic field lines. The magnetic field satisfies the force free equation $jxB = 0, \nabla.B = 0$ which can be rewritten

$$-\Delta^* \Psi = \frac{1}{\mu_0 R} \frac{\partial}{\partial \Psi} f_H^2(\Psi)$$

in an axisymmetric configuration. The function $f_H(\Psi)$ is supposed to be known. This simple model has been implemented in CEDRES++ and first tests have been done. This first simplified model has to be improved to get more realistic simulations and to be validated. The choice of the function f_H, the value of the total halo current, the geometry and the size of the halo region need to be enhanced with respect to experimental data.

5.2.2. *Equilibrium reconstruction and current density profile identification*

Participants: Jacques Blum, Cédric Boulbe, Blaise Faugeras.

EQUINOX is a real-time equilibrium reconstruction code. It solves the equation satisfied by the poloidal flux in a computation domain, which can be the vacuum vessel for example, using a P1 finite element method and solves the inverse problem of the identification of the current density profile by minimizing a least square cost-function. It uses as minimal input the knowledge of the flux and its normal derivative on the boundary of the computation domain. It can also use supplementary constraints to solve the inverse problem: interferometric, polarimetric and MSE measurements. Part of the work reported here has been done in the frame of a RTM-JET contract.

5.2.2.1. *Direct use of the magnetic measurements*

Equinox was not originally designed to take as magnetic inputs directly the magnetic measurements, as it should be the case in the ITM, but some outputs from the real-time codes Apolo at ToreSupra and Xloc at JET. These codes provide Equinox with the values of the flux and its normal derivative on a closed contour defining the boundary of the computation domain (this contour can be the limiter for example). As a consequence the main difficulty arising in the objective of integrating the code Equinox in the ITM structure was to interpolate between the magnetic measurements (flux loops and poloidal B-probes) with a machine independent method. This has already been achieved by using toroidal harmonic functions, as a basis for the decomposition of the poloidal flux in the vacuum region, in complement to the contribution of the PF coils. This method can provide an alternative tool, comparable to APOLO (for Tore Supra) and FELIX (for JET), to compute the plasma boundary in real time from the magnetic measurements. Some twin experiments for WEST (Tore Supra upgrade) have been successfully conducted. In a first step the equivalents of magnetic measurements were generated using the FBE code CEDRES++. In a second step these measurements were used by the toroidal harmonics algorithm to reconstruct the plasma boundary. The results are very promising and the work on this subject is ongoing for JET.

5.2.2.2. *Boundary conditions for EQUINOX*

In the present version of EQUINOX the boundary condition is a flux condition (Dirichlet boundary condition) and the tangential component of the poloidal field is incorporated in the cost-function to be minimized. This is a constant criticism which is made on EQUINOX. The idea was to inverse these two boundary conditions in order to determine if this choice is determinant in the results. We tried to use the tangential poloidal field (Neumann boundary condition for the flux) as boundary condition for the boundary value problem, and to put the flux (or its tangential derivative linked to the normal component of the poloidal field) in the cost function. However no convincing results could be obtained because the numerical resolution of the boundary value problem associated with Neumann boundary conditions proved to be unstable. This might be explained by the fact that a compatibility condition has to be satisfied between the Neumann conditions and the current density in the plasma which evolves during the mixed fixed-point and optimization iterations.

5.2.2.3. *Induced currents in EQUINOX*

In a disruption when the total plasma current disappears, there are very important induced currents, for example in the toroidal pumped limiter. These currents are in the domain of resolution of EQUINOX. Therefore it is necessary to take them into account in the resolution of the equilibrium reconstruction problem. This has been tested on a Tore Supra disruption case. The mesh generation has been modified in order to incorporate the real structure of the limiter. The structure of the equations being solved in the code also had to be modified in order to take into account the measured induced currents.

5.3. Turbulence models

5.3.1. *Hybrid RANS/LES models*

Participants: Hubert Alcin [Tropics], Alain Dervieux, Bruno Koobus [University of Montpellier 2], Carine Moussaed [University of Montpellier 2], 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, a novel dynamic formulation has been introduced in our models and tested. the new model has been adapted to very high Reynolds number. Carine Moussaed has presented her results in ECCOMAS (Vienna). In our set of benchmark test cases which are also ECINADS test cases, the flow past a circular cylinder at Reynolds number from 3900 to 1 Million could be passed with improved predictions of main properties like mean drag, root mean square of lift fluctuation, and base pressure.

5.3.2. *Acoustics*

Participants: ILya Abalakin [IMM-Moscou], Alain Dervieux [Tropics], Alexandre Carabias [Tropics], Tatyana Kozubskaya [IMM-Moscow], Bruno Koobus [University of Montpellier 2].

A method for the simulation of aeroacoustics on the basis of hybrid RANS/LES 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. This year, this scheme is also introduced in the Gamma-Sciport mesh adaptation loop.

5.4. Environmental flows

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

Mobile bed and sediment transport

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 published in [14] and is now applied to the study of the Nador Lagoon in Morocco.

COFFEE Project-Team (section vide)

<div align="center" style="color:red">

COPRIN Project-Team

</div>

6. New Results

6.1. Interval analysis

6.1.1. *A Contractor Based on Convex Interval Taylor*

Participants: Gilles Trombettoni [correspondant], Bertrand Neveu.

Interval Taylor has been proposed in the sixties by the interval analysis community for relaxing continuous non-convex constraint systems. However, it generally produces a non-convex relaxation of the solution set. A simple way to build a convex polyhedral relaxation is to select a *corner* of the studied domain/box as expansion point of the interval Taylor form, instead of the usual midpoint. The idea has been proposed by Neumaier to produce a sharp range of a single function and by Lin and Stadtherr to handle $n \times n$ (square) systems of equations.

This paper presents an interval Newton-like operator, called `X-Newton`, that iteratively calls this interval convexification based on an endpoint interval Taylor. This general-purpose contractor uses no preconditioning and can handle any system of equality and inequality constraints. It uses Hansen's variant to compute the interval Taylor form and uses two opposite corners of the domain for every constraint.

The `X-Newton` operator can be rapidly encoded, and produces good speedups in constrained global optimization and constraint satisfaction. First experiments compare `X-Newton` with affine arithmetic[31], [19], [20]

6.2. Robotics

6.2.1. *Robotics*

6.2.1.1. *Kinematics of wire-driven parallel robots*

Participants: Laurent Blanchet, Jean-Pierre Merlet [correspondant].

The kinematics of wire robot is a complex problem because it involves both the geometrical constraints and the static equilibrium constraints as only positive tensions in the wire are possible. A major issue, that has not been addressed in the literature [16], [15], is that for a robot having n wires the forward kinematic problem (FK) (determining the possible pose(s) of the robot knowing the wire lengths, a problem that is crucial to solve for controlling the robot) cannot be solved by assuming that all n wires are under tension as the current pose of the robot may be such that only a subset of the wires may be under tension. Hence the FK problem has to be solved for **all** robots that may be derived from the initial one by removing 1 one to $n - 1$ wires, each solving leading to a set of possible poses for the platform. Solving the FK for 1 wire is trivial, while for 6 wires the FK solving may be based on the already complex FK of parallel robot with rigid legs. For 2 wires it can be shown that the FK solutions can be found by solving a 12th order univariate polynomial, while for 3 wires we have shown last year by using an elimination procedure that the solutions are obtained by solving a 158th order polynomial. A very recent result of this year is that for 4 wires the order of this polynomial is 216, while no known result has been established for 5 wires (note that for 3 to 5 wires the size of the system of equations that has to be solved for the FK is larger than the one for the FK of 6-dof robot with rigid legs, a problem that has required 20 years to be solved).

Drawbacks of the elimination approach is that it does not take into account 1) that the solution should be mechanically stable, 2) that the wire tensions at the solution(s) must be positive. Hence, assuming that all solutions may be computed by the elimination approach, an a-posteriori analysis has to be performed to sort out the solutions that verify 1) and 2). We have proposed this year an efficient method to determine if a solution was mechanically stable [13]. But another major issue with the elimination method is that it leads to high order polynomial that cannot be safely numerically solved. To address this problem and 2) we are considering a numerical algorithm based on interval analysis, that consider also the tension as unknowns, hence allowing to search only for solution(s) with positive tensions [21].

Another issue for wire-driven parallel robots is the concept of redundancy. Having more wires than dof to be controlled is interesting for increasing the workspace of the robot. But it is believed that redundant wires may also be used to better distribute the load among the wires. Unfortunately we have shown for the $N-1$ robot (all N wires connected at the same point on the platform) with non elastic wires that whatever N there will be at most 3 wires under tension simultaneously [25] and consequently that tension management is not possible (with 3 wires the tensions is uniquely determined). If the wires are elastic, then tension management is possible but the positioning error is very sensitive to errors in the stiffness model [24]. Hence new method for tension management should be devised and we have explored some possibilities [23]. Still there is some magic in wire-driven parallel robots: in spite of all the uncertainties prototypes work quite well, a phenomenon which has been explained through a sensitivity analysis [24], [22].

Finally we address the management of modular robots, whose geometry can be adapted to various tasks and different objects to be manipulated, especially for very large scale robot [28], that may be used in industry for maintenance and logistics (see the Cablebot project in section 8.2.1).

6.2.1.2. Robot Calibration
Participants: Thibault Gayral, David Daney [correspondant], Jean-Pierre Merlet.

6.2.1.2.1. Experimental calibration of a high-accuracy space telescope

A collaborative work began in October 2010 with Thales Alenia Space on the calibration of the mechanical structure of a space telescope. Its architecture is based on a parallel manipulator (of the active wrist 6-PUS type, which has been patented by COPRIN) and is used to correct the relative position of two mirrors. The aim is to reach a micrometer accuracy in order to obtain a suitable quality of the images provided by the telescope. Thus, a complete model of the space telescope needs to be developed and validated through calibration. Since high velocity is not required in such an application, the dynamic effects can be neglected and only geometric and/or static calibration has to be considered. Moreover, measurements for calibration were performed in a clean room under controlled pressure, temperature and humidity conditions to minimize the influence of the non-geometric errors. Thus, two possible static inaccuracy sources were identified and modeled: one from the deformation of the mobile platform and the other resulting from the behavior of the flexure joints. Three incremental models of the flexure joints were developed and compared: a spherical joint model, a model issued from the beam theory and a stiffness model. Results of calibration using an accurate measurement system of photogrammetry showed that the flexure joints can be modeled by perfect spherical joints due to the small workspace of the telescope. Concerning the mobile platform deformation, two models were developed. Good accuracy results were obtained for both models. The developed models allowed us to explain how the model errors are directly accounted in the parameter identification during calibration. This resulted in different sets of identified parameters which all enable a good positioning accuracy. Those differences were explained and results of calibration allow a proper choice of the model of the mobile platform deformation. Considering this model, a positioning accuracy of some micrometers was finally reached after calibration with only position and orientation measurements of the mobile platform, which should allow the calibration of the telescope in space [33]. This is currently under study using interferometric measurements on the prototype of the space telescope.

6.2.1.2.2. Calibration of a cable-driven robot

To improve the accuracy of a cable manipulator, it is necessary to identify the uncertainties of its model. The cable robots, studied in the ANR funded project Cogiro (see section 8.1.1.2), are theoretically redundantly actuated: the number of powered wires is larger than the number of degrees of freedom of the manipulator (however see section 6.2.1.1 about the reality of this redundancy).

In 2011 an over-constrained prototype was self-calibrated (the identification of the parameters does not need additional external measurement), under some assumptions on the cable properties [17], [29]. We will apply our recent calibration methods on the large scale robot prototype developed for the Cogiro project at the very end of this year.

6.2.1.2.3. Cable properties

Quite often cable-driven robot analysis assume mass-less and non-elastic wires. We proposed a method based on interval analysis to judge the validity of this assumption for a particular robot in a specific workspace. Our aim is to use this method in order to determine a a region within the robot workspace for which the hypothesis is valid and consequently for which self calibration of the robot is possible. Indeed, the assumption on the cable properties is not acceptable over the full workspace of the large scale robot developed in the Cogiro project. Still a self-calibration is possible if calibration poses are chosen within a specific subpart of the workspace. A more efficient calibration approach is in progress with additional measures and a more complex model (static and elasticity). The results has been published in [18], [30].

6.2.1.3. Assistance robotics

Participants: David Daney, Claire Dune, Jean-Pierre Merlet [correspondant], Yves Papegay, Odile Pourtallier.

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 smart devices that may help disabled, elderly and handicapped people in their personal life, provide also assistance to family and caregivers while allowing doctors to get better and objective information on the health state of the end-user. Our credo is that these devices 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). As for cost reasons we intend to use only standard hardware adaptation has to be taken into account at the very early stage of the system design and uncertainties in the physical instances of our systems are also to be considered.

For validation purposes we have developed a 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 [1]. Our activity in this field is concentrated on transfer, manipulation, walking monitoring, rehabilitation and the use of virtual reality. We are also investigating how such complex environments with multiples smart agents, quite heterogeneous from a computing viewpoint, but that have to cooperate, may be programmed. All these topics are in accordance with the one of the large scale initiative PAL [2] of which we are an active member.

6.2.1.3.1. Transfer and manipulation

Participants: François Chaumette [Lagadic], Jean-Pierre Merlet, Rémy Ramadour.

For transfer operation we are using the `MARIONET-ASSIST` robot (see section 6.2.1.4) that is installed in our flat. Currently we use 4 winches with the wires connected at the same point on the platform, hence providing 3 translational degrees of freedom. This low-cost, low-intrusivity robot has proved to be very effective for transfer operation. Apart of transfer operation robot may be used for manipulation. Adding one or several low-cost cameras (the cost being here a fundamental constraint), visual-servoing control is used to provide a whole new set of useful services such as grasping objects in order to bring them to the end-user (if they are too heavy, too far, high or low), or cleaning the table after lunch. Using a parallel crane robot, we are able to cover a large workspace, the vision-control allowing us to obtain the precision required by the manipulation of daily-life objects. The collaborative implementation of the vision and the kinematic control of the robot gives us a way to make best use of the advantages of both parts, while overcoming their respective drawbacks.

Given a region where the object of interest belongs, the first step is to detect it in an evolutive environment. A segmentation is made, robust to luminance variations and perspective projections. The vision is then used to move the platform toward a desired position relatively to the target. In order to execute this task, some carefully chosen features are measured, allowing to estimate the incremental displacement required to move the end-effector to the desired place. We use the library ViSP for both the detection and the visual-control part.

---

[1]pictures of this assistive flat are available at http://www-sop.inria.fr/coprin/prototypes/main.html
[2]http://pal.inria.fr

Experimental results were obtained using a platform with 3 degrees of freedom and a single camera, grasping a single object and moving it from a place to another. We used for that basic image data such as 2D moments, allowing a fast computing and yet robustness in measurements. We currently address to generalize this manipulation to other configurations and to evaluate its robustness to calibration errors and other uncertainty sources. We also are looking for a global paradigm merging both the vision-based kinematic model and the mechanical one, which could significantly improve the efficiency of the experiment, while reducing the mathematical complexity behind each kinematic model considered on its own.

6.2.1.3.2. Walking monitoring
Participants: Claire Dune [Handibio], Jean-Pierre Merlet.

We use the walking aids `ANG-light` and `ANG-II` (see section 6.2.1.4) to monitor the trajectory of the walking aid. The on-board sensors of these aids allow to evaluate the step pattern, gait asymmetry,...during daily walking, hence providing an health monitoring system that is always available. `ANG-light` has been tested last year with 24 subjects that were themselves instrumented (accelerometers in the wrists and knees, pressure sensors in the shoes) and were asked to perform two different trajectories twice with/without the walking aid. The purposes were:

- to determine pertinent walking indicators
- to obtain a "gold" standard of these indicators for non pathological walking, taking into account the normal variability of the walking pattern
- to determine if indicators obtained with the walking aid may led to an accurate estimation of the indicators when the walking aid is not used

Several indicators have been determined after the analysis of these data. In a second phase the inclusion test of elderly people (30 subjects) is taking place at the CHU of Nice-Cimiez and will last until the first trimester of 2013. An analysis of the collected data, in close collaboration with the doctors, will allow to determine if the proposed walking indicators are pertinent.

Another interest of the walking aids `ANG` is that they allow to collect significant information for mobility during their daily use: slope and surface quality of the sidewalks and automatic detection of lowered kerbs with a ranking of their convenience. It will be interesting for a community to share such information that is collected by the community members. For that purpose we propose to use collective maps, such as OpenStreeMap, which allow for map annotation. To validate this concept we have used `ANG-light` to automatically annotate the map of the Inria Sophia site with pertinent information for walking aid and wheelchair users [3].

6.2.1.3.3. Rehabilitation
Participants: David Daney, Mandar Harshe, Sami Bennour, Jean-Pierre Merlet [correspondant].

The focus of our work is on analyzing knee joint motion during a walking activity. The main principle of the system is to observe relative motions of the collars attached to tibia and femur. The measurement of the motion of these collars is based on the wire actuated parallel robot architecture (using the `MARIONET-REHAB` robot, see section 6.2.1.4). 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.

Measurements in the global frame and collar specific local frames give precise data to reconstruct collar (and thus, knee joint) motion. The system developed already incorporates the optical motion capture, inertial measurement units and the wire sensors for comprehensive coordinated measurements of the motion of the knee.We have performed preliminary trials on three subjects for walking motion.

In the past year, we worked on processing the data to obtain pose and orientation information of knee joint. Data obtained from the trials was analyzed and post-processing steps were implemented to reduce noise and errors. In order to perform sensor fusion, we implemented a probabilistic estimation based method to estimate the pose.

---

[3]see http://www-sop.inria.fr/coprin/prototypes/main.html#ang

The results from these analysis have allowed us to identify the merits of our approach and also helped us identify improvements that are needed. We have also identified the possible changes to our mathematical model that could allow use of interval analysis tools along with probabilistic estimation methods. We have identified the changes needed to the hardware setup that will help reduce the sensor noise and error. These changes once implemented will allow us to improve the usability of the system and also point us towards newer areas for further investigation, including, for example, effect of sensor placement, collar design, and interval based extended Kalman Filters for pose estimation[8], [9].

6.2.1.3.4. Virtual reality

Virtual reality has proved to be an effective mean for dealing with rehabilitation, provided that motion is added to the 3D visual feedback. The MARIONET-VR robot, together with our motion base (see section 6.2.1.4) may provide very realistic motion in an immersive room and we will start in 2013 a collaboration with the VR4I and REVES project-teams on this issue. A first task will be to install a moving walking treadmill in the immersive room at Sophia and to combine motion of the treadmill with 3D viewing.

6.2.1.3.5. Programming

In our opinion there will not be a single assistance device that will be able to offer all the required help that may be needed but rather numerous redundant smart agents that are able to perform very efficiently (at low cost and with a small energy consumption) a set of specific tasks. Such agents must be able to communicate and possibly will have to cooperate in some cases (e.g. after a fall). They will be heterogeneous from a computer view point as the used agents will change according to technological advances and to the trajectory of life of the end-users. If the manual programming of a single agent is possible (although quite complex for some of them) the overall system cannot be managed in that way: we need an unifying framework for this development. For that purpose we are currently investigating the use of HOP, a multi-tier programming language for the diffuse Web developed in the INDES project-team. Already one of our wire-driven parallel robot MARIONET-SCHOOL has been programmed in HOP and can be seen as a web resource.

6.2.1.4. Prototypes

Participants: Julien Alexandre Dit Sandretto, David Daney, Claire Dune, Jean-Pierre Merlet [correspondant].

Experimental works are a key point in the field of service robotics: it allows for validating concepts, getting feedback from the end-users and discovering new problems. We have extensively developed prototypes this year [4]

6.2.1.4.1. Wire-driven parallel robots

The MARIONET family is now constituted of

- MARIONET-REHAB (2004-): using up to 7 linear actuators it is mainly used for rehabilitation and health monitoring although it is also a very fast pick-and place robot
- MARIONET-CRANE (2008-): a very large 6-dof rescue crane, portable and autonomous that can be deployed in 10 minutes and has a lifting capability of 2 tons
- MARIONET-ASSIST (2011-): a robot deployed in our flat with up to 6 winches, that is used for transfer operation and health monitoring, with a lifting ability of about 2000 kg
- MARIONET-VR (2011-): using up to 6 linear actuators it will be used in the Sophia immersive room for simulation and rehabilitation. It allows to fully lift a person and has been physically installed in the immersive room this year although it is not fully operational
- MARIONET-SCHOOL (2012-): a set of very low-cost robots that are intended to be used for dissemination. They fit in a small suitcase and can be deployed on a table or over a full classroom. We believe that such robots may be used by roboticists but also by researchers from other domains that may use the motion of the robot(s) to illustrate scientific concepts. Currently we have 3 such robots (one using Lego components, one with step motors and one with servos)

---

[4]pictures and videos of our prototypes are available at http://www-sop.inria.fr/coprin/prototypes/main.html or on our YouTube channel http://www.youtube.com/user/CoprinTeam/videos?flow=grid&view=0

6.2.1.4.2. Walking aids

The *Assistive Navigation Guide* (ANG) family is 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

- 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

For reaching these objectives we have developed two walking aids:

- `ANG-light`: a walking aid with encoders in the wheels, 3D accelerometer, gyrometer and GPS. These sensors allow to measure the trajectory of the walking aid and several features of the user's gait. This walking aid is currently being used at CHU Nice, see section 6.2.1.3.2 . A replica of `ANG-light` is currently under development at the Handibio laboratory of Toulon and will include force sensors in the handles to get measurement of muscular activities while walking.

- `ANG-II`: this aid is an evolution of the motorized walker `ANG`, with a lower weight and better integration

6.2.1.4.3. Other devices

As seen in the previous sections we have focused our work on mobility as it has been identified as a major priority during our two years interview period. Another priority is fall management: it is adressed with the `ANG`'s) but requires that the patient uses a walking aid. To obtain a better coverage we have developed an instrumented vest that includes an Arduino Lilypad microcontroller to monitor a 3D accelerometer and verticality sensors in order to detect a fall. This system may communicate an alarm through wifi, zigbee or even infrared signals. Apart of its low cost the system is washable and has a very low power consumption (we are considering energy harvesting to further increase its energy autonomy).

Two other important issues for assistance robotics is activities monitoring and patient localization. In cooperation with the STARS project-team an activity monitoring system within the Dem@care project. Its purpose is to log specif events such as taking a pen, setting on a kettle, ...using only simple sensors such as proximity and distance sensors, switches, that are more reliable and less complex than using a vision system. We have provided to the CHU a system that allows to monitor up to 208 different events. In the same manner we are working on a localization system of elderly in a room based only on distance sensors, that will be less intrusive and/or may complement the measurements of a vision system.

Rehabilitation is also part of our activities. Apart of the `MARIONET-REHAB` system we plan to investigate the use of Sophia immersive room for that purpose. To complement the motion that will be provided by the `MARIONET-VR` robot we have bought a 6-dof motion base from Servos with a nominal load of 150kg, that we have modified to accommodate our needs. We will use also two lifting columns with a load of 100 kg each, that will allow to manage motion for rehabilitation apparatus such as treadmill.

6.3. Miscellaneous results

6.3.1. *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 [10].

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 functionnalities.

- a C code generator which, using these models, automatically generates the numerical real-time simulation engines

- a technical documentation generator

To finalize the transfer of the technology demonstrated by our prototype to our industrial partner, an extensive collection of testing and corresponding improvements have been done in 2012. This step has delayed the delivery of the final version of our modeling and simulation environment to Airbus until November 2012.

6.3.2. *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 aircraft have been developed as user cases for the project.

In 2012 we focused on automatic generation of agent code based on models that is ready for integration in the ID4CS platform prototype.

6.3.2.1. 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 has pursued the study of equilibrium model for coupled electricity and CO2 allowance exchange markets (see also Section 7.3). We have mainly focused on the determination of Nash equilibrium for the coupled electricity and carbon markets with the assumption that the producers maximize their market shares. Nash equilibrium have been obtained by using as an intermediary step a decreasing auction mechanism.

We have also pursued an indifference pricing methodology which is presented in more details in Inria -TOSCA Team section.

DEMAR Project-Team

6. New Results

6.1. Modelling and Identification

6.1.1. *Subject-specific Center of Mass estimation in human subjects*

Participants: Alejandro González, Mitsuhiro Hayashibe, Philippe Fraisse.

Center of mass position (CoM) in humanoid robots can be used to generate a joint trajectory suitable for walking and standing. Oscillations of the CoM while maintaining a standing posture have been observed in older patients. These oscillations are thought to occur due a change in the subject's balancing strategy. This is why the motion of the CoM in generally considered as good metric to be used while diagnosing pathologies which affect gait. With this in mind we propose the use of the statically equivalent serial chain (SESC) to provide a subject specific estimate of CoM position. The state of the art techniques for CoM estimation in humans involve the use of expensive equipment in a laboratory setting, making it difficult to use as a clinical tool or inside the home environment. Current work for diagnosing a subject's balance moves away from this, focusing on wearable and minimally invasive sensors that obtain information during the subject's daily activities. We propose the use of widely available sensors like the Kinect camera, for tracking the subject's movements, and the Wii balance board during the calibration of the SESC.
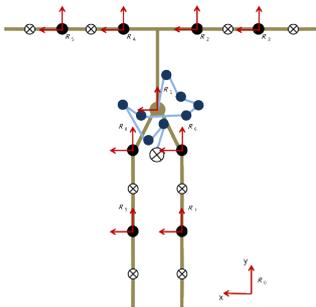
a)   b)

Figure 1. Center of mass estimation can be done using the statically equivalent serial chain. We assume a 9 rigid-link model (a) with spherical joints, capable of three dimensional movements. The SESC estimation can be performed in real time and is driven by joint angular measurements (b).

We have focused in improving the reliability of the identification during a study of the CoM trajectory in the sagital plane [14]. We studies small value for the condition number of the used data as well as of the parameter relative standard deviation ($\sigma_{\widehat{\mathbf{R}}_r}$ %) are useful to determine the validity of the estimate. Subsequent works have extended the human model to three-dimensional motion (Fig. 1 .a). In order to observe the tracking of the center of mass, we have developed a 3D visualization tool which represents the subject's skeleton and SESC in real time (Fig. 1 .b) [16]; it is also possible to observe the CoM history. Finally we are now comparing the performance of the inexpensive Kinect sensor and the traditional video based motion capture Vicon system, with reasonably good results [15]. Current work is also focused in improving the speed of the identification phase by simplifying the assumed human model and using physical constraints to reduce the complexity of the SESC. Also the development of a simple to follow identification protocol which takes into account multiple supporting surfaces is desired.

6.1.2. Recursive estimation of SESC parameters for human Center of Mass estimation

Participants: Alejandro González, Mitsuhiro Hayashibe, Philippe Fraisse.

A human's center of mass (CoM) trajectory is useful to evaluate the dynamic stability during daily life activities such as walking and standing up. To estimate the subject-specific CoM position in the home environment, we make use of a statically equivalent serial chain (SESC) developed with a portable mesurement system. In order to adapt to the subject's physical capacities we implement a constrained Kalman filter to achieve an online parameter estimation of the SESC parameters. By accounting for the human body bilateral symmetry we hope to reduce the identification time. This results in constrainig SESC parameters to be consistent with the human skeletal model used. The Kinect camera is used as a markerless motion capture system for measuring limb orientations while the Wii board is used to measure the subject's center of pressure (CoP) during the identification phase. For his experiment CoP measurements and Kinect data were recorded for five able-bodied subjects. The data was then given to the proposed recursive algorithm to identify the parameters of the SESC online (Fig.2). This method of online identification allows the subject or the therapist to know the quality of the on-going CoM identification while giving postures; by doing this the time needed to perform the identification can be reduced. A cross-validation was performed to verify the identification performance.
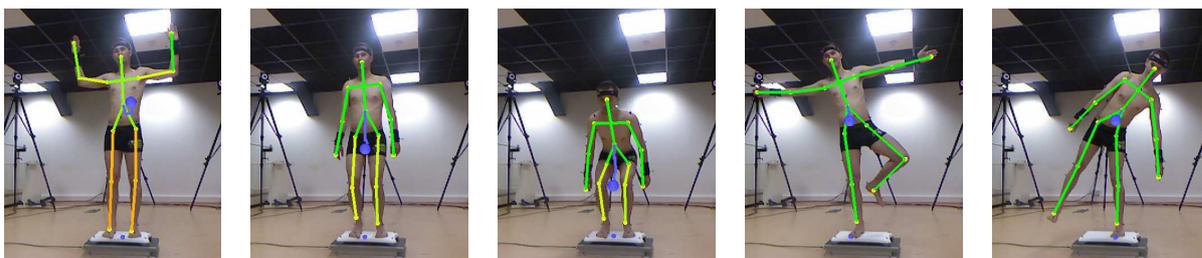
Figure 2. Online statically equivalent serial chain (SESC) parameter estimation. The length of each of the SESC's link is updated when a static pose has been found and can be observed in real time. The color of the skeleton can be updated as a cue to the subject and/or therapist.

6.1.3. FES-Induced Torque Prediction with Evoked EMG Synthesized by Recurrent Neural Network

Participants: Zhan Li, Mitsuhiro Hayashibe, David Guiraud.

A NARX-type recurrent neural network (NARX-RNN) model is proposed for identification and prediction of FES-induced muscular dynamics with eEMG. Such NARX-RNN model is with a novel architecture for prediction, with robust prediction performance. To make fast convergence for identification of such NARXRNN, directly-learning pattern is exploited during the learning phase. Due to difficulty of choosing a proper forgetting factor of Kalman filter for predicting time-variant torque with eEMG, such NARX-RNN may be considered to be a better alternative as torque predictor. Data gathered from two SCI patients is used to evaluate the proposed NARX-RNN model. The NARX-RNN model shows promising estimation and prediction performance only based on eEMG [23].

6.1.4. Inverse Estimation of Muscle Activations with Weights Optimization

Participants: Zhan Li, Mitsuhiro Hayashibe, David Guiraud.
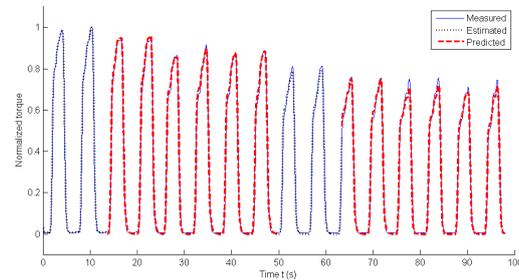
Figure 3. Performance of NARX-RNN model with eEMG for prediction with identification on periodically

Inverse estimation of activations of muscle groups at human lower leg in random movement condition is investigated with merely the ankle joint torque used. Optimization technique for the relationship between muscle activations and torque is exploited. Such optimization is able to rebuilt the relationship between muscle activations and torque inversely based on experimental data obtained from five healthy subjects, and the optimal weight matrix can indicate each muscle's contribution for producing the torque. Further cross-validation on prediction of muscle activations with joint torque with optimal weights shows such approach may possess promising performance [22].
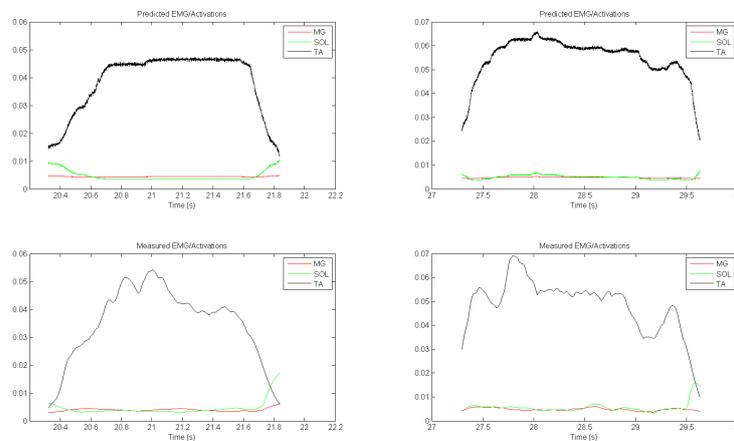
Figure 4. Prediction of muscle activations

6.1.5. 3D Volumetric Muscle Modeling For Real-time Deformation Analysis With FEM
Participants: Yacine Berranen, Mitsuhiro Hayashibe, Benjamin Gilles, David Guiraud.

Computer simulators are promising numerical tools to study muscle volumetric deformations but most models are facing very long computation time and thus are based on simplified Hill model versions. The purpose of this study is to develop a real-time three-dimensional biomechanical model of fusiform muscle based on

modified Hill model for the active stress which is controlled from EMG recordings. Finite element model is used to estimate the passive behavior of the muscle and tendons during contraction. We show that this 3D model implementation is very cost effective with respect to the computation time and the simulation gives good results compared to real measured data. Thus, this effective implementation will allow implementing much more complex and realistic models, with moderate computation time.
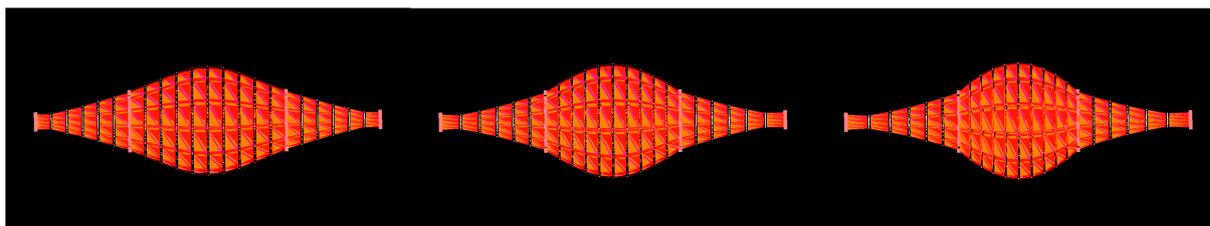
Figure 5. Real-time Simulation of Volumetric Muscle Deformation.

6.1.6. *Principal Geodesic Dynamics*

Participants: Maxime Tournier, Lionel Reveret.

This paper presents a new integration of a data-driven approach using dimension reduction and a physically-based simulation for real-time character animation. We exploit Lie group statistical analysis techniques (Principal Geodesic Analysis, PGA) to approximate the pose manifold of a motion capture sequence by a reduced set of pose geodesics. We integrate this kinematic parametrization into a physically-based animation approach of virtual characters, by using the PGA-reduced parametrization directly as generalized coordinates of a Lagrangian formulation of mechanics. In order to achieve real-time without sacrificing stability, we derive an explicit time integrator by approximating existing variational integrators. Finally, we test our approach in task-space motion control. By formulating both physical simulation and inverse kinematics time stepping schemes as two quadratic programs, we propose a features-based control algorithm that interpolates between the two metrics. This allows for an intuitive trade-off between realistic physical simulation and controllable kinematic manipulation.

6.1.7. *An improved kinematic model of the spine for three-dimensional motion analysis in the Vicon system*

Participants: Pawel Maciejasz, Wieslaw Chwala (University School of Physical Education, Krakow), Miroslawa Dlugosz, Daria Panek, Witold Alda (AGH University of Science and Technology, Krakow).

The mechanism of creation and pathomechanics of lateral spinal deformation is still not fully explained. Modern medical imaging techniques give scientists possibility to understand some aspects, but vast majority of those techniques is based on static trials. A motion capture system belongs to techniques which enable visualization of a spine during dynamic trials; however, due to lack of appropriate computational model, it is unsuitable for scoliosis imaging.

A few years ago a kinematic model of the spine has been proposed to be used with Vicon Motion Capture System (Master thesis of P. Maciejasz). This model was based on Bézier curves and allowed for much more precise investigation of spinal kinematics during dynamic trials as compared with other computational models. However, it did not allowed to restrict only selected movements for particular segments of the spine (e.g. axial rotation for lumbar spine). The aim of the current work is to improve the proposed model in order to be able to restrict selected movements according to the knowledge concerning spinal anatomy and spinal range of motion. The new kinematic model of the spine was written in BodyBuilder for Biomechanics Language.
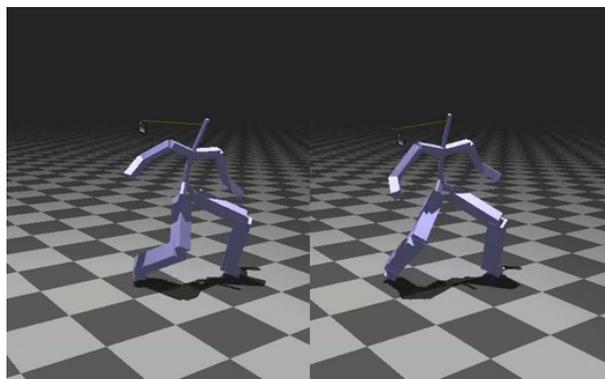
Figure 6. Real-time user interaction with a virtual character maintaining balance, animated using our approach: based on a break-dance motion capture sequence, our dimension reduction process allows one to compute both dynamics and task-space control in a low-dimensional, data-driven subspace.

For the purpose of visualization also an accurate graphical representation of each vertebra (polygon mesh) was computed and adapted to be compatible with the kinematic model. Using a new version of the model it is possible to perform precise analysis of movement of all vertebrae during such dynamic activities as e.g. gait and forward or lateral bending, as well as to present the results not only on the charts, but also as a 3D animation of movements of a realistically looking spine [12].

6.1.8. *Methodology of automated detection and classification of action potentials in nerve fiber based on multichannel recordings*

Participants: Thomas Guiho, Pawel Maciejasz, David Guiraud.

For some neuroprosthetic applications it would be beneficial to be able to automatically detect when particular nerve fibers (axons) are in "firing" (i.e. when an action potential is propagating along them). Due to limitations of currently available recording electrodes on one side, and the attempt to be as little invasive as possible, at the moment in practical application it is not possible to record signals coming from a single nerve fiber. In signal recorded using typical electrodes placed close to nerve fibers, action potentials coming from various nerve fibers, as well as noise coming from outside of the nerve, may be detected. One of the possibilities allowing to distinguish action potentials coming from various nerve fibers in such a case is to record signals at a few places along the nerve and compare them.

We have proposed an algorithm that allows to automatically detect and classify evoked action potentials in a simple earthworm model. The signals were recorded concurrently at 2 places along the giant nerve fibers. In the first step the algorithm tries to identify only does recordings in which action potentials generated by various nerve fibers can be easily distinguished. Afterwards, the most significant features (such as amplitude, duration, propagation velocity, etc.) to distinguish between different populations of fibers are identified. Finally, the action potentials in all signals are identified and classified using the features determined in the previous step.

The proposed method was implemented using MATLAB software and tested on the file containing almost 200 signals record in response to various stimuli. The same data ware later inspected manually and the action potentials were manually classified. More than 99% of action potentials were classified to the same nerve fiber when performing automatic and manual classification.

6.2. Function control and synthesis

6.2.1. FES assisted sitting pivot transfer

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

Transferring from a wheelchair to a treatment table, bed, tub/shower bench, toilet seat, car seat and vice versa represent typical Sitting Pivot Transfer (SPT) realized by individuals with Spinal Cord Injury (SCI). Individuals with SCI, perform this postural task around fifteen times a day using upper extremities. In the chronic stage after SCI, soft tissue structures are exposed to overuse in activities of daily living, such as, transfer task in which the shoulder becomes a weight-bearing joint. Therefore, the risk of shoulder pain and musculoskeletal disorders is higher in persons with paraplegia compared with an able-bodied population. A lot of scientific effort has been focused on experimental studies in which the kinetic and the kinematic of the SPT movement have been analyzed. To our best knowledge, the scientists have focused their attention only on the performance of SPT; the influence of Functional Electrical Stimulation (FES) on SPT maneuver has not been investigated so far.

Therefore, we investigate the influence of FES on SPT motion of a paraplegic person. First, we develop dynamic optimization method in order to predict SPT motion of an able-bodied subject. This approach have been validated by comparing computed SPT trajectories with the ones measured during the experiment with an able-bodied subject (see Fig. 7). After validating our method, we used the optimization tool for analyzing the influence of FES on SPT maneuver in paraplegic persons. Our results suggest that FES can decrease arm participations during the transfer motion of a paraplegic person. [6], [21].
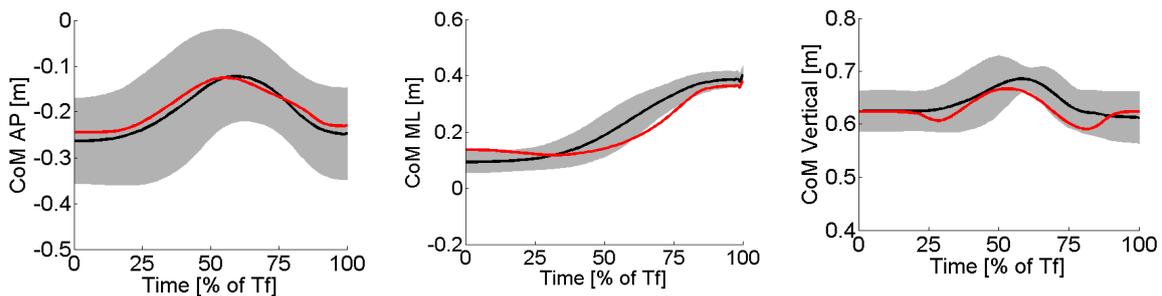
Figure 7. Computed (red line) CoM position, mean value of CoM positions estimated from experimental data (black line) and its plus/minus standard deviation (gray line) in AP, ML, and vertical direction in able-bodied subject.

6.2.2. FES assisted prolonged standing

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

Prolonged immobilization which occurs after spinal cord injury results in many physiological problems. Standing therapy can ameliorate many of those problems. The approaches proposed in the literature for restoration of standing in paraplegic population based on Functional Electrical Stimulation (FES) focuse on the control of each individual joint, i.e. joint space control. In those cases the balance of the postural system is not directly controlled. This could be problematic especially when only the lower limbs are controlled. During paraplegic's quiet standing two concurrent controllers are acting in parallel, the physiological system under control of Central Nervous System (CNS), and artificial FES system. Upper part of the paraplegic's body is under voluntary control, therefore artificial controllers should be designed in the way to take into account actions of the intact part of the body and to assist users in their task.

For human beings the Center of Mass (CoM) provides an indicator of stability and it is an essential parameter in human postural stability. By controlling CoM position in paraplegic person the voluntary motions under CNS control are taken into account. Therefore, in we propose a whole body controller based on control of the CoM position. The goal was to develop a simple balance controller which would, by means of FES, enable quiet standing of individuals suffering from SCI while taking into account the voluntary motion of the upper limbs. The controller should enable prolonged standing by simulating the behavior of an able-bodied subject during the standing task, i.e. by imposing posture switching and in that way allowing the stimulated muscles to relax. The proposed approach is based on a 10 DoF biomechanical model and Proportioan Integral (PI) controller (see Fig. 8). The validity of the approach is tested, in computer simulations, using human CoM trajectories estimated from experimental data and by applying perturbations in simulation during quiet standing in order to simulate voluntary upper body movements.The results show that proposed controller is able to track desired CoM position with sufficient precision and to maintain stability even in the presence of simulated movements of the upper body [20].
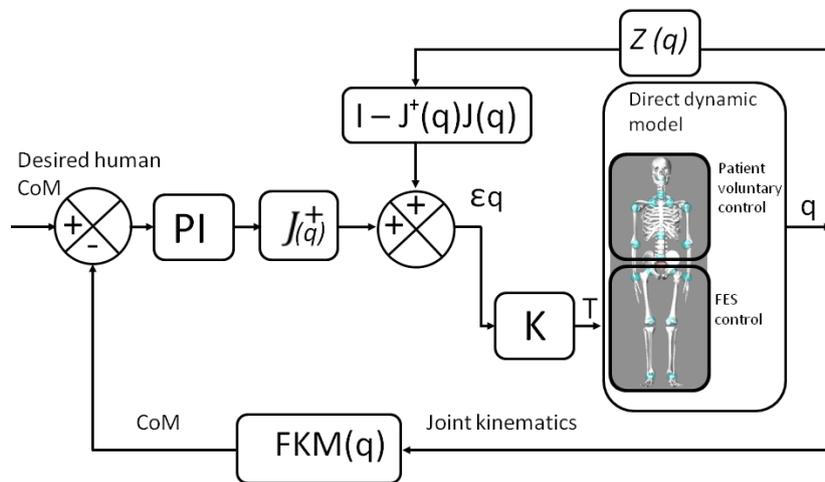
Figure 8. Block diagram of the proposed postural controller. Controller follows desired 3D CoM positions and controls the lower limbs by applying torque at ankle, knee and hip joints (10 DoF).

6.2.3. Bimanual reeducation assisted by FES in post-stroke patients

Participants: Nader Rouis, Christine Azevedo Coste, Philippe Fraisse, Isabelle Laffont, Denis Mottet.

This starting project will investigate the possibility to stimulate the deficient upper limb of a post-stroke patient in order to reproduce the movements observed on the valid upper limb or in order to achieve a bimanual task in cooperation with the valid upper limb. The aim is to improve the bimanual training tasks classically used in fictional rehabilitation. Both embedded sensors and kinect type systems will be investigated as possible ways to observe the valid upper limbs.

6.2.4. Freezing detection in Parkinson Disease patients

Participants: Maud Pasquier, Christine Azevedo Coste, Christian Geny, Bernard Espiau.

This work intends to apply the results of Maud Pasquier thesis about data segmentation and locomotion analysis to the detection as soon as possible of freezing episodes in Parkinson Disease (PD) patients. PD is a chronic degenerative disease of the central nervous system. One of the consequence is walking troubles and increased postural instability and falling risks. Freezing concerns at least half of PD patients, it is characterized by the transitory incapacity to make a step. It classically occurs at the gait initiation, turn around and passing doors. This freezing of gait (FOG) strongly impacts patient's mobility. As an example, in figure 9 , data recorded by an inertial sensor placed at the ankle is presented.
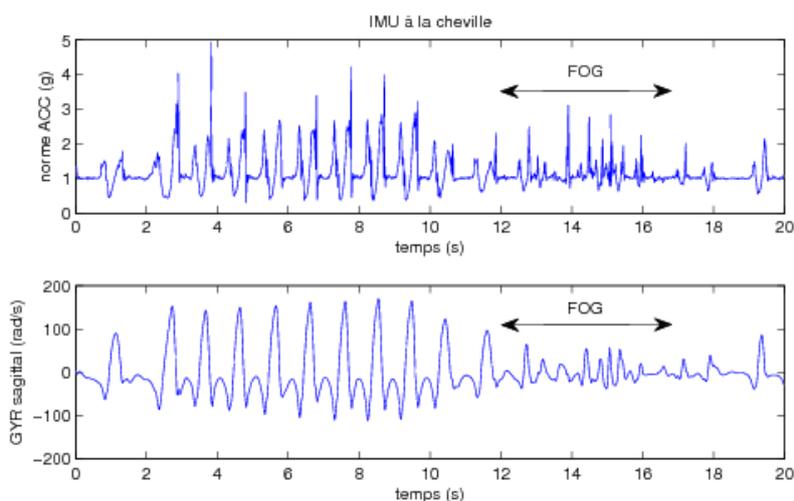
Figure 9. Signals recorded with an inertial sensor placed at the ankle during a walking trial. Top: acceleration norm, Bottom: sagittal plane rotation velocity

Assistive devices have been proposed, it has been shown that providing the patient with an auditory metronome or visual lines on the ground allows to reduce FOG occurrence but the effects are not maintained within the time. An approach would be to present this signals only when a freezing episode occurs and this implies to be able to detect it. Several authors have shown that FOG are in general associated to walking rhythm variability. Moore et al have used an accelerometer placed on lower limv in order to detect the presence of high frequencies. Indeed, the tremor pattern which can be observed during freezing is located at 3 et 8Hz, whereas normal locomotion is around 3Hz. These authors are able to detect a large part of the FOG but only those presenting high frequency patterns which is not systematic. Furthermore the detection delay (FFT) is very high and cannot be compatible with assistive device control constraints. We have proposed to observe stride properties in an online manner and compute a criterion which value informs about the FOG occurrence. The criterion is based on two variables: the stride length and the cadence. When stride length diminishes and cadence increases a FOG may be upcoming. 3 patients have been involved in this study and the criterion proposed has been shown to be as efficient than Moore's method in terms of number of detected FOG but the detection time is strongly improve with our method.

6.2.5. "Awake surgery" of slow-growing tumors and cortical excitability measured by EEG recordings.
Participants: François Bonnetblanc, Guillaume Herbet, Pom Charras, Mitsuhiro Hayashibe, David Guiraud, Hugues Duffau, Bénédicte Poulin-Charronnat.

Using direct electrical stimulation, real-time functional mapping of the brain can be used to perform resections of slow-growing infiltrative tumors in awake patients and to prevent the resection of essential areas near the tumor. To investigate interhemispheric imbalance following " awake surgeries" of slow-growing tumors we recorded EEG in a visuo- manual RT paradigm. Increase of cortical excitability within the ipsilesional hemisphere was signed by increased event related potentials (ERPs) amplitude for two patients. The cortical excitability in the lesioned hemisphere may be increased to maintain performances and cerebral plasticity.
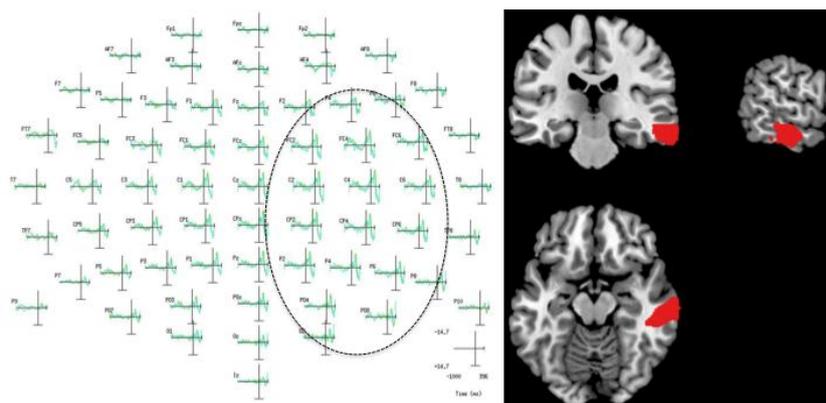
Figure 10. Left : Post-operative Event-Related Potentials (ERPs) for Patient 1. Increased ERPs amplitude can be seen in the right ipsilesional hemisphere (ellipse) in comparison to the contralesional hemisphere and homologous recording sites. The patient had to respond to visual stimuli occurring in the right or left hemifield with his right hand. The vertical line on each ERP indicates the occurrence of the visual go- signal Right : Lesion mapping for the same patient after the surgery.

6.3. Neuroprostheses

6.3.1. Distributed Measurement Unit for Closed-Loop Functional Electrical Stimulation: Prototype for Muscular Activity Detection

Participants: Guillaume Coppey, David Andreu, David Guiraud.

One way to face centralized Functional Electrical Stimulation (FES) architecture limitations is to distribute electronics close to electrodes. These Distributed Stimulation and Measurement Units (DSU and DMU) are interconnected by a network. Different DSU have been designed and prototyped. We started the design and prototyping of a DMU dedicated to ElectroMyoGramm (EMG) activity reading.

To validate both the digital architecture of the DMU and the digital processing it performs, we prototyped a DMU in charge of muscular activity detection. This DMU is able to detect a threshold crossing on an EMG input signal. The experimental setup is schematically represented on figure 11 , showing also the digital architecture of the prototyped DMU.

This DMU is able to accurately detect EMG activity after filtering and then processing a rolling average. Figure 12 shows intermediate signals: (a) is the absolute value of the filtered EMG signal, (b) is the rolling average, and (c) is the threshold used for activity detection.

This DMU prototype showed that digital processing chain dedicated to EMG activity detection can be embedded within a distributed measurement unit using a programmable logical device (FPGA), like we did for distributed stimulation unit. The embedded architecture of this unit is designed according to a Petri Net based methodology. This allows to exploit effective parallelism offered by FPGA devices, and to reach expected
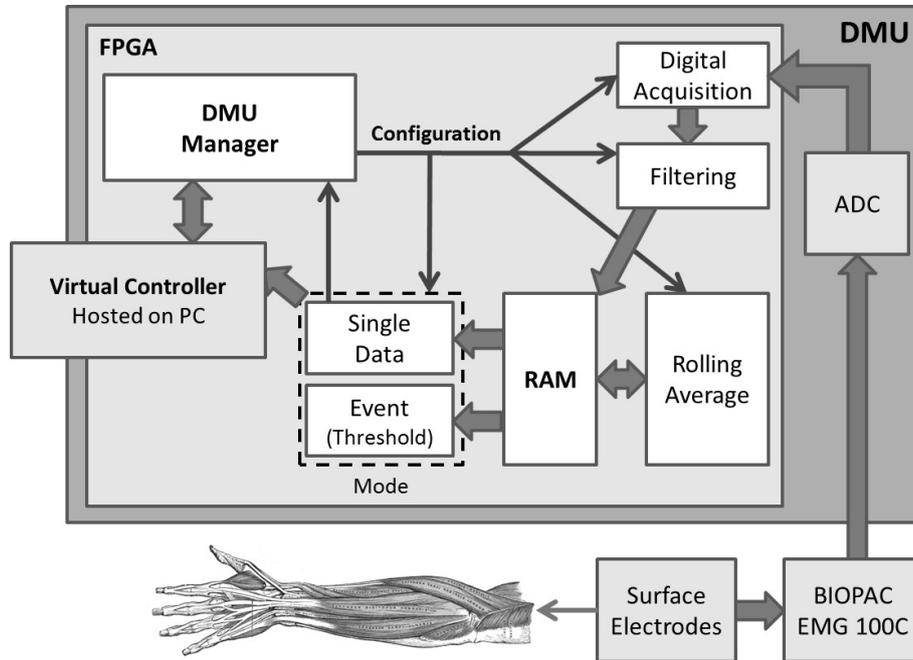
Figure 11. Experimental setup
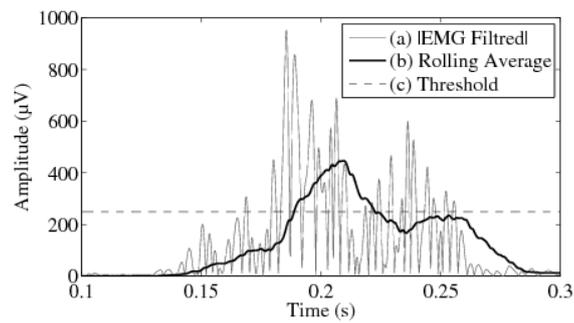
Figure 12. Intermediate signals for muscular activity detection on extensor digitorum communis

performances even at low frequency. The embedded processing chain is configurable and parameters can be adjusted, in order to optimize performance.

Future works will consist in adding the protocol stack to the digital architecture of the DMU, allowing integrating it within our distributed FES architecture. This will allow us to measure effective latencies and other performances from a closed-loop point of view. This work is necessary to ensure that such a distributed EMG activity detection is adequate with FES requirements. After that, we will investigate the trade-off between the global performances versus the implantable device constraints, like its size and power consumption.

6.3.2. *Abstraction and composition for formal design of neuroprotheses*

Participants: Hélène Leroux, David Andreu, Karen Godary.

In the framework of specification and implementation of complex digital systems on FPGA, we have developed an approach based on components whose behavior and composition are specified by generalized interpreted T-time Petri nets. One of the inherent difficulties for designer is to take into account, on the behavioral part, exceptions. This leads often to a complex modeling and is a source of human errors. Indeed, it is intricate to express all the possible situations (i.e. current state of model). We have defined a way to model exception handling by integrating the well-know concept of macroplace into the formalism. The analysability of the model and the efficiency of the implementation on FPGA (reactivity and surface, ie number of logic blocks) have been preserved. An example of macroplace is given in figure 13 ; it contains a sub-net (set of places of its refinement) from which exception handling is simply described by a dedicated output transition (transition t_e on fig. 13), whatever is the current state of the sub-net.
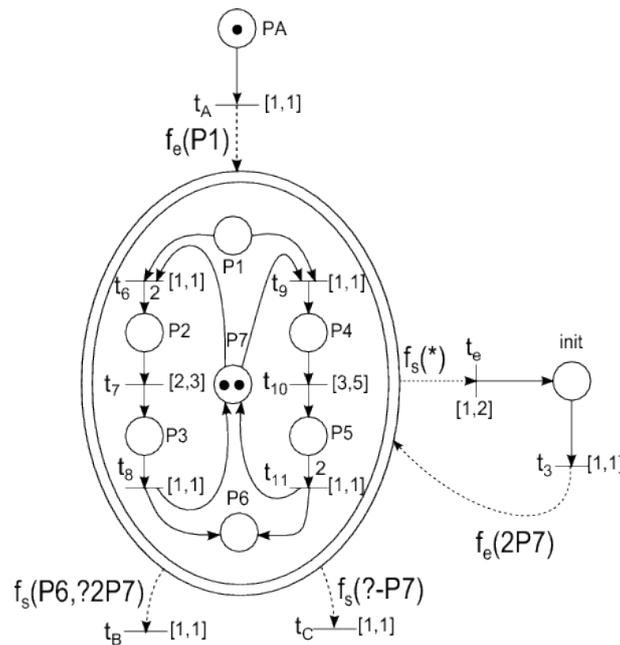
Figure 13. An example of macroplace and exception catching

6.3.3. *Increasing stimulation selectivity using the Transversal Intrafascicular Multichannel Electrode (TIME)*

Participants: Pawel Maciejasz, David Andreu, David Guiraud, Xavier Navarro, Jordi Badia (Universitat Autònoma de Barcelona), Winnie Jensen, Kristian Rauhe Harreby, Aritra Kundu, Bo Geng (Aalborg University), Thomas Stieglitz, Tim Boretius (University of Freiburg), Ken Yoshida (Purdue University Indianapolis).

The electrical stimulation of nerve fibres may allow to restore or augment some body functions lost due to disease or injury. However, in typical peripheral nerves there are thousands of nerve fibres innervating various organs. Therefore, it is necessary to develop interfaces and methods allowing for selective activation of only desired population of nerve fibres. Various neural interfaces have been already proposed for that purpose, including multipolar cuff electrodes, longitudinal intrafascicular electrodes (LIFE) and the Utah Slanted Electrode Array (USEA), all with different selectivity and invasiveness ratios. Recently a new electrode concept of a transversal intrafascicular multichannel electrode (TIME) has been proposed (Fig. 14). This electrode has been developed in frame of the European Project TIME in which the DEMAR team is participating (grant CP-FP-INFSO 224012 from the European Union). It is intended to be implanted transversally in the nerve and address several fascicles or subgroups of nerve fibres with one device. It has longitudinal shape and has several independent stimulation sites equally spread on both sides of the electrode.
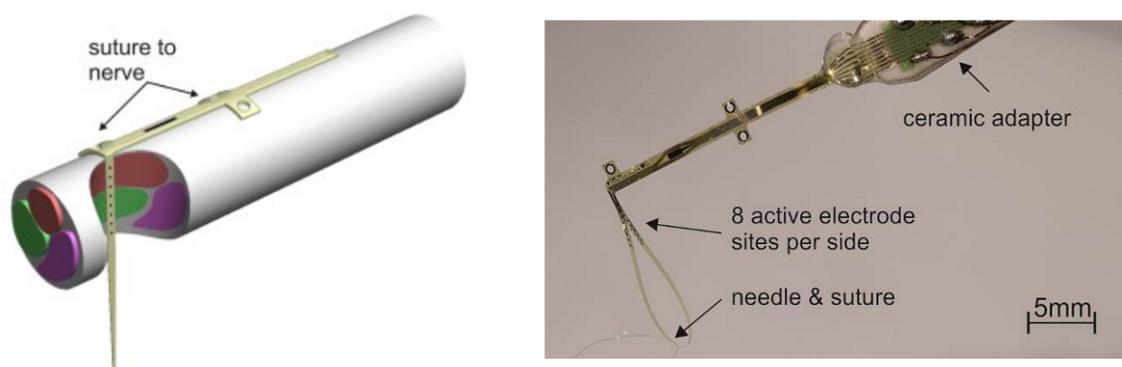
Figure 14. Left: The schematic representation of the TIME electrode implanted into the peripheral nerve. Right: detailed photograph of an TIME-3H electrode with pre-attached loop-thread (Source: Boretius et al. Proc. IEEE Biomedical Robotics and Biomechatronics Conference, Rome 2012)

It has been already shown that TIME allows to achieve high selectivity of stimulation when using monopolar configuration, i.e. when current is delivered through one of the sites of the TIME against small needle electrode placed in the proximity of the nerve. We have performed investigation in the sciatic nerve of rat to verify if the use of bipolar configuration, i.e. when current is delivered through one of the TIME sites against an other site of the same electrode, could allow to further enhance selectivity of stimulation. The results of our studies suggest that using bipolar configuration do allow to increase selectivity of stimulation. However, higher charge of the stimulation may be necessary to achieve similar level of muscle activation, as compared to the monopolar configuration [24].

When applied in the rat model the Transverse Intrafascicular Multi-channel Electrode (TIME) showed selective nerve fascicle recruitment. But results from the larger and poly-fasicular median nerves in pigs indicated that a single TIME could not reach the entire nerve and could only selectively recruit a subset of the nerve fasicles. The use of multiple TIME structures could offer a means to achieve highly selective fascicular stimulation while reaching a larger percentage of the fascicles in the nerve. Therefore we have investigated the use of pairs of TIMEs implanted in the median nerves of anesthesized pigs. TIME structures were implanted at different angles relative to each other or in parallel with one another. Electrical stimuli was passed through each contact of each TIME and the resulting electromyograms were recorded from seven muscles innervated

by the median nerve. The ability to recruit these muscles was used to assess the stimulation selectivity of each contact using a selectivity index comparing the root-mean-square of the the evoked EMG of individul muscles. Results showed a signifcant increase in the selectivity index, when using two TIMEs compared to one. The optimal improvement was obeserved when TIMEs were placed in parallel to each other in such a way that they interfaced non-overlapping nerve regions [18].

6.3.4. *Nerve model for ENG recording*

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

In the context of selective electroneurogram recording, we showed last year the efficiency of a *small tripole* filtering (the distance between contacts is $375\,\mu$m) thanks to simulated signals. This recording is locally sensitive and greatly increases the selectivity of the electrode. This year, we realized an experiment to verify the simulated results. Theoritical study of the *small tripole* sensitivity was realized for a single-fiber action potential (SFAP). We wanted to proceed in the same way for the experiment by trying to measure a SFAP.

However, actual biological SFAP would be hardly measurable by a *small tripole* in an in-vitro experiment. So, we decided to choose an approach based on an artificial axon. In this artificial model, every parameter is under control. The position of the fiber, the nodes of Ranvier, the position of the measuring electrode, as well as the involved currents are perfectly known. This allows us to implement and to estimate with accuracy the filtering realized by the *small tripole* with exactly the same configurations as in simulations. And by performing measures for several radial distances, we can verify the influence of this distance, in order to estimate the sensibility the *small tripole*. Moreover, we can increase the activity amplitude to be higher than in real fiber and then achieve a beneficial signal to noise ratio.

6.3.4.1. Method
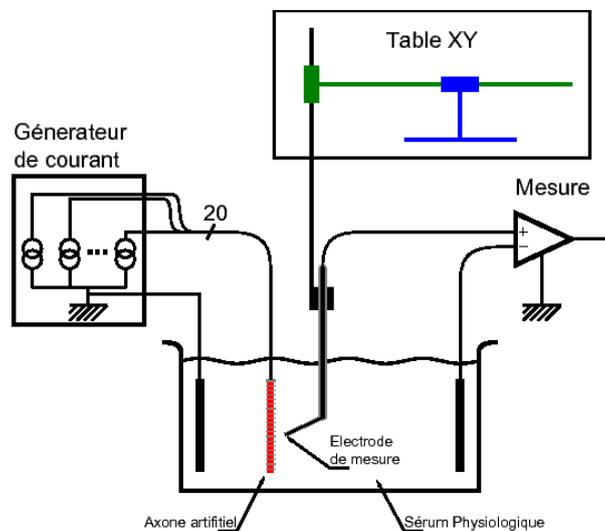
Figure 15. Principle of the experimental setup using the artificial axon.

The experimental setup consists in an artificial model emulating an axon and in a system measuring the action potential across the surrounding medium, as sketched in Fig. 15 . Biological tissues are modeled by a saline solution having a conductivity close to the human body. The space and time behavior of the current generated by a natural axon is reproduced on the artificial axon. The latter is emulated by an cochlear electrode. It exhibits 20 contacts, being an accurate image of the nodes of Ranvier on an axon of $8.7\,\mu$m diameter regarding the spatial periodicity.

In order to generate electric activity on several contacts that can be compared to the one of several nodes of Ranvier during the conduction of an action potential, we realized a custom multi-current generator with asynchronous outputs. The chosen amplitude was $35 \cdot 10^4$ times that of an human axon.

The measure of the SFAP in the space, is done using a punctual monopolar electrode. The position of the measuring electrode relative to the fiber is automatically set by a programmable micromanipulator. Then, the measure is repeated every $50 \, \mu$m along $2 \, mm$ on the radial axis.

6.3.4.2. Monopolar and tripolar sensitivity
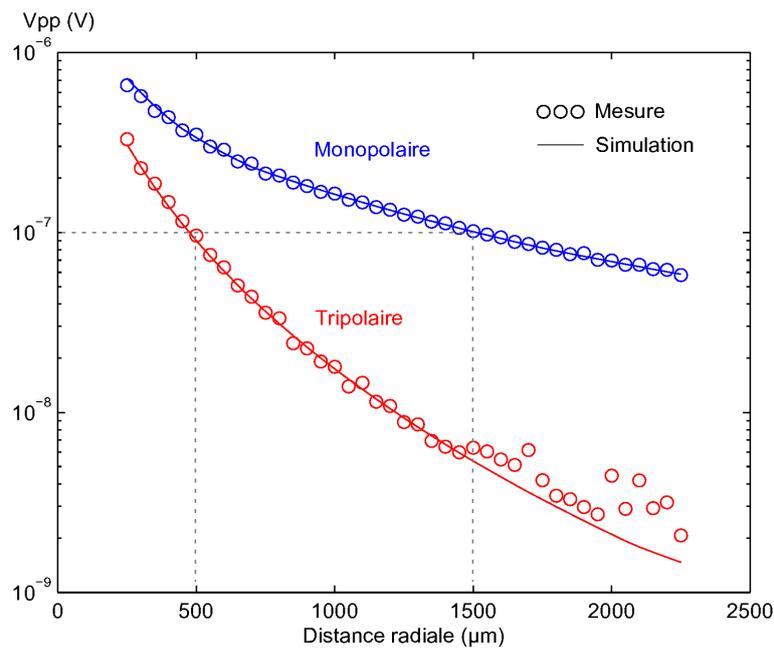
Figure 16. Comparison between monopolar and small-tripolar recording sensitivities. Measures are plotted with circles and simulation results are solid lines. The monopolar data are in blue and tripolar are in red.

To verify quantitatively the sensibility results obtained by simulation for monopolar and tripolar electrodes, we have first estimated the amplitude of a SFAP measured by a monopolar electrode. The measured SFAP amplitude directly gives the monopolar recording that varies according to the radial distance between the artificial axon and the electrode. Then, the *small-tripole* recording results from an off-line process combining three points of measure spaced out of $375 \, \mu$m. The two kind of recordings are represented in Fig. 16 .

By comparing the recording sensitivities either estimated or measured, we can conclude that measures perfectly fit the theoretical results. We can also notice that the attenuation relative to the radial distance is more important for the *small tripole* than for the monopolar measure. This confirms that the measure performed by one *small tripole* electrode is exclusively sensitive to the closest fibers.

<div align="center">

FOCUS Project-Team

</div>

6. New Results

6.1. Service-oriented computing

Participants: Mila Dalla Preda, Ornela Dardha, Maurizio Gabbrielli, Elena Giachino, Claudio Guidi, Jacopo Mauro, Fabrizio Montesi, Davide Sangiorgi.

6.1.1. Primitives

In the context of Service-Oriented Architectures (SOAs), the integration of services is an important aspect that is usually addressed by using specific tools, such as Enterprise Service Bus (ESB). Although widely used these ad-hoc solutions do not exploit the possibility of using a mechanism of interface extension to foster the rapid prototyping and deployment of in-the-middle entities that compose services abstracting from the order in which they exchange messages. We have proposed [29] a framework to perform service integration, based on the extension of service interfaces, capturing a class of service integrators that are decoupled from the services they integrate in an SOA. We also provide a reference implementation for the primitive of service integration by extending the Jolie language, thus allowing for the experimentation with real SOA scenarios. We have shown [30] how our methodology differs from the standard practice with ESB.

6.1.2. Contracts and sessions

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.

Contracts specify the expected dialogue in a session and they can be expressed as types, usually called *session types* in this context.

A session type describes communication by specifying type and direction of data exchanged between two parties. When session types and session primitives are added to the syntax of the types and terms of a language, they give rise to additional separate syntactic categories. As a consequence, there may be duplication of efforts in the theory: the proofs of properties must be checked both on ordinary types and on session types. We have shown [32] that this duplication is not necessary, by exhibiting an encoding of (dyadic) session types into ordinary types. Using the encoding, the properties of session types are derived as straightforward corollaries.

We have also studied the problem of handling unexpected or unwanted conditions in sessions, that may change the default execution of distributed communication protocols. We have proposed [14] a global escape mechanism; it can handle such events while preserving compatibility of multiparty conversations. This flexibility enables us to model complex exceptions such as criss-crossing global interactions and error handling for distributed cooperating threads. Guided by multiparty session types, our semantics is proven to provide a termination algorithm for global escapes, as well as further safety properties, such as progress within the session and atomicity of escapes with respect to the subset of involved participants.

6.2. Adaptability and faults

Participants: Mario Bravetti, Elena Giachino, Ivan Lanese, Michael Lienhardt, Jacopo Mauro, Davide Sangiorgi, Gianluigi Zavattaro.

6.2.1. *Reversibility*

We have continued the study of reversibility started in the past years, aimed at developing programming abstractions for reliable distributed systems. We have shown [39] preliminary results on the interplay between reversibility and compensations, which are a main ingredient in many existing techniques for reliability, in particular long running transactions.

We have then applied [43] our reversibility theory to μOz, a concurrent programming language defined by a stack-based abstract machine, and we make it reversible. This is a first step towards the definition of reversible variants of more complex languages. As additional result we show that the memory overhead due to reversibility is optimal as an order of magnitude.

6.2.2. *Primitives for adaptable and evolving components*

We study primitives for adaptable and evolving components both in an abstract algebraic setting and in a more concrete setting based on the ABS object-oriented language.

We have defined [13] adaptable processes, a concurrent higher-order calculus where processes have a location, and are sensible to actions of update at runtime. This allows us to express a wide range of evolvability patterns. We have also defined [24] a temporal logic over adaptable processes, with examples of the expressiveness of the logic and of its significance in relation to the calculus of adaptable processes.

A different direction has focused on ABS, a concurrent object-oriented language based on futures for asynchronous method invocations and on object groups for concurrency control. We have given [38] an overview of the architectural aspects of ABS: a feature-driven development workflow, a formal notion of deployment components for specifying environmental constraints, and a dynamic component model that is integrated into the language. We have employed an industrial case study to demonstrate how the various aspects work together in practice. In [40] we have focused on the component model and studied techniques allowing safe dynamic reconfiguration. Our approach adds to ABS: i) output ports to represent variability points, ii) critical sections to control when updates of the software can be made and iii) hierarchy to model locations and distribution. These different notions work together to allow dynamic and safe update of a system.

6.2.3. *Reconfigurability in the cloud*

The cloud is a relevant application domain for FOCUS. We have considered [35] the problem of deploying and (re)configuring resources in a cloud setting, where interconnected software components and services can be deployed on clusters of heterogeneous (virtual) machines that can be created and connected on-the-fly. We introduce a component model to capture similar scenarii from realistic cloud deployments, and instrument automated planning of day-to-day activities such as software upgrade planning, service deployment, elastic scaling, etc. We formalize the model and characterize the feasibility and complexity of configuration achievability.

6.2.4. *Delta-Oriented Programming and Software Product Lines*

Delta-oriented programming (DOP) provides a technique for implementing Software Product Lines based on modifications (add, remove, modify) to a core program. Unfortunately, such modifications can introduce errors into a program, especially when type signatures of classes are modified in a non-monotonic fashion. To deal with this problem in we have designed [42] a type system for delta-oriented programs based on row polymorphism. This exercise elucidates the close correspondence between delta-oriented programs and row polymorphism.

In [41] we have studied the notion of conflict for a variant of DOP without features, separating out the notions of hard and soft conflict. Specifically, we have defined a language for this subset of DOP and give a precise, formal definition of these notions. We have then extended the type system in [42] to ensure that the computation of a well-typed product will always succeed and has an unambiguous result.

6.3. Resource Control

Participants: Michele Alberti, Ugo Dal Lago, Marco Gaboardi, Daniel Hirschkoff, Simone Martini, Paolo Parisen Toldin, Giulio Pellitta, Barbara Petit, Davide Sangiorgi, Marco Solieri.

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.

During 2012, we have continued our work on intensionally complete techniques for the complexity analysis of functional programs. In [15] a relatively complete type system from which complexity of call-by-name terms has been introduced, while in [25] the same approach is used in a call-by-value setting. The introduced methodology allows us to reduce the problem at hand to the verification of a set of first-order proof-obligations. No information is lost along the reduction.

The interpretation method has been the object of a couple of investigations. On the one hand, we have proved a necessary condition for a first-order program to admit a quasi-interpretation [12]: it must be blind, namely it must be somehow insensible to its argument value, but only sensible to their length. Moreover, we have introduced a new methodology for the complexity analysis of higher-order programs based on an higher-order generalizations of ordinary polynomial interpretations and quasi-interpretations [23].

Among the most foundational works in this area, we should also mention those about invariance results on cost models [16], [22]: we proved that in many different cases, the number of beta-reduction steps is an adequate cost-model for the lambda calculus. These results are potentially useful for complexity analysis, in that they show that a natural and quite intuitive cost model is indeed reasonable, meaning that evaluation can be simulated by finer-grained models of computation within a polynomial overhead.

Finally, some of our works have to do with the semantics of various sorts of lambda calculi with linearity constraints: a non-deterministic extension of the call-by-value lambda calculus, which corresponds to the additive fragment of the linear-algebraic lambda-calculus [36]; a lambda calculus with constructors that decomposes the pattern matching à la ML into some atomic rules [44]; a categorical approach to model the programming language SlPCF that has been conceived in order to program only linear functions between Coherence Spaces [20].

We have also continued our work on techniques for ensuring termination of programs, studying [34] 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.

6.4. Verification of extensional properties

Participants: Mario Bravetti, Daniel Hirschkoff, Cosimo Laneve, Jean-Marie Madiot, 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. A substantial amount of the work carried out this year has to do with the transfer of techniques from the area of concurrency theory to the investigation of properties in adaptable systems, obiect-oriented concurrent systems, and systems based on specific synchronization mechanisms.

6.4.1. *Adaptability*

We mentioned earlier the process calculus of adaptable processes [13] and the related temporal logic [24]. In the same papers, we have addressed the (un)decidability of two safety properties related to error occurrence, and we have explained how the proof techniques in [13] can be extended to prove (un)decidability results for the temporal logic.

6.4.2. Object-orientation

We have considered concurrent object-oriented languages with futures and cooperative scheduling. Verification of deadlock in such systems is a nontrivial task due to the dynamic and unbounded creation of futures. We have introduced [45] a technique to prove deadlock freedom for such systems by translating a concrete program to an abstract version of the program, and then encoding such abstract program into a Petri net. Deadlock can be detected on Petri nets via checking the reachability of a distinct marking: absence of deadlocks in the Petri net constitutes deadlock freedom of the concrete system.

6.4.3. Synchronization primitives

We have investigated [33] the impact of node and communication failures on the decidability and complexity of parametric verification of a formal model of ad hoc networks, in which finite state processes communicate via selective broadcast. We have considered three possible kinds of node failures –intermittence, restart, and crash– and three cases of communication failures –nondeterministic message loss, message loss due to conflicting emissions, and detectable conflicts. Interestingly, we have proved that the considered decision problem (reachability of a control state) is decidable for node intermittence and message loss (either nondeterministic or due to conflicts) while it turns out to be undecidable for node restart/crash, and conflict detection. The conclusion is that verification is decidable only when processes are unaware of the occurrence of a failure.

In another line of work, we have studied the impact of dualities and symmetries in synchronization primitives for message-passing processes [37]. We have shown that in languages where input and output are dualisable (e.g., variants of the π-calculus such as πI and fusion), duality breaks with the addition of ordinary input/output types. We have then considered the minimal symmetrical conservative extension of π-calculus input/output types. We have proved duality properties for it. As example of application of the dualities, we have used this language to relate two encodings of λ-calculus, by Milner and by van Bakel and Vigliotti, syntactically quite different from each other. Thus, results on one encoding can be transferred onto the other one.

6.4.4. Coinduction

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 [48], [51] two comprehensive textbooks on bisimulation and coinduction (in [51], Sangiorgi is an editor, and author of two chapter contributions [49], [47]). 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.

6.5. Expressiveness of computational models

Participants: Cosimo Laneve, Maurizio Gabbrielli, Gianluigi Zavattaro.

Expressiveness refers to the study of the expressive power of computational models.

We have studied [46] the expressiveness of an actor-based language similar to the language ABS developed in Hats. We have identified the presence/absence of fields as a relevant feature: the dynamic creation of names in combination with fields gives rise to Turing completeness. On the other hand, restricting to stateless actors gives rise to systems for which properties such as termination are decidable. Such decidability result holds in actors with states when the number of actors is finite and the state is read-only.

Our other study of expressiveness has been made on Constraint Handling Rules (CHR), a committed-choice declarative language originally designed for writing constraint solvers and that is nowadays a general purpose language. The study of CHR is interesting within Focus as this kind of languages, having both constraint solving and concurrency features, allow us to express in a natural way quantitative aspects related to resources. Moreover, constraints may be used to describe dynamic adaptation and evolution of systems. CHR programs

consist of multi-headed guarded rules which allow one to rewrite constraints into simpler ones until a solved form is reached. Many empirical evidences suggest that multiple heads augment the expressive power of the language (somehow, it can be considered similar to multiple synchronization gathering n processes simultaneously), however no formal result in this direction had been proved so far. We have proved [18] a number of expressiveness results to support such a claim. First, we have analyzed the Turing completeness of CHR with respect to the underlying constraint theory. If the constraint theory is powerful enough then restricting to single head rules does not affect Turing completeness. On the other hand, differently from the case of the multi-headed language, the single head CHR language is not Turing powerful when the underlying signature (for the constraint theory) does not contain function symbols. Then we have proved that, no matter which constraint theory is considered, under some reasonable assumptions it is not possible to encode the CHR language (with multi-headed rules) into a single headed language while preserving the semantics of the programs. Moreover, under some stronger assumptions, considering an increasing number of atoms in the head of a rule augments the expressive power of the language.

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

6. New Results

6.1. Algebraic representations for geometric modeling

6.1.1. *Fitting ideals and multiple-points of surface parameterizations*

Participants: Nicolàs Botbol, Laurent Busé.

Given a birational parameterization ϕ of an algebraic surface $\mathcal{S}$ in the projective space $\mathbb{P}^3$, the purpose of this ongoing work is to investigate the sets of points $D_k(\phi)$ on $\mathcal{S}$ whose preimage consists in k or more points, counting multiplicity. Our main result is an explicit description of these algebraic sets $D_k(\phi)$ in terms of Fitting ideals of some graded parts of a symmetric algebra associated to the parameterization ϕ.

This work is done in collaboration with Marc Chardin (University Pierre et Marie Curie).

6.1.2. *Algebraic geometry tools for the study of entanglement: an application to spin squeezed states*

Participant: Alessandra Bernardi.

In [18] a short review of Algebraic Geometry tools for the decomposition of tensors and polynomials is given from the point of view of applications to quantum and atomic physics. Examples of application to assemblies of indistinguishable two-level bosonic atoms are discussed using modern formulations of the classical Sylvester's algorithm for the decomposition of homogeneous polynomials in two variables. In particular, the symmetric rank and symmetric border rank of spin squeezed states is calculated as well as their Schrödinger-cat-like decomposition as the sum of macroscopically different coherent spin states; Fock states provide an example of states for which the symmetric rank and the symmetric border rank are different.

This is a joint work with I. Carusotto (University of Trento, Italy).

6.1.3. *A partial stratification of secant varieties of Veronese varieties via curvilinear subschemes.*

Participant: Alessandra Bernardi.

In [11] we give a partial quasi-stratification of the secant varieties of the order d Veronese variety $X_{m,d}$ of $\mathbb{P}^m$. It covers the set $\sigma_t(X_{m,d})^\dagger$ of all points lying on the linear span of curvilinear subschemes of $X_{m,d}$, but two quasi-strata may overlap. For low border rank, two different quasi-strata are disjoint and we compute the symmetric rank of their elements. Our tool is the Hilbert schemes of curvilinear subschemes of Veronese varieties. To get a stratification we attach to each $P \in \sigma_t(X_{m,d})^\dagger$ the minimal label of a quasi-stratum containing it.

This is a joint work with E. Ballico (University of Trento, Italy).

6.1.4. *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 0 and suppose that F belongs to the s-th secant variety of the d-uple Veronese embedding of $\mathbb{P}^m$ into $\mathbb{P}^{\binom{m+d}{d}-1}$ but that its minimal decomposition as a sum of d-th powers of linear forms $M_1, ..., M_r$ is $F = M_1^d + \cdots + M_r^d$ with $r > s$. In [12], we show that if $s + r \leq 2d + 1$ then such a decomposition of F can be split into 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 r is at most d and a mild condition is satisfied.

This is a joint work with E. Ballico (University of Trento, Italy).

6.1.5. Higher secant varieties of $\mathbb{P}^n \times \mathbb{P}^1$ embedded in bi-degree (a, b)

Participant: Alessandra Bernardi.

In [15], we compute the dimension of all the higher secant varieties to the Segre-Veronese embedding of $\mathbb{P}^n \times \mathbb{P}^1$ via the section of the sheaf $\mathcal{O}(a, b)$ for any $n, a, b \in \mathbb{Z}^+$. We relate this result to the Grassmann Defectivity of Veronese varieties and we classify all the Grassmann $(1, s-1)$-defective Veronese varieties.

This is a joint work with E. Ballico, M. V. Catalisano (University of Trento, Italy).

6.1.6. 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 $\tau(X_{m,d}, t) \subseteq \mathbb{P}^N$, be the join of $\tau(X_{m,d})$ and $t-2$ copies of $X_{m,d}$. In [13], we prove that if $m \geq 2$, $d \geq 7$ and $t \leq 1 + \lfloor \binom{m+d-2}{m} / (m+1) \rfloor$, then for a general $P \in \tau(X_{m,d}, t)$ there are uniquely determined $P_1, \cdots, P_{t-2} \in X_{m,d}$ and a unique tangent vector ν of $X_{m,d}$ such that P is in the linear span of $\nu \cup \{P_1, \cdots, P_{t-2}\}$. In other words, a degree d linear form f (a symmetric tensor T of order d) associated to P may be written as

$$f = L_{t-1}^{d-1} L_t + \sum_{i=1}^{t-2} L_i^d, \quad (T = v_{t-1}^{\otimes(d-1)} v_t + \sum_{i=1}^{t-2} v_i^{\otimes d})$$

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 E. Ballico (University of Trento, Italy).

6.1.7. General tensor decomposition, moment matrices and applications.

Participants: Alessandra Bernardi, Bernard Mourrain.

In [17] the tensor decomposition addressed may be seen as a generalisation of Singular Value Decomposition of matrices. We consider general multilinear and multihomogeneous tensors. We show how to reduce the problem to a truncated moment matrix problem and give a new criterion for flat extension of Quasi-Hankel matrices. We connect this criterion to the commutation characterisation of border bases. A new algorithm is described. It applies for general multihomogeneous tensors, extending the approach of J.J. Sylvester to 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 J. Brachat and P. Comon (i3S, CNRS).

6.1.8. On the cactus rank of cubic forms

Participant: Alessandra Bernardi.

In this work, we prove that the smallest degree of an apolar 0-dimensional scheme of a general cubic form in $n+1$ variables is at most $2n+2$, when $n \geq 8$, and therefore smaller than the rank of the form. For the general reducible cubic form the smallest degree of an apolar subscheme is $n+2$, while the rank is at least $2n$.

This is a joint work with K. Ranestad (University of Oslo, Norway) that will be published in 2013 in the Journal of Symbolic Computation. The preprint is available at http://hal.inria.fr/inria-00630456.

6.1.9. Tensor ranks on tangent developable of Segre varieties

Participant: Alessandra Bernardi.

In [14] we describe the stratification by tensor rank of the points belonging to the tangent developable of any Segre variety. We give algorithms to compute the rank and a decomposition of a tensor belonging to the secant variety of lines of any Segre variety. We prove Comon's conjecture on the rank of symmetric tensors for those tensors belonging to tangential varieties to Veronese varieties.

This is a joint work with E. Ballico (University of Trento, Italy).

6.1.10. On the dimension of spline spaces on planar T-meshes
Participant: Bernard Mourrain.

In [33], we analyze the space of bivariate functions that are piecewise polynomial of bi-degree $\leq (m, m')$ and of smoothness r along the interior edges of a planar T-mesh. We give new combinatorial lower and upper bounds for the dimension of this space by exploiting homological techniques. We relate this dimension to the weight of the maximal interior segments of the T-mesh, defined for an ordering of these maximal interior segments. We show that the lower and upper bounds coincide, for high enough degrees or for hierarchical T-meshes which are enough regular. We give a rule of subdivision to construct hierarchical T-meshes for which these lower and upper bounds coincide. Finally, we illustrate these results by analyzing spline spaces of small degrees and smoothness.

6.1.11. On the problem of instability in the dimension of a spline space over a T-mesh
Participant: Bernard Mourrain.

In [23], we discuss the problem of instability in the dimension of a spline space over a T-mesh. For bivariate spline spaces $S(5, 5, 3, 3)$ and $S(4, 4, 2, 2)$, the instability in the dimension is shown over certain types of T-meshes. This result could be considered as an attempt to answer the question of how large the polynomial degree (m, m') should be relative to the smoothness (r, r') to make the dimension of a spline space stable. We show in particular that the bound $m \geq 2r + 1$ and $m' \geq 2r' + 1$ are optimal.

This is a joint work with Berdinsky Dmitry, Oh Min-Jae and Kim Taewan (Department of Naval Architecture and Ocean Engineering Seoul National University, South Korea).

6.1.12. Homological techniques for the analysis of the dimension of triangular spline spaces
Participant: Bernard Mourrain.

The spline space $C_k^r(\Delta)$ attached to a subdivided domain Δ of $\mathbb{R}^d$ is the vector space of functions of class C^r which are polynomials of degree $\leq k$ on each piece of this subdivision. Classical splines on planar rectangular grids play an important role in Computer Aided Geometric Design, and spline spaces over arbitrary subdivisions of planar domains are now considered for isogeometric analysis applications. In [34], we address the problem of determining the dimension of the space of bivariate splines $C_k^r(\Delta)$ for a triangulated region Δ in the plane. Using the homological introduced by Billera (1988), we number the vertices and establish a formula for an upper bound on the dimension. There is no restriction on the ordering and we obtain more accurate approximations to the dimension than previous methods. Furthermore, in certain cases even an exact value can be found. The construction makes it also possible to get a short proof for the dimension formula when $k \geq 4r + 1$, and the same method we use in this proof yields the dimension straightaway for many other cases.

This is a joint work with Nelly Villamizar (CMA, University of Oslo, Norway).

6.1.13. Analysis-suitable volume parameterization of multi-block computational domain in isogeometric applications
Participants: Bernard Mourrain, André Galligo.

Parameterization of computational domain is a key step in isogeometric analysis just as mesh generation is in finite element analysis. In [36], we study the volume parameterization problem of multi-block computational domain in isogeometric version, i.e., how to generate analysis-suitable parameterization of the multi-block computational domain bounded by B-spline surfaces. Firstly, we show how to find good volume parameterization of single-block computational domain by solving a constraint optimization problem, in which the constraint condition is the injectivity sufficient conditions of B-spline volume parametrization, and the optimization term is the minimization of quadratic energy functions related to the first and second derivatives of B-spline volume parameterization. By using this method, the resulted volume parameterization has no self-intersections, and the isoparametric structure has good uniformity and orthogonality. Then we extend this method to the multi-block case, in which the continuity condition between the neighbor B-spline volume should be added to the constraint term. The effectiveness of the proposed method is illustrated by several examples based on three-dimensional heat conduction problem.

This is a joint work with Régis Duvigneau (Inria, EPI OPALE) and Xu Gang (College of computer - Hangzhou Dianzi University, China).

6.1.14. A new error assessment method in isogeometric analysis of 2D heat conduction problems
Participants: Bernard Mourrain, André Galligo.

In [35], we propose a new error assessment method for isogeometric analysis of 2D heat conduction problems. A posteriori error estimation is obtained by resolving the isogeometric analysis problem with several k-refinement steps. The main feature of the proposed method is that the resulted error estimation surface has a B-spline form, according to the main idea of isogeometric analysis. Though the error estimation method is expensive, it can be used as an error assessment method for isogeometric analysis. Two comparison examples are presented to show the efficiency of the proposed method.

This is a joint work with Régis Duvigneau (Inria, EPI OPALE) and Xu Gang (College of computer - Hangzhou Dianzi University, China).

6.1.15. On the cut-off phenomenon for the transitivity of randomly generated subgroups
Participant: André Galligo.

Consider $K \geq 2$ independent copies of the random walk on the symmetric group S_N starting from the identity and generated by the products of either independent uniform transpositions or independent uniform neighbor transpositions. At any time $n \in \mathbb{N}$, let G_n be the subgroup of S_N generated by the K positions of the chains. In the uniform transposition model, we prove in [28] that there is a cut-off phenomenon at time $N \ln(N)/(2K)$ for the non-existence of fixed point of G_n and for the transitivity of G_n, thus showing that these properties occur before the chains have reached equilibrium. In the uniform neighbor transposition model, a transition for the non-existence of a fixed point of G_n appears at time of order $N^{1+\frac{2}{K}}$ (at least for $K \geq 3$), but there is no cut-off phenomenon. In the latter model, we recover a cut-off phenomenon for the non-existence of a fixed point at a time proportional to N by allowing the number K to be proportional to $\ln(N)$. The main tools of the proofs are spectral analysis and coupling techniques.

This is a joint work with Laurent Miclo (University of Toulouse).

6.2. Algebraic algorithms for geometric computing

6.2.1. On the isotopic meshing of an algebraic implicit surface
Participant: Bernard Mourrain.

In [22], we present a new and complete algorithm for computing the topology of an algebraic surface given by a squarefree polynomial in $\mathbb{Q}[X, Y, Z]$. Our algorithm involves only subresultant computations and entirely relies on rational manipulation, which makes it direct to implement. We extend the work in [15], on the topology of non-reduced algebraic space curves, and apply it to the polar curve or apparent contour of the surface S. We exploit simple algebraic criterion to certify the pseudo-genericity and genericity position of the surface. This gives us rational parametrizations of the components of the polar curve, which are used to lift the topology of the projection of the polar curve. We deduce the connection of the two-dimensional components above the cell defined by the projection of the polar curve. A complexity analysis of the algorithm is provided leading to a bound in $\widetilde{\mathcal{O}}_B(d^{15}\tau)$ for the complexity of the computation of the topology of an implicit algebraic surface defined by integer coefficients polynomial of degree d and coefficients size τ. Examples illustrate the implementation in Mathemagix of this first complete code for certified topology of algebraic surfaces.

This is a joint work with Daouda Niang Diatta, Olivier Ruatta (XLIM, University of Limoges).

6.2.2. *Moment matrices, border basis and real radical computation*
Participant: Bernard Mourrain.

In [32], we describe new methods to compute the radical (resp. real radical) of an ideal, assuming its complex (resp. real) variety is finite. The aim is to combine approaches for solving a system of polynomial equations with dual methods which involve moment matrices and semi-definite programming. While border basis algorithms are efficient and numerically stable for computing complex roots, algorithms based on moment matrices allow the incorporation of additional polynomials, e.g., to restrict the computation to real roots or to eliminate multiple solutions. The proposed algorithm can be used to compute a border basis of the input ideal and, as opposed to other approaches, it can also compute the quotient structure of the (real) radical ideal directly, i.e., without prior algebraic techniques such as Gröbner bases. It thus combines the strengths of existing algorithms and provides a unified treatment for the computation of border bases for the ideal, the radical ideal and the real radical ideal.

This is a joint work with Jean-Bernard Lasserre (LAAS, Toulouse), Monique Laurent (CWI, Amsterdam, Netherland), Philipp Rostalski (University of California, Berkeley, US) Philippe Trébuchet (APR, LIP6, Paris).

6.2.3. *On the computation of matrices of traces and radicals of ideals*
Participant: Bernard Mourrain.

Let $f_1, ..., f_s \in \mathbb{K}[x_1, ..., x_m]$ be a system of polynomials generating a zero-dimensional ideal I, where $\mathbb{K}$ is an arbitrary algebraically closed field. In [31], we study the computation of "matrices of traces" for the factor algebra $\mathcal{A} := \mathbb{C}[x_1, ..., x_m]/I$, i.e. matrices with entries which are trace functions of the roots of I. Such matrices of traces in turn allow us to compute a system of multiplication matrices $\{M_{x_i} | i = 1, ..., m\}$ of the radical $\sqrt{I}$. We first propose a method using Macaulay type resultant matrices of $f_1, ..., f_s$ and a polynomial J to compute moment matrices, and in particular matrices of traces for $\mathcal{A}$. Here J is a polynomial generalizing the Jacobian. We prove bounds on the degrees needed for the Macaulay matrix in the case when I has finitely many projective roots in $\mathbb{P}^m_{\mathbb{C}}$. We also extend previous results which work only for the case where $\mathcal{A}$ is Gorenstein to the non-Gorenstein case. The second proposed method uses Bezoutian matrices to compute matrices of traces of $\mathcal{A}$. Here we need the assumption that $s = m$ and $f_1, ..., f_m$ define an affine complete intersection. This second method also works if we have higher dimensional components at infinity. A new explicit description of the generators of $\sqrt{I}$ are given in terms of Bezoutians.

This is a joint work with Itnuit Janovitz-Freireich (Departamento de Matemáticas, Mexico), Lajos Ron](Hungarian Academy of Sciences and Budapest University of Technology and Economics, Budapest), Agnes Szanto (Department of Computer Science, North Carolina State University, US).

6.2.4. *Border basis representation of a general quotient algebra*
Participant: Bernard Mourrain.

In [40], we generalized the construction of border bases to non-zero dimensional ideals for normal forms compatible with the degree, tackling the remaining obstacle for a general application of border basis methods. First, we give conditions to have a border basis up to a given degree. Next, we describe a new stopping criterion to determine when the reduction with respect to the leading terms is a normal form. This test based on the persistence and regularity theorems of Gotzmann yields a new algorithm for computing a border basis of any ideal, which proceeds incrementally degree by degree until its regularity. We detail it, prove its correctness, present its implementation and report some experimentations which illustrate its practical good behavior.

This is a joint work with Philippe Trébuchet (APR, LIP6, Paris).

6.2.5. *Voronoï diagrams of algebraic distance fields*
Participant: Bernard Mourrain.

In [25], we design and implement an efficient and certified algorithm for the computation of Voronoi Diagrams (VD's) constrained to a given domain. Our framework is general and applicable to any VD-type where the distance field is given explicitly or implicitly by a polynomial, notably the anisotropic VD or VD's of non-punctual sites. We use the Bernstein form of polynomials and DeCasteljau's algorithm to subdivide the initial domain and isolate bisector, or domains that contain a Voronoi vertex. The efficiency of our algorithm is due to a filtering process, based on bounding the field over the subdivided domains. This allows us to exclude functions (thus sites) that do not contribute locally to the lower envelope of the lifted diagram. The output is a polygonal description of each Voronoi cell, within any user-defined precision, isotopic to the exact VD. Correctness of the result is implied by the certified approximations of bisector branches, which are computed by existing methods for handling algebraic curves. First experiments with our C++ implementation, based on double precision arithmetic, demonstrate the adaptability of the algorithm.

This is a joint work with Ioannis Emiris (ERGA, National Kapodistrian University of Athens, Greece), Angelos Mantzaflaris (RICAM, Austrian Academy of Sciences, Austria).

6.2.6. *Rational invariants of scalings from Hermite normal forms*
Participant: Evelyne Hubert.

Scalings form a class of group actions that have both theoretical and practical importance. A scaling is accurately described by an integer matrix. In [39] tools from linear algebra are exploited to compute a minimal generating set of rational invariants, trivial rewriting and rational sections for such a group action. The primary tools used are Hermite normal forms and their unimodular multipliers. With the same line of ideas, a complete solution to the scaling symmetry reduction of a polynomial system is also presented.

This is joint work with George Labahn (University of Waterloo, Canada).

6.2.7. *Scaling invariants and symmetry reduction of dynamical systems*
Participant: Evelyne Hubert.

The motivation for this subject is to offer an algorithmic scheme for reducing the number of parameters in physical, chemical or biological models. This comes as a special case of a symmetry reduction scheme that can be fully realized by linear algebra over the integers. See http://hal.inria.fr/hal-00668882. We provide there the algebraic determination of the scaling symmetry of a dynamical system and an complete explicit symmetry reduction scheme with polynomial complexity.

This is joint work with George Labahn (University of Waterloo, Canada).

6.2.8. *A computational approach to the discriminant of homogeneous polynomials*
Participant: Laurent Busé.

In this work, the discriminant of homogeneous polynomials is studied in two particular cases: a single homogeneous polynomial and a collection of $n-1$ homogeneous polynomials in n variables. In these two cases, the discriminant is defined over a large class of coefficient rings by means of the resultant. Many formal properties and computational rules are provided and the geometric interpretation of the discriminant is investigated over a general coefficient ring, typically a domain.

This work is done in collaboration with Jean-Pierre Jouanolou (University of Strasbourg). A preprint is available at http://hal.inria.fr/hal-00747930/en/.

6.2.9. Intersection between rational curves and surfaces by means of matrix representations
Participant: Laurent Busé.

In [37], we propose a survey of matrix representations for parameterized curves and surfaces. Illustrations of the properties of these representations are given for intersection problems. In particular, we focus on the ray/surface intersection which is an important step in ray-tracing algorithms.

6.2.10. A root isolation algorithm for sparse univariate polynomials
Participant: André Galligo.

In [38], we consider a univariate polynomial f with real coefficients having a high degree N but a rather small number $d + 1$ of monomials, with $d \ll N$. Such a sparse polynomial has a number of real roots smaller or equal to d. Our target is to find for each real root of f an interval isolating this root from the others. The usual subdivision methods, relying either on Sturm sequences or Moebius transform followed by Descartes's rule of signs, destruct the sparse structure. Our approach relies on the generalized Budan-Fourier theorem of Coste, Lajous, Lombardi, Roy and the techniques developed in some previous works of Galligo. To such a f is associated a set of $d + 1 \mathbb{F}$-derivatives. The Budan-Fourier function $V_f(x)$ counts the sign changes in the sequence of $\mathbb{F}$-derivatives of f evaluated at x. The values at which this function jumps are called the $\mathbb{F}$-virtual roots of f. These include the real roots of f. We also consider the augmented $\mathbb{F}$-virtual roots of f and introduce a genericity property which eases our study. We present a real root isolation method and an algorithm which has been implemented in Maple. We rely on an improved generalized Budan-Fourier count applied to both the input polynomial and its reciprocal, together with Newton like approximation steps.

This is a joint work with Maria Emilia Alonso (University of Madrid).

6.2.11. Deformation of roots of polynomials via fractional derivatives
Participant: André Galligo.

In [26], we first recall the main features of Fractional calculus. In the expression of fractional derivatives of a real polynomial $f(x)$, we view the order of differentiation q as a new indeterminate; then we define a new bivariate polynomial $Pf(x, q)$. For $0 \leq q \leq 1$, $Pf(x, q)$ defines a homotopy between the polynomials $f(x)$ and $xf'(x)$. Iterating this construction, we associate to $f(x)$ a plane spline curve, called the stem of f. Stems of classic random polynomials exhibits intriguing patterns; moreover in the complex plane $Pf(x, q)$ creates an unexpected correspondence between the complex roots and the critical points of $f(x)$. We propose 3 conjectures to describe and explain these phenomena. Illustrations are provided relying on the Computer Algebra System Maple.

GEOMETRICA Project-Team

6. New Results

6.1. Mesh Generation and Geometry Processing

6.1.1. New bounds on the size of optimal meshes

Participant: Donald Sheehy.

The theory of optimal size meshes gives a method for analyzing the output size (number of simplices) of a Delaunay refinement mesh in terms of the integral of a sizing function over the input domain. The input points define a maximal such sizing function called the feature size. This work aims to find a way to bound the feature size integral in terms of an easy to compute property of a suitable ordering of the point set. The key idea is to consider the pacing of an ordered point set, a measure of the rate of change in the feature size as points are added one at a time. In previous work, Miller et al. showed that if an ordered point set has pacing ϕ, then the number of vertices in an optimal mesh will be $O(\phi^d n)$, where d is the input dimension. We give a new analysis of this integral showing that the output size is only $\Theta(n + n \log \phi)$. The new analysis tightens bounds from several previous results and provides matching lower bounds. Moreover, it precisely characterizes inputs that yield outputs of size $O(n)$ [20].

6.1.2. State of the art in quad meshing

Participant: David Bommes.

Triangle meshes have been nearly ubiquitous in computer graphics, and a large body of data structures and geometry processing algorithms based on them has been developed in the literature. At the same time, quadrilateral meshes, especially semi-regular ones, have advantages for many applications, and significant progress was made in quadrilateral mesh generation and processing during the last several years. In this work, we discuss the advantages and problems of techniques operating on quadrilateral meshes, including surface analysis and mesh quality, simplification, adaptive refinement, alignment with features, parametrization, and remeshing [23].

6.1.3. Meshing the hyperbolic octagon

Participants: Mathieu Schmitt, Monique Teillaud.

We propose a practical method to compute a mesh of the octagon, in the Poincaré disk, that respects its symmetries. This is obtained by meshing the Schwartz triangle $T(8, 3, 2)$ and applying relevant hyperbolic symmetries (ie., Euclidean reflexions or inversions). The implementation is based on CGAL 2D meshes and on the ongoing implementation on CGAL hyperbolic Delaunay triangulations [44]. Further work will include solving robutsness issues and generalizing the method to any Schwartz triangle [62].

6.1.4. Index-based data structure for 3D polytopal complexes

Participant: David Bommes.

OpenVolumeMesh is a data structure which is able to represent heterogeneous 3-dimensional polytopal cell complexes and is general enough to also represent non-manifolds without incurring undue overhead [30]. Extending the idea of half-edge based data structures for two-manifold surface meshes, all faces, i.e. the two-dimensional entities of a mesh, are represented by a pair of oriented half-faces. The concept of using directed half-entities enables inducing an orientation to the meshes in an intuitive and easy to use manner. We pursue the idea of encoding connectivity by storing first-order top-down incidence relations per entity, i.e. for each entity of dimension d, a list of links to the respective incident entities is stored. For instance, each half-face as well as its orientation is uniquely determined by a tuple of links to its incident half-edges or each 3D cell by the set of incident half-faces. This representation allows for handling non-manifolds as well as mixed-dimensional mesh configurations. No entity is duplicated according to its valence, instead, it is shared by all incident entities in order to reduce memory consumption. Furthermore, an array-based storage layout is used in combination with direct index-based access. This guarantees constant access time to the entities of a mesh. Although bottom-up incidence relations are implied by the top-down incidences, our data structure provides the option to explicitly generate and cache them in a transparent manner. This allows for accelerated navigation in the local neighbor- hood of an entity. We provide an open-source and platform-independent implementation of the proposed data structure written in C++ using dynamic typing paradigms. The library is equipped with a set of STL compliant iterators, a generic property system to dynamically attach properties to all entities at run-time, and a serializer/deserializer supporting a simple file format. Due to its similarity to the OpenMesh data structure, it is easy to use, in particular for those familiar with OpenMesh. Since the presented data structure is compact, intuitive, and efficient, it is suitable for a variety of applications, such as meshing, visualization, and numerical analysis. OpenVolumeMesh is open-source software licensed under the terms of the LGPL [29].

6.1.5. *Editable SQuad representation for triangle meshes*
Participant: Olivier Devillers.

In collaboration with Luca Castelli Aleardi (LIX, Palaiseau) and Jarek Rossignac (Georgia Tech).

We consider the problem of designing space efficient solutions for representing the connectivity information of manifold triangle meshes. Most mesh data structures are quite redundant, storing a large amount of information in order to efficiently support mesh traversal operators. Several compact data structures have been proposed to reduce storage cost while supporting constant-time mesh traversal. Some recent solutions are based on a global re-ordering approach, which allows to implicitly encode a map between vertices and faces. Unfortunately, these compact representations do not support efficient updates, because local connectivity changes (such as edge-contractions, edge-flips or vertex insertions) require re-ordering the entire mesh. Our main contribution is to propose a new way of designing compact data structures which can be dynamically maintained. In our solution, we push further the limits of the re-ordering approaches: the main novelty is to allow to re-order vertex data (such as vertex coordinates), and to exploit this vertex permutation to easily maintain the connectivity under local changes. We describe a new class of data structures, called Editable SQuad (ESQ), offering the same navigational and storage performance as previous works, while supporting local editing in amortized constant time. As far as we know, our solution provides the most compact dynamic data structure for triangle meshes. We propose a linear-time and linear-space construction algorithm, and provide worst-case bounds for storage and time cost [25].

6.1.6. *Surface reconstruction through point set structuring*
Participants: Pierre Alliez, Florent Lafarge.

We present a method for reconstructing surfaces from point sets. The main novelty lies into a structure-preserving approach where the input point set is first consolidated by structuring and resampling the planar components, before reconstructing the surface from both the consolidated components and the unstructured points. The final surface is obtained through solving a graph-cut problem formulated on the 3D Delaunay triangulation of the structured point set where the tetrahedra are labeled as inside or outside cells. Structuring facilitates the surface reconstruction as the point set is substantially reduced and the points are enriched with structural meaning related to adjacency between primitives. Our approach departs from the common dichotomy between smooth/piecewise-smooth and primitive-based representations by gracefully combining

canonical parts from detected primitives and free-form parts of the inferred shape. Our experiments on a variety of inputs illustrate the potential of our approach in terms of robustness, flexibility and efficiency [59].

6.1.7. *Feature-preserving surface reconstruction and simplification from defect-laden point sets*

Participants: Pierre Alliez, David Cohen-Steiner, Julie Digne.

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

We introduce a robust and feature-capturing surface reconstruction and simplification method that turns an input point set into a low triangle-count simplicial complex. Our approach starts with a (possibly non-manifold) simplicial complex filtered from a 3D Delaunay triangulation of the input points. This initial approximation is iteratively simplified based on an error metric that measures, through optimal transport, the distance between the input points and the current simplicial complex, both seen as mass distributions. Our approach is shown to exhibit both robustness to noise and outliers, as well as preservation of sharp features and boundaries (Figure 1). Our new feature-sensitive metric between point sets and triangle meshes can also be used as a post-processing tool that, from the smooth output of a reconstruction method, recovers sharp features and boundaries present in the initial point set [58].

6.1.8. *Similarity based filtering of point clouds*

Participant: Julie Digne.

Denoising surfaces is a crucial step in the surface processing pipeline. This is even more challenging when no underlying structure of the surface is known, that is when the surface is represented as a set of unorganized points. We introduce a denoising method based on *local similarities*. The contributions are threefold: first, we do not denoise directly the point positions but use a low/high frequency decomposition and denoise only the high frequency. Second, we introduce a local surface parameterization which is proved stable. Finally, this method works directly on point clouds, thus avoiding building a mesh of a noisy surface which is a difficult problem. Our approach is based on denoising a height vector field by comparing the neighborhood of the point with neighborhoods of other points on the surface (Figure 2). It falls into the non-local denoising framework that has been extensively used in image processing, but extends it to unorganized point clouds [26].

6.1.9. *Progressive compression of manifold polygon meshes*

Participant: Pierre Alliez.

In collaboration with Adrien Maglo, Clément Courbet and Céline Hudelot from Ecole Centrale Paris.

We present a new algorithm for the progressive compression of surface polygon meshes. The input surface is decimated by several traversals that generate successive levels of detail through a specific patch decimation operator which combines vertex removal and local remeshing. This operator encodes the mesh connectivity through a transformation that generates two lists of Boolean symbols during face and edge removals. The geometry is encoded with a barycentric error prediction of the removed vertex coordinates. In order to further reduce the size of the geometry and connectivity data, we propose a curvature prediction method and a connectivity prediction scheme based on the mesh geometry. We also include two methods that improve the rate-distortion performance: a wavelet formulation with a lifting scheme and an adaptive quantization technique. Experimental results demonstrate the effectiveness of our approach in terms of compression rates and rate-distortion performance. Our approach compares favorably to compression schemes specialized to triangle meshes [31].

6.2. Topological and Geometric Inference

6.2.1. *Homological reconstruction and simplification in $\mathbb{R}^3$*

Participants: Olivier Devillers, Marc Glisse.

In collaboration with Dominique Attali (Gipsa-lab), Ulrich Bauer (Göttingen Univ.), and André Lieutier (Dassault Systèmes).
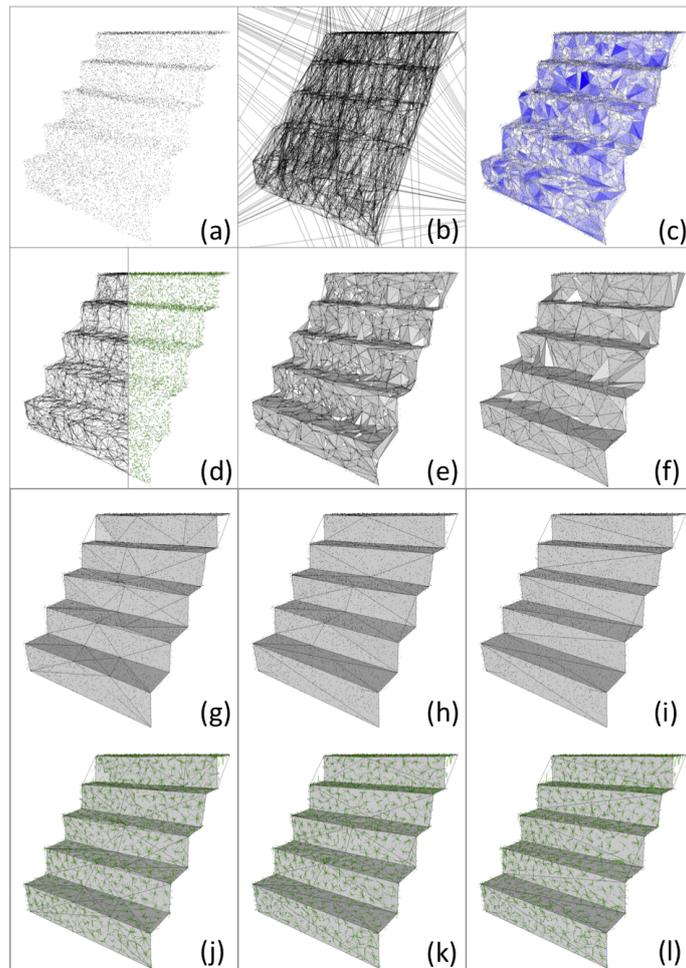
Figure 1. Steps of our algorithm: (a) Initial point set; (b) 3D Delaunay triangulation of a random subset containing 10% of the input points; (c) Initial simplicial complex constructed from facets of the 3D triangulation with non-zero measure; (d) Initial transport plan assigning point samples to bin centroids (green arrows); (e-f) Intermediary decimation steps; (g-i) Reconstruction with 100, 50, and 22 vertices, respectively; (j-l) Final transport plan with 100, 50, and 22 vertices, respectively.
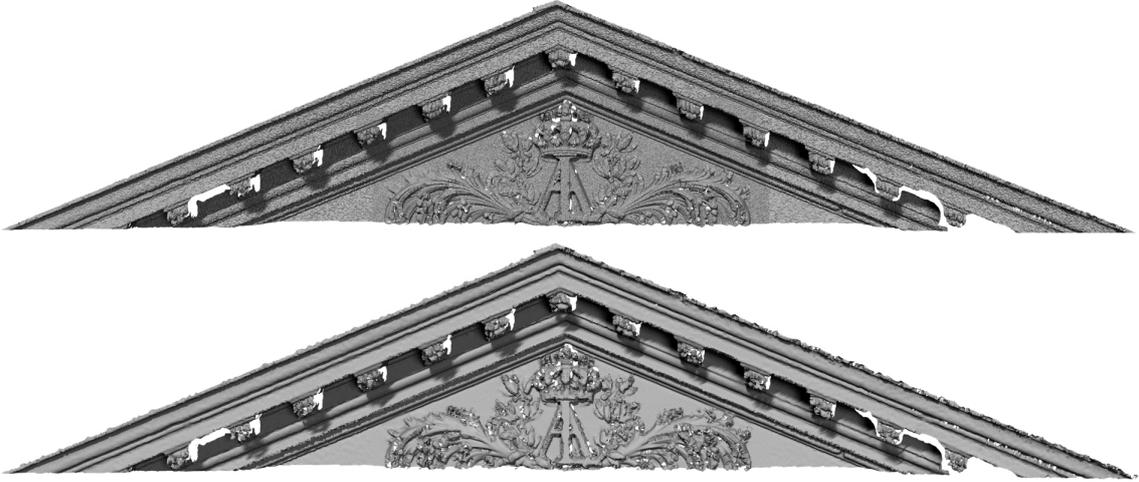
Figure 2. Similarity-based denoising. Top: input point set. Bottom: point set after denoising.

We consider the problem of deciding whether the persistent homology group of a simplicial pair (K, L) can be realized as the homology $H_*(X)$ of some space X with $L \subset X \subset K$. We show that this problem is NP-complete even if K is embedded in $\mathbb{R}^3$.

As a consequence, we show that it is NP-hard to simplify level and sublevel sets of scalar functions on $\mathbb{S}^3$ within a given tolerance constraint. This problem has relevance to the visualization of medical images by isosurfaces. We also show an implication to the theory of well groups of scalar functions: not every well group can be realized by some level set, and deciding whether a well group can be realized is NP-complete [43].

6.2.2. *The structure and stability of persistence modules*
Participants: Frédéric Chazal, Marc Glisse, Steve Oudot.

In collaboration with Vin de Silva (Pomona College)

We give a self-contained treatment of the theory of persistence modules indexed over the real line. We give new proofs of the standard results. Persistence diagrams are constructed using measure theory. Linear algebra lemmas are simplified using a new notation for calculations on quiver representations. We show that the stringent finiteness conditions required by traditional methods are not necessary to prove the existence and stability of the persistence diagram. We introduce weaker hypotheses for taming persistence modules, which are met in practice and are strong enough for the theory still to work. The constructions and proofs enabled by our framework are, we claim, cleaner and simpler [54].

6.2.3. *Persistence stability for geometric complexes*
Participants: Frédéric Chazal, Steve Oudot.

In collaboration with Vin de Silva (Pomona College)

We study the properties of the homology of different geometric filtered complexes (such as Vietoris–Rips, Čech and witness complexes) built on top of precompact spaces. Using recent developments in the theory of topological persistence [54] we provide simple and natural proofs of the stability of the persistent homology of such complexes with respect to the Gromov–Hausdorff distance. We also exhibit a few noteworthy properties of the homology of the Rips and Čech complexes built on top of compact spaces [53].

6.2.4. *Zigzag zoology: rips zigzags for homology inference*

Participants: Steve Oudot, Donald Sheehy.

For points sampled near a compact set X, the persistence barcode of the Rips filtration built from the sample contains information about the homology of X as long as X satisfies some geometric assumptions. The Rips filtration is prohibitively large, however zigzag persistence can be used to keep the size linear. We present several species of Rips-like zigzags and compare them with respect to the signal-to-noise ratio, a measure of how well the underlying homology is represented in the persistence barcode relative to the noise in the barcode at the relevant scales. Some of these Rips-like zigzags have been available as part of the Dionysus library for several years while others are new. Interestingly, we show that some species of Rips zigzags will exhibit less noise than the (non-zigzag) Rips filtration itself. Thus, the Rips zigzag can offer improvements in both size complexity and signal-to-noise ratio.

Along the way, we develop new techniques for manipulating and comparing persistence barcodes from zigzag modules. We give methods for reversing arrows and removing spaces from a zigzag. We also discuss factoring zigzags and a kind of interleaving of two zigzags that allows their barcodes to be compared. These techniques were developed to provide our theoretical analysis of the signal-to-noise ratio of Rips-like zigzags, but they are of independent interest as they apply to zigzag modules generally [60].

6.2.5. *A space and time efficient implementation for computing persistent homology*

Participants: Jean-Daniel Boissonnat, Clément Maria.

In collaboration with Tamal Dey (Ohio State University)

The persistent homology with Z_2-coefficients coincides with the same for cohomology because of duality. Recently, it has been observed that the cohomology based algorithms perform much better in practice than the originally proposed homology based persistence algorithm. We have implemented a cohomology based algorithm that attaches binary labels called annotations with the simplices. This algorithm fits very naturally with our recently developed data structure called simplex tree to represent simplicial complexes [49], [22]. By taking advantages of several practical tricks such as representing annotations compactly with memory words, using a union-find structure that eliminates duplicate annotation vectors, and a lazy evaluation, we save both space and time cost for computations. The complexity of the procedure, in practice, depends almost linearly on the size of the simplicial complex and on the variables related to the maximal dimension of the local homology groups we maintain during the computation, which remain small in practice. We provide a theoretical analysis as well as a detailed experimental study of our implementation. Experimental results show that our implementation performs several times better than the existing state-of-the-art software for computing persistent homology in terms of both time and memory requirements and can handle very large (several hundred million simplices in high-dimension) complexes efficiently [45].

6.2.6. *Minimax rates for homology inference*

Participant: Donald Sheehy.

In collaboration with Sivaraman Balakrishnan and Alessandro Rinaldo and Aarti Singh and Larry A. Wasserman (Carnegie Mellon University)

Often, high dimensional data lie close to a low-dimensional submanifold and it is of interest to understand the geometry of these submanifolds. The homology groups of a manifold are important topological invariants that provide an algebraic summary of the manifold. These groups contain rich topological information, for instance, about the connected components, holes, tunnels and sometimes the dimension of the manifold. We consider the statistical problem of estimating the homology of a manifold from noisy samples under several different noise models. We derive upper and lower bounds on the minimax risk for this problem. Our upper bounds are based on estimators which are constructed from a union of balls of appropriate radius around carefully selected points. In each case, we establish complementary lower bounds using Le Cam's lemma [15].

6.2.7. *Linear-size approximations to the Vietoris-Rips filtration*

Participant: Donald Sheehy.

The Vietoris-Rips filtration is a versatile tool in topological data analysis. Unfortunately, it is often too large to construct in full. We show how to construct an $O(n)$-size filtered simplicial complex on an n-point metric space such that the persistence diagram is a good approximation to that of the Vietoris-Rips filtration. The filtration can be constructed in $O(n \log n)$ time. The constants depend only on the doubling dimension of the metric space and the desired tightness of the approximation. For the first time, this makes it computationally tractable to approximate the persistence diagram of the Vietoris-Rips filtration across all scales for large data sets. Our approach uses a hierarchical net-tree to sparsify the filtration. We can either sparsify the data by throwing out points at larger scales to give a zigzag filtration, or sparsify the underlying graph by throwing out edges at larger scales to give a standard filtration. Both methods yield the same guarantees [34].

6.2.8. *A multicover nerve for geometric inference*

Participant: Donald Sheehy.

We show that filtering the barycentric decomposition of a Čech complex by the cardinality of the vertices captures precisely the topology of k-covered regions among a collection of balls for all values of k. Moreover, we relate this result to the Vietoris-Rips complex to get an approximation in terms of the persistent homology [33].

6.2.9. *Computing well diagrams for vector fields on $\mathbb{R}^n$*

Participant: Frédéric Chazal.

In collaboration with Primoz Skraba (Lubiana Univ.), Amit Patel (Rutgers Univ.)

Using topological degree theory, we present and prove correctness of a fast algorithm for computing the well diagram, a quantitative property, of a vector field on Euclidean space [17].

6.3. Data Structures and Robust Geometric Computation

6.3.1. *Straight-line graph drawing on the torus*

Participant: Olivier Devillers.

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

We extend the notion of canonical orderings to cylindric triangulations. This allows us to extend the incremental straight-line drawing algorithm of de Fraysseix et al. to this setting. Our algorithm yields in linear time a crossing-free straight-line drawing of a cylindric triangulation T with n vertices on a regular grid $\mathbb{Z}/w\mathbb{Z} \times [0, h]$, with $w \le 2n$ and $h \le n(2d + 1)$, where d is the (graph-) distance between the two boundaries. As a by-product, we can also obtain in linear time a crossing-free straight-line drawing of a toroidal triangulation with n vertices on a periodic regular grid $\mathbb{Z}/w\mathbb{Z} \times \mathbb{Z}/h\mathbb{Z}$, with $w \le 2n$ and $h \le 1 + n(2c + 1)$, where c is the length of a shortest non-contractible cycle. Since $c \le \sqrt{2n}$, the grid area is $O(n^{5/2})$ [24].

6.3.2. *Qualitative symbolic perturbation*

Participants: Olivier Devillers, Monique Teillaud.

In collaboration with Menelaos Karavelas (University of Crete).

In the literature, the generic way to address degeneracies in computational geometry is the *Symbolic Perturbation* paradigm: the input is made dependent of some parameter ε so that for ε positive and close to zero, the input is close to the original input, while at the same time, in non-degenerate position. A geometric predicate can usually be seen as the sign of some function of the input. In the symbolic perturbation paradigm, if the function evaluates to zero, the input is perturbed by a small positive ε, and the sign of the function evaluated at the perturbed input is used instead.

The usual way of using this approach is what we will call *Algebraic Symbolic Perturbation* framework. When the function to be evaluated is a polynomial of the input, its perturbed version is seen as a polynomial in ε, whose coefficients are polynomials in the input. These coefficients are evaluated by increasing degree in ε until a non-vanishing coefficient is found. The number of these coefficients can be quite large and expressing them in an easily and efficiently computable manner (e.g., factorized) may require quite some work.

We propose to address the handling of geometric degeneracies in a different way, namely by means of what we call the *Qualitative Symbolic Perturbation* framework. We no longer use a single perturbation that must remove all degeneracies, but rather a sequence of perturbations, such that the next perturbation is being used only if the previous ones have not removed the degeneracies. The new perturbation is considered as *symbolically smaller* than the previous ones. This approach allows us to use simple elementary perturbations whose effect can be analyzed and evaluated: (1) by geometric reasoning instead of algebraic development of the predicate polynomial in ε, and (2) independently of a specific algebraic formulation of the predicate.

We apply our framework to predicates used in the computation of Apollonius diagrams in 2D and 3D, as well as the computation of trapezoidal maps of circular arcs [57].

6.3.3. *Covering spaces and Delaunay triangulations of the 2D flat torus*
Participants: Mikhail Bogdanov, Monique Teillaud.

In collaboration with Gert Vegter (Johan Bernoulli Institute, Groningen University)

A previous algorithm was computing the Delaunay triangulation of the flat torus, by using a 9-sheeted covering space [64]. We propose a modification of the algorithm using only a 8-sheeted covering space, which allows to work with 8 periodic copies of the input points instead of 9. The main interest of our contribution is not only this result, but most of all the method itself: this new construction of covering spaces generalizes to Delaunay triangulations of surfaces of higher genus.

6.3.4. *Hyperbolic Delaunay complexes and Voronoi diagrams made practical*
Participants: Mikhail Bogdanov, Olivier Devillers, Monique Teillaud.

We study Delaunay complexes and Voronoi diagrams in the Poincaré ball, a confomal model of the hyperbolic space, in any dimension. We elaborate on our earlier work on the space of spheres [65], giving a detailed description of algorithms, and presenting a static and a dynamic variants. All proofs are based on geometric reasoning, they do not resort to any use of the analytic formula of the hyperbolic distance. We also study algebraic and arithmetic issues, observing that only rational computations are needed. This allows for an exact and efficient implementation in 2D. All degenerate cases are handled. The implementation will be submitted to the CGAL editorial board for future integration into the CGAL library [44].

6.3.5. *The stability of Delaunay triangulations*
Participants: Jean-Daniel Boissonnat, Ramsay Dyer.

In collaboration with Arijit Ghosh (Indian Statistical Institute, Kolkata, India)

We introduce a parametrized notion of genericity for Delaunay triangulations which, in particular, implies that the Delaunay simplices of δ-generic point sets are thick. Equipped with this notion, we study the stability of Delaunay triangulations under perturbations of the metric and of the vertex positions. We quantify the magnitude of the perturbations under which the Delaunay triangulation remains unchanged. We also present an algorithm that takes as input a discrete point set in $\mathbb{R}^m$, and performs a small perturbation that guarantees that the Delaunay triangulation of the resulting perturbed point set has quantifiable stability with respect to the metric and the point positions. There is also a guarantee on the quality of the simplices: they cannot be too flat. The algorithm provides an alternative tool to the weighting or refinement methods to remove poorly shaped simplices in Delaunay triangulations of arbitrary dimension, but in addition it provides a guarantee of stability for the resulting triangulation [21], [47].

6.3.6. *Constructing intrinsic Delaunay triangulations of submanifolds*
Participants: Jean-Daniel Boissonnat, Ramsay Dyer.

In collaboration with Arijit Ghosh (Indian Statistical Institute, Kolkata, India)

This work is the algorithmic counterpart of our previous paper [21]. We describe an algorithm to construct an intrinsic Delaunay triangulation of a smooth closed submanifold of Euclidean space. We also provide a counterexample to the results announced by Leibon and Letscher on Delaunay triangulations on Riemannian manifolds. In general the nerve of the intrinsic Voronoi diagram is not homeomorphic to the manifold. The density of the sample points alone cannot guarantee the existence of a Delaunay triangulation. To circumvent this issue, we use results established in our companion paper on the stability of Delaunay triangulations on δ-generic point sets. We establish sampling criteria which ensure that the intrinsic Delaunay complex coincides with the restricted Delaunay complex and also with the recently introduced tangential Delaunay complex. The algorithm generates a point set that meets the required criteria while the tangential complex is being constructed. In this way the computation of geodesic distances is avoided, the runtime is only linearly dependent on the ambient dimension, and the Delaunay complexes are guaranteed to be triangulations of the manifold [46].

6.3.7. *Equating the witness and restricted Delaunay complexes*

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

In collaboration with Arijit Ghosh (Indian Statistical Institute, Kolkata, India)

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 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 [37].

6.3.8. *Simpler complexity analysis of random geometric structures*

Participants: Olivier Devillers, Marc Glisse.

In collaboration with Xavier Goaoc (EPI VEGAS*).*

Average-case analysis of data-structures or algorithms is commonly used in computational geometry when the, more classical, worst-case analysis is deemed overly pessimistic. Since these analyses are often intricate, the models of random geometric data that can be handled are often simplistic and far from "realistic inputs". We present a new simple scheme for the analysis of geometric structures. While this scheme only produces results up to a polylog factor, it is much simpler to apply than the classical techniques and therefore succeeds in analyzing new input distributions related to smoothed complexity analysis.

We illustrate our method on two classical structures: convex hulls and Delaunay triangulations. Specifically, we give short and elementary proofs of the classical results that n points uniformly distributed in a ball in $\mathbb{R}^d$ have a convex hull and a Delaunay triangulation of respective expected complexities $\widetilde{\Theta}(n^{\frac{d-1}{d+1}})$ and $\widetilde{\Theta}(n)$. We then prove that if we start with n points well-spread on a sphere, e.g. an (ϵ, κ)-sample of that sphere, and perturb that sample by moving each point randomly and uniformly within distance at most δ of its initial position, then the expected complexity of the convex hull of the resulting point set is $\widetilde{\Theta}\left((\sqrt{n})^{1-\frac{1}{d}} \left(\frac{1}{\sqrt[4]{\delta}} \right)^{d-\frac{1}{d}} \right)$ [55].

6.3.9. *Analysis of cone vertex walk in Poisson Delaunay triangulation*

Participants: Olivier Devillers, Ross Hemsley.

In collaboration with Nicolas Broutin (EPI RAP*).*

Walking strategies are a standard tool for point location in a triangulation of size n. Although often claimed to be $\Theta(\sqrt{n})$ under random distribution hypotheses, this conjecture has only been formally proved by Devroye, Lemaire, and Moreau [*Comp Geom–Theor Appl*, vol. 29, 2004], in the case of the so called *straight walk* which has the very specific property that deciding whether a given (Delaunay) triangle belongs to the walk may be determined without looking at the other sites. We analyze a different walking strategy that follows vertex neighbour relations to move towards the query. We call this walk *cone vertex walk*. We prove that cone vertex walk visits $\Theta(\sqrt{n})$ vertices and can be constructed in $\Theta(\sqrt{n})$ time. We provide explicit bounds on the hidden constants [50].

6.3.10. The monotonicity of f-vectors of random polytopes

Participants: Olivier Devillers, Marc Glisse.

In collaboration with Xavier Goaoc and Guillaume Moroz (EPI VEGAS*) and Matthias Reitzner (Universität Osnabrück, Germany).*

Let K be a compact convex body in $\mathbb{R}^d$, let K_n be the convex hull of n points chosen uniformly and independently in K, and let $f_i(K_n)$ denote the number of i-dimensional faces of K_n.

We show that for planar convex sets, $E[f_0(K_n)]$ is increasing in n. In dimension $d \geq 3$, we prove that if $\lim_{n\to\infty} \frac{E[f_{d-1}(K_n)]}{An^c} = 1$ for some constants A and $c > 0$ then the function $n \mapsto E[f_{d-1}(K_n)]$ is increasing for n large enough. In particular, the number of facets of the convex hull of n random points distributed uniformly and independently in a smooth compact convex body is asymptotically increasing. Our proof relies on a *random sampling* argument [57].

6.3.11. Efficient Monte Carlo sampler for detecting parametric objects in large scenes

Participants: Florent Lafarge, Yannick Verdie.

Point processes have demonstrated efficiency and competitiveness when addressing object recognition problems in vision. However, simulating these mathematical models is a difficult task, especially on large scenes. Existing samplers suffer from average performances in terms of computation time and stability. We propose a new sampling procedure based on a Monte Carlo formalism. Our algorithm exploits Markovian properties of point processes to perform the sampling in parallel. This procedure is embedded into a data-driven mechanism such that the points are non-uniformly distributed in the scene. The performances of the sampler are analyzed through a set of experiments on various object recognition problems from large scenes, and through comparisons to the existing algorithms [35], [63].

6.4. Applications

6.4.1. Creating large-scale city models from 3D-point clouds: a robust approach with hybrid representation

Participant: Florent Lafarge.

We present a novel and robust method for modeling cities from 3D-point data. Our algorithm provides a more complete description than existing approaches by reconstructing simultaneously buildings, trees and topologically complex grounds. A major contribution of our work is the original way of modeling buildings 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. Our approach is experimentally validated on complex buildings and large urban scenes of millions of points, and is compared to state-of-the-art methods [19].

6.4.2. The sticky geometry of the cosmic web

Participant: Monique Teillaud.

In collaboration with Johan Hidding, Rien van de Weygaert, Bernard J.T. Jones (Kapteyn Institute, Groningen University) and Gert Vegter (Johan Bernoulli Institute, Groningen University)

We highlight the application of Computational Geometry to our understanding of the formation and dynamics of the Cosmic Web. The emergence of this intricate and pervasive weblike structure of the Universe on Megaparsec scales can be approximated by a well-known equation from fluid mechanics, the Burgers' equation. The solution to this equation can be obtained from a geometrical formalism. We have extended and improved this method by invoking weighted Delaunay and Voronoi tessellations. The duality between these tessellations finds a remarkable and profound reflection in the description of physical systems in Eulerian and Lagrangian terms [28].

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

6. New Results

6.1. Ontology-Based Query Answering with Existential Rules

Participants: Jean-François Baget, Mélanie König, Michel Leclère, Marie-Laure Mugnier, Michaël Thomazo.

Note that for this section, as well as all sections in New Results, participants are given in alphabetical order.

In collaboration with: Sebastian Rudolph (Karlsruhe Institute of Technology)

We have pursued the work on the existential rule framework in the context of Ontology-Based Query Answering. See the 2011 activity report for details on this framework also known as Datalog+/-. The ontology-based query answering problem consists of querying data while taking into account inferences enabled by an ontology (described by existential rules in our case).

From 2009 to 2011, we mainly investigated decidability and complexity issues. In 2012, while still interested in deepening decidability and complexity results, we tackled the next step: algorithms. Our aim is to develop algorithms with good theoretical properties (at least they should run in "the good worst-case complexity class") and with good performance in practice. There are two main ways of processing rules, namely forward chaining and backward chaining. In forward chaining, rules are applied to enrich the initial facts and query answering is solved by evaluating the query against the "saturated" facts (as in a classical database system). When it is finite, the backward chaining process can be divided into two steps: first, the query is rewritten into a first-order query (typically a union of conjunctive queries) using the rules; then the rewritten query is evaluated against the initial facts (again, as in a classical database system).

6.1.1. Forward Chaining Algorithms

Considering the expressive class of greedy bounded-treewidth set of rules (in short *gbts*), which we defined in 2011, we have designed a query answering algorithm which has several advantages over 2011 algorithm, while staying optimal with respect to worst-case combined and data complexities.

1. It is much more implementable (previous algorithm was using an oracle).
2. It is generic in the sense that it works for any class of rules that fulfills the gbts property, but it can also be easily specialized for specific gbts subclasses whith lower complexities, such as frontier-guarded or guarded rules, in such a way that it runs in the good complexity class.
3. It allows for separation between offline and online processing steps: the knowledge base can be compiled independently from queries, which are evaluated against the compiled form.

One of the lightweight description logics used for ontology-based query answering is $\mathcal{EL}$. We designed a subclass of existential rules that covers $\mathcal{EL}$ with the same complexity of reasoning, while allowing for any predicate arity and some cycles on variables. We also added complex role inclusions like transitivity and right/left identity rules to enhance expressivity, while staying polynomial in data complexity and generalizing existing results.

- *Results published in [36], [37] and [32] (invited conference). See also our research report [49] for a longer version.*
- *A journal version extending the papers at IJCAI 2011 and KR 2012 is in preparation, to be submitted to a major artificial intelligence journal.*

6.1.2. Backward Chaining Algorithms

We consider query rewriting techniques that output a union of conjunctive queries, which we see as a set of conjunctive queries. More specifically, only the most general elements of this set need to be kept in the output. We first proved that all sound and complete query rewriting algorithms necessarily produce the same result (up to redundancy) when restricted to their most general elements. It follows that comparing existing algorithms with respect to the size of the produced query is pointless.

Existing query rewriting algorithms accept only specific classes of existential rules (mainly corresponding to the translation of some lightweight description logics). We designed an algorithm that accept as input any set of existential rules and stops if this set of rules fufills so-called *fus* property (meaning that the set of most general rewritings of any initial conjunctive query is finite). This algorithm has been implemented and first experimentations have been led on rule bases obtained by translating description logic bases.

- *Results published in [31] (best paper price)*

6.1.3. *Querying Optimization (Work in Progress)*

Our current work aims at improving previous algorithms, in particular: the online querying step in the gbts algorithm; the query rewriting algorithm, by avoiding generating several times equivalent rewritings; for specific subclasses, query rewriting into a set of so-called semi-conjunctive queries instead of conjunctive queries, which reduces the size of the output query.

6.2. Reasoning with Imperfect Information and Priorities

Participants: Madalina Croitoru, Jérôme Fortin, Souhila Kaci, Tjitze Rienstra, Rallou Thomopoulos.

In collaboration with: Joël Abecassis (IATE/INRA), Patrice Buche (IATE/INRA), Nir Oren (Univ. of Aberdeen, Scotland), Leon van der Torre (University of Luxembourg) and Nouredine Tamani (post-doc IATE).

This year, we mainly investigated decision support based on argumentations systems and preferences, either in relation with application needs in agronomy or on more fundamental aspects.

6.2.1. *Argumentation for Decision Making in Agronomy*

Historically, scientific investigations in this axis are guided by applications of our partners in agronomy (IATE laboratory). 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.

In the context of the EcoBioCap project (see Sect. 8.2), the different stakeholders have expressed conflicting preferences for the packaging quality. However, when discussing with domain experts they have raised the need for a tool which allows them to highlight a conflict and see the reasons behind it. In order to achieve this goal two steps were taken. First we have instantiated a popular logical argumentation framework (ASPIC+) with a simple preference logic. This allowed the different experts to express arguments about their preferences. We can then extract maximal consistent subsets of preferences by the means of extensions.

- *This work was performed in collaboration with the University of Aberdeen (Dr. Nir Oren) and the results were published and presented at the COMMA conference [24].*

Second, a negotiation phase was introduced to the previously described system in order for the domain experts to refine and extend their preferences. This tool was the aim of the master thesis of Patricio Mosse.

- *This work was published and presented at the Effost conference [23], based upon Patricio Mosse's Master Thesis [48]. A detailed journal article reporting on the two steps is under preparation and will be submitted beginning 2013.*

Let us mention additional results related to the applications in agronomy on semi-automatic data extraction from web data (tables) [39], [40], [41], data reliability, and the representation and flexible querying of imprecise data with fuzzy sets [42], [15]. These investigations are complementary to the above mentioned results on argumentation and generally relate to other aspects in the same applicative projects.

6.2.2. *Conditional Acceptance Functions*

Dung-style abstract argumentation theory centers on argumentation frameworks and acceptance functions. The latter take as input a framework and return sets of labelings. A labeling assigns "in", "out" or "undecided" to each arguments. Arguments having "in" assignment are acceptable arguments. This methodology however assumes full awareness of the arguments relevant to the evaluation. There are two reasons why this is not satisfactory. Firstly, full awareness is, in general, not a realistic assumption. Second, frameworks have explanatory power, which allows us to reason abductively or counterfactually, but this is lost under the usual

semantics. To recover this aspect, we generalized conventional acceptance, and we present the concept of a conditional acceptance function which copes with the dynamics of argumentation frameworks.

- *Results published in [28].*

6.2.3. Foundational Aspects of Preferences

Preferences are the backbone of various fields as they naturally arise and play an important role in many real-life decisions. Preferences are fundamental in scientific research frameworks as well as applications. One of the main problems an individual faces when expressing her preferences lies in the number of variables (or attributes or criteria) that she takes into account to evaluate the different outcomes. Indeed, the number of outcomes increases exponentially with the number of variables. Moreover, due to their cognitive limitation, individuals are generally not willing to compare all possible pairs of outcomes or evaluate them individually. These facts have an unfortunate consequence that any preference representation language that is based on the direct assessment of individual preferences over the complete set of outcomes is simply infeasible. Fortunately, individuals can abstract their preferences. More specifically, instead of providing preferences over outcomes (by pairwise comparison or individual evaluation), they generally express preferences over partial descriptions of outcomes. Often such statements take the form of qualitative comparative preference statements e.g., "I like London more than Paris" and "prefer tea to coffee". Conditional logics aim at representing such partial descriptions of individual preferences which we refer to as comparative preference statements. They use different completion principles in order to compute a preference relation induced by a set of preference statements. In particular they use various more or less strong semantics to interpret comparative preference statements. So far the main objective in artificial intelligence has been to rank-order the set of outcomes given a set of comparative preference statements and one or several semantics. We addressed this problem from a different angle. We considered a set of postulates studied in preference logics and non-monotonic reasoning which formalize intuition one may have regarding the behavior of preference statements. We analyzed the behavior of the different semantics w.r.t. these postulates. Our analysis gives a complete picture of the behavior of our (five) semantics.

In the last decade, AI researchers have pointed out the existence of two types of information: positive information and negative information. This distinction has also been asserted in cognitive psychology. Distinguishing between these two types of information may be useful in both knowledge and preference representation. In the first case, one distinguishes between situations which are not impossible because they are not ruled out by the available knowledge, and what is possible for sure. In the second case, one distinguishes between what is not rejected and what is really desired. Besides it has been shown that possibility theory is a convenient tool to model and distinguish between these two types of information. Knowledge/Preference representation languages have also been extended to cope with this particular kind of information. Nevertheless despite solid theoretical advances in this topic, the crucial question of "which reading (negative or positive) one should have" remains a real bottleneck. We focused on comparative statements and presented a set of postulates describing different situations one may encounter. We provided a representation theorem describing which sets of postulates are satisfied by which kind of information (negative or positive) and conversely. One can then decide which reading to apply depending on which postulates she privileges.

- *Results published in [29] and [30].*

6.2.4. Argumentation for Inconsistency-Tolerant Query Answering (Work in Progress)

Argumentation allows to encode by the means of extensions maximal subsets of the knowledge base which are consistent (given the logic chosen). We are currently investigating the link between different argumentation extensions and the notion of a maximal repair as introduced by [51], [50] in the context of the positive existential subset of first order logic we are mainly working with. We are then interested in comparing the semantics proposed in the literature for query answering with inconsistent knowledge bases and argumentation reasoning paradigms. This study has been performed jointly with the University of Luxembourg during a research visit during end of November. We plan to submit our results at a conference beginning January.

6.3. Semantic Data Integration

Participants: Michel Chein, Madalina Croitoru, Léa Guizol, Michel Leclère, Rallou Thomopoulos.

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 [52], 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 Saïs, Nathalie Pernelle and Marie-Christine Rousset proposed the first logical approach [53]. Many experiments on public data are underway, in France (cf. DataLift[2] and ISIDORE[3] projects) or internationally (e.g., VIAF project[4] led by OCLC[5], whose aim is to interconnect authority files coming from 18 national organizations).

Three 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 based 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 metedata 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"* [54].

6.3.1. *Implementation of an Entity Identification Service*

Last year, we have developed a method and a prototype to perform entity resolution between on one hand the authors of a new bibliographic notice, and, on the other the domain experts of an authority catalog (and namely the Sudoc catalogue from the ABES agency). The prototype providing this service has been implemented on top of Cogui and experiments have been led in the context of the SudocAd project (jointly conducted by ABES and GraphIK). This work has been continued this year on the following issues as part of the Qualinca project:

- *generalizing the developed method with the aim to define a generic combined (numerical/logical) framework for entity resolution. This work is reported in the research report [44] that we plan to submit to a conference in January.*
- *Defining evaluation measures of the quality of an entity resolution tool. This work is still on-going.*

6.3.2. *Quality of Document Catalogs*

The SudocAd project showed the feasability and pertinence of a mixed approach for data interlinking problems. It also showed the immediate necessity of taking into account the existence of human errors already present in document catalogues. This led us to propose Qualinca, an ANR Contint project, accepted beginning 2012 and started in April 2012. The partners include two major actors in the document catalogues field: ABES and INA, as well as three academic research groups.

In this context we currently investigate a formal approach to the notion of a "key" in the web of data. Our immediate objective is to define the notion of a discovered key used then in order to evaluate the quality of data inter linking of a meta data catalogue.

---

[1] http://www.hcp.med.harvard.edu/statistics/survey-soft/docs/WinklerReclinkRef.pdf
[2] DataLift, http://datalift.org/
[3] ISIDORE, http://www.rechercheisidore.fr/
[4] The Virtual International Authority File, http://www.oclc.org/research/activities/viaf/
[5] Online Computer Library Center, http://www.oclc.org

We also study the methodology of linking error detection and fixing based on a partitioning (clustering) method on authors of bibliographic records. This study is part of the PhD thesis of Léa Guizol (jointly funded by GraphIK and ABES). The above mentioned methodology is based on a set of criteria which will allow us to cluster "similar" authors together. Each criterion represents a point of view on the author: name, publication time span, publication domain etc. The first challenge consists of defining for each of such view points the respective criteria. The second challenge is to propose an aggregation semantics of such criteria which is well adapted for the problem at hand.

- *The methodology of using such clustering techniques for this problem has been published in [25]. A certain number of criteria have already been implemented and different partitioning semantics proposed. We are currently evaluating these on the ABES data.*

6.3.3. Multi Agent Knowledge Allocation

The assumption behind semantic data integration and querying is that different agents accessing the integrated data repository will have equal interest in the querying results. This is not always true in a data sensitive scenario where the knowledge provider might want to allocate the query answers to the agents based on their valuations. Furthermore, the agents might want some information exclusively (and thus offer a valuation that allows it) while others might want it shared. To this end we have proposed a new mechanism of allocation of query answers inspired from combinatorial auctions. We have defined the newly introduced scenario of Multi Agent Knowledge Allocation and proposed a graph based method, inspired on network flows, for solving it.

- *These results were published in [26] and [35]. We are currently investigating the mechanism design aspects of such valuations in collaboration with the University of Athens (Dr. Iannis Vetsikas).*

INDES Project-Team

6. New Results

6.1. Security

Participants: Ilaria Castellani, Zhengqin Luo, Tamara Rezk [correspondant], José Santos, Manuel Serrano.

6.1.1. Session types with security

We have pursued our work on integrating security constraints within session types, in collaboration with our colleagues from Torino University. This resulted in the journal paper [8]. This article extends a previous conference paper with full proofs, additional examples and further results. In particular, [8] presents new properties of information-flow security, which is stronger and more compositional (*i.e.*, more robust with respect to parallel composition of processes) than that originally proposed, while being still ensured by the same session type system.

All the work on session types was partially funded by the ANR-08- EMER-010 grant PARTOUT. It is expected to continue within the starting COST Action BETTY.

6.1.2. Mashic Compiler: Mashup Sandboxing Based on Inter-frame Communication

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 constraints 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.

This work appeared in CSF'12 [14]. 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 [6].

6.2. Models, semantics, and languages

Participants: Pejman Attar, Gérard Berry, Gérard Boudol, Frédéric Boussinot, Ilaria Castellani, Johan Grande, Cyprien Nicolas, Tamara Rezk, Manuel Serrano [correspondant].

6.2.1. Memory Models

As regards the theory of multithreading, we have extended our operational approach to capture more relaxed memory models than simple write buffering. A step was made in this direction by formalizing the notion of a speculative computation, but this was not fully satisfactory as an operational approach to the theory of memory models: indeed, in the speculative framework one has to reject a posteriori some sequences of executions as invalid. In [13] we have defined a truly operational semantics, by means of an abstract machine, for extremely relaxed memory models like the one of PowerPC. In our new framework, the relaxed abstract machine features a "temporary store" where the memory operations issued by the threads are recorded, in program order. A memory model then specifies the conditions under which a pending operation from this sequence is allowed to be globally performed, possibly out of order. The memory model also involves a "write grain," accounting for architectures where a thread may read a write that is not yet globally visible. Our model is also flexible enough to account for a form of speculation used in PowerPC machines, namely branch prediction. To experiment with our framework, we found it useful to design and implement a simulator that allows us to exhaustively explore all the possible relaxed behaviors of (simple) programs. The main problem was to tame the combinatory explosion due to the massive non-deterministic interleaving of the relaxed semantics. Introducing several optimizations described in [13], we were able to run a large number of litmus tests successfully.

6.2.2. Dynamic Synchronous Language with Memory

We have investigated the language DSLM (Dynamic Synchronous Language with Memory), based on the synchronous reactive model. In DSLM, systems are composed of several sites, each of which runs a number of agents. An agent consists of a memory and a script. This script is made of several parallel components which share the agent's memory. A simple form of migration is provided: agents can migrate from one site to another. Since sites have different clocks, a migrating agent resumes execution at the start of the next instant in the destination site. Communication between a migrating agent and the agents of the destination site occurs via (dynamically bound) events. The language uses three kinds of parallelism: 1) synchronous, cooperative and deterministic parallelism among scripts within an agent, 2) synchronous, nondeterministic and confluent parallelism among agents within a site, and 3) asynchronous and nondeterministic parallelism among sites. Communication occurs via both shared memory and events in the first case, and exclusively via events in the other two cases. Scripts may call functions or modules which are handled in a host language. Two properties are assured by DSLM: reactivity of each agent and absence of data-races between agents. Moreover, the language offers a way to benefit from multi-core and multi-processor architectures, by means of the notion of synchronized scheduler which abstractly models a computing resource. Each site may be expanded and contracted dynamically by varying its number of synchronized schedulers. In this way one can model the load-balancing of agents over a site.

A secure extension of the language DSLM, called DSSLM (Dynamic Secure Synchronous Language with Memory), is currently under investigation. This language uses the same deterministic parallel operator for scripts as DSLM. It adds to DSLM a let operator that assigns a security level to the defined variable. Security levels are also assigned to events and sites, to allow information flow control during interactions and migrations. The study of different security properties (both sensitive and insensitive to the passage of the instants) and of type systems ensuring these properties is currently under way.

6.2.3. jthread

The jthread library (working name) is a Bigloo library featuring threads and mutexes and most notably a deadlock-free locking primitive. The jthread library appears as an alternative to Bigloo's pthread (POSIX threads) library and relies on it for its implementation.

The locking primitive is the following: `(synchronize* ml [:prelock mlp] expr1 expr2 ...)` where ml and mlp are lists of mutexes.

This primitive evaluates the expressions that constitute its body after having locked the mutexes in ml and before unlocking them back. The meaning of the *prelock* argument is to be explained below.

The absence of deadlocks is guaranteed by two complementary mechanisms:

- Each mutex belongs to a *region* defined by the programmer. Regions form a lattice which is inferred at runtime. A thread owning a mutex belonging to region R0 can only lock a mutex belonging to region R1 if R1 is lower than R0 in the lattice. This rule is enforced at runtime and guarantees the absence of deadlocks involving mutexes belonging to different regions.
- Under the previous condition, a thread owning a mutex M1 can lock a mutex M2 belonging to the same region only provided that M2 appeared in the *prelock* list of the `synchronize*` that locked M1. This rule is enforced at runtime and allows a *deadlock-avoiding* scheduling of threads based on previous work by Gérard Boudol and on Lamport's Bakery algorithm.

The library has been implemented. It is currently being integrated to Bigloo and benchmarked. It has not been released yet.

6.3. Web programming

Participants: Zhengqin Luo, Cyprien Nicolas, Tamara Rezk, Bernard Serpette, Manuel Serrano [correspondant].

6.3.1. *Reasoning about Web Applications: An Operational Semantics for HOP*

We propose a small-step operational semantics to support reasoning about web applications written in the multi-tier language HOP. The semantics covers both server side and client side computations, as well as their interactions, and includes creation of web services, distributed client-server communications, concurrent evaluation of service requests at server side, elaboration of HTML documents, DOM operations, evaluation of script nodes in HTML documents and actions from HTML pages at client side. We also model the browser same origin policy (SOP) in the semantics. We propose a safety property by which programs do not get stuck due to a violation of the SOP and a type system to enforce it. This work appeared in TOPLAS [7].

6.3.1.1. *Hiphop*

We pursued the development of the Hiphop orchestration language. The first version was written as a DSL with very few connection to Hop. During this year, we changed Hiphop syntax to blend it better with Hop. All Hiphop objects are now Hop values, and thus Hiphop programs can benefit from all Hop featues. The Hiphop development has enabled us to improve Hop stability and quality in client code generation.

We have found a new use-case for Hiphop: Robotics. We are currently working with the Inria Coprin team to pilot their robot using Hop and Hiphop. We have already used Hop to program with low-level motors API (using the Phidget libraries). Hop enabled us to distribute the robot control application over HTTP, in order to control the robot from a smartphone or tablet.

6.3.2. *A CPS definition of HipHop*

Since the Esterel model is used very dynamically in the HipHop framework, we have begun studying new frameworks of computations. We designed a definition of a HipHop-core, which is similar to Esterel-core, based on continuations. This approach allows a specification close to the implementation. The main problem was to define a predicate assuring the *absence* of a specific signal in the current instant. For this, we have designed a static analysis that predicts, for each program point and for each signal, the number of emissions remaining to be done until the end of the instant. The prediction may be over-estimated but when a null value is reached the corresponding signal can be considered as absent for the analyzed instant.

Contrary to existing analyses, this prediction can be done at compile time. Nevertheless some extra computations must be inserted in the evaluator to adjust a runtime prediction. For example. this is done when one branch of a conditional is dynamically taken, but adjusting the prediction only involves subtractions on global counters.

The continuation based definition doesn't prevent a space efficiency implementation. Esterel is known to be compiled to hardware and thus able to run a program in a fixed space of silicon; in the same manner we have implemented an evaluator than doesn't allocate extra memories while running a program: all the continuations can be allocated at compile time.

We have also extended the language to reach the HipHop definition. Some dynamic extensions (`mappar`) may dynamically allocate some resources but we were able to tune the static analysis to insure both confluence and constructive absence detection.

LAGADIC Project-Team

6. New Results

6.1. Visual tracking

6.1.1. 3D model-based tracking
Participants: Antoine Petit, Eric Marchand.

Our 3D model-based tracking algorithm [2] was used in various contexts. We began a collaboration with Astrium EADS in 2010 in order to build a more versatile algorithm able to consider complex objects. The main principle is to align the projection of the 3D model of the object with observations made in the image for providing the relative pose between the camera and the object using a non-linear iterative optimization method. The approach proposed takes advantage of GPU acceleration and 3D rendering. From the rendered model, visible edges are extracted, from both depth and texture discontinuities. Potential applications would be the final phase of space rendezvous mission, in-orbit servicing, large debris removal using visual navigation, or airborne refuelling [41], [40], [32].

6.1.2. Omnidirectional vision system
Participant: Eric Marchand.

In this study performed in collaboration with Guillaume Caron and El Mustapha Mouaddib from Mis in Amiens, we have been interested by the redundancy brought by stereovision in omnidirectional vision sensors. This has been obtained by combining a single camera and multiple mirrors. Within this framework, we proposed to extend the 3D model-based tracking algorithm [2] for such system [15].

Thanks to a collaboration with Esiea in Laval, France, and the Inria and Irisa Hybrid team, we developed a system named Flyviz that has been patented. It is composed of a helmet mounted catadioptric camera coupled with an immersive display. The image acquired by the sensor is processed to give the user a full 360-degree panoramic view [27].

6.1.3. Pose estimation using mutual information
Participant: Eric Marchand.

Our work with Amaury Dame related to template tracking using mutual information [17] as registration criterion has been extended to 3D pose estimation using a 3D model. Since a homography was estimated, the tracking approach presented in [17] was usable for planar scenes. The new approach [45] can be considered for any scene or camera motion. Considering mutual information as similarity criterion, this approach is robust to noise, lighting variations and does not require a statistically robust estimation process. It has been used for visual odometry in large scale environment.

6.1.4. Pseudo-semantic segmentation
Participants: Rafik Sekkal, François Pasteau, Marie Babel.

To address the challenge of tracking initialization issues, we investigate joint segmentation and tracking approaches characterized by resolution and hierarchy scalability as well as a low computational complexity. Through an original scalable Region Adjacency Graph (RAG), regions can be adaptively processed at different scale representations according to the target application [42]. The results of this pseudo-semantic segmentation process are further used to initialize the object tracker (patch, visual objects, planes...) on several scales of resolutions.

6.1.5. Augmented reality using RGB-D camera
Participants: Hideaki Uchiyama, Eric Marchand.

We consider detection and pose estimation methods of texture-less planar objects using RGB-D cameras. It consists in transforming features extracted from the color image to a canonical view using depth data in order to obtain a representation invariant to rotation, scale, and perspective deformations. The approach does not require to generate warped versions of the templates, which is commonly needed by existing object detection techniques [35].

We also investigate the use of RGB-D sensors for object detection and pose estimation from natural features. The proposed method exploits depth information to improve keypoint matching of perspectively distorted images. This is achieved by generating a projective rectification of a patch around the keypoint, which is normalized with respect to perspective distortions and scale [34].

6.2. Visual servoing

6.2.1. *Visual servoing using the sum of conditional variance*
Participants: Bertrand Delabarre, Eric Marchand.

Within our study of direct visual servoing, we propose a new similarity function: the use of the sum of conditional variance [31] that replace SSD or mutual information [3]. It has been shown to be invariant to non-linear illumination variations and inexpensive to compute. Compared to other direct approaches of visual servoing, it is a good trade off between techniques using the pixels luminanc,e which are computationally inexpensive but non robust to illumination variations, and other approaches using the mutual information, which are more complicated to compute but offer more robustness towards the variations of the scene.

6.2.2. *Photometric moment-based visual servoing*
Participants: Manikandan Bakthavatchalam, Eric Marchand, François Chaumette.

The direct visual servoing approaches that have been developed in the group in the recent years, either using the luminance of each pixel, or the mutual information [3], or the sum of conditional variance described just above, allows reaching an excellent positioning accuracy. This good property is however counterbalanced by a small convergence domain due to the strong non linearities involved in the control scheme. To remedy to these problems, we started a study on using photometric moments as inputs of visual servoing. We expect to find again the nice decoupling and large convergence domain that we obtained for binary moments, without the need of any object segmentation.

6.2.3. *Visual servoing using RGB-D sensors*
Participants: Céline Teulière, Eric Marchand.

We propose a novel 3D servoing approach [43] that uses dense depth maps to perform robotic tasks. With respect to pose-based approaches, our method does not require the estimation of the 3D pose, nor the extraction and matching of 3D features. It only requires dense depth maps provided by 3D sensors. Our approach has been validated in servoing experiments using the depth information from a low cost RGB-D sensor. Thanks to the introduction of M-estimator in the control law, positioning tasks are properly achieved despite the noisy measurements, even when partial occlusions or scene modifications occur.

6.2.4. *Visual servoing of cable-driven parallel robot*
Participant: François Chaumette.

This study is realized in collaboration with Rémy Ramadour and Jean-Pierre Merlet from EPI Coprin at Inria Sophia Antipolis. Its goal is to adapt visual servoing techniques for cable-driven parallel robot in order to achieve acurate manipulation tasks. This study is in the scope of the Inria large-scale initiative action Pal (see Section 8.2.7).

6.2.5. *Micro-Nanomanipulation*
Participants: Eric Marchand, Le Cui.

In collaboration with Femto-ST in Besançon, we developed an accurate nanopositioning system based on direct visual servoing [20]. This technique relies only on the pure image signal to design the control law, by using the pixel intensity of each pixel as visual features. The proposed approach has been tested in terms of accuracy and robustness in several experimental conditions. The obtained results have demonstrated a good behavior of the control law and very good positioning accuracy: 89 nm, 14 nm, and 0.001 degrees in the x, y and θ_z axes of a positioning platform, respectively.

We begin a work, within the ANR P2N Nanorobust project (see Section 8.2.4), on the development of micro- and nano-manipulation within SEM (Scanning Electron Microscope). Our goal is to provide visual servoing techniques for positioning and manipulation tasks with a nanometer precision.

6.2.6. *Autonomous landing by visual servoing*

Participants: Laurent Coutard, François Chaumette.

This study was realized in collaboration with Dassault Aviation with the financial support of DGA. It was concerned with the autonomous landing of fixed wing aircrafts on carrier by visual servoing. A complete system has been developed [12]. The vision part consists in detecting the carrier in the image sequence and then tracking it using either dense template tracking or our 3D model-based tracker [2]. The visual servoing part consists in computing particular visual features able to correctly handle the aircraft degrees of freedom. Perturbations due to the wind and carrier motions have also been considered. The complete system has been validated in simulation using synthetic images provided by Xplane simulator and a dynamic model of the aircraft provided by Dassault Aviation.

6.3. Visual navigation of mobile robots

6.3.1. *Visual navigation using mutual information*

Participants: Eric Marchand, Bertrand Delabarre.

We have developed a visual navigation scheme based on the mutual information between the images acquired by an onboard camera and a visual memory to control the orientation of a vehicle during its navigation [18].

We also proposed to extend this approach to visual servoing with vision systems that consider the unified sphere model for central cameras using a normalized version of the mutual information. This permitted to apply the technique to large fields of view with a more reliable similarity function [30].

6.3.2. *3D Mapping and real time navigation*

Participants: Maxime Meilland, Patrick Rives.

This study was realized in collaboration with Andrew Comport from I3S in Sophia Antipolis. Our approach relies on a monocular camera on board the vehicle and the use of a database of spherical images of the scene acquired during an offline step [14]. This geo-referenced database allows us to obtain a robust **drift free** localization. Basically, the database is composed of spherical images augmented by depth that are positioned in a GIS (Geographic information system). This spherical robot centered representation accurately represents all necessary information for vision-based navigation and mapping [37]. During the online navigation, the vehicle pose is computed by aligning the current image acquired by the camera with the closest reference sphere extracted from the database [26].

6.3.3. *Indoors Slam*

Participants: Cyril Joly, Patrick Rives, Pierre Martin, Eric Marchand.

We developed in Sophia Antipolis a new Slam method fusing laser scan data with the spherical images provided by an omnidirectional camera. Thanks to the trace of the laser scan projected onto the spherical view, we are able to compute a RGB-D model of the environment by using a dense visual Slam approach.

In Rennes and in collaboration with Orange Labs, we considered the development of a visual Slam algorithm. Since the targeted platforms in this this study are Android Smartphone, sequential Slam approaches have been studied.

6.3.4. *Topological navigation*

Participants: Alexandre Chapoulie, Patrick Rives.

This study is realized in collaboration with David Filliat from Ensta in Paris. Navigation algorithms are often sensitive to the robot orientation involving an impossibility to detect a place already visited from a different point of view. In order to alleviate this drawback, panoramic or omnidirectional cameras are often used. We have developed a loop closure detection algorithm based on an ego-centric spherical view that satisfies, in addition to other properties, a robot orientation independence [11].

A topological model captures the accessibility of the different places in the environment and allows a coarse localization. From a sequence of spherical views, we have developed a context-based segmentation algorithm. We hence define a topological place as having a structure which does not change, variation leading to a place change. The structure variations are detected with an efficient change-point detection algorithm [28].

6.3.5. *Development of an autonomous shopping cart*

Participants: Luca Marchetti, Patrick Rives.

This work is realized in collaboration with Pascal Morin from Isir in Paris. It consists in developing a shopping cart with autonomy capabilities (automatic user following, obstacle avoidance, etc), as part of the Inria Large-scale initiative action Pal, which aims at developing robotic tools for disabled persons or elderlies (see Section 8.2.7). 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 in Sophia Antipolis [36].

6.3.6. *Automous navigation of wheelchairs*

Participants: Rafik Sekkal, François Pasteau, Marie Babel.

This study is aimed at designing a robotic vision-based system dedicated to assisted navigation of electrical wheelchair in an unkown environment. In particular, going through doors, taking the elevator in a secure way without risking collision because of hazardous wheelchair motions remain a relevant issue. The idea is here to provide an embedded and flexible system able to ensure the immediate compatibility of the proposed system with existing electrical wheelchairs. From the platform described in Section 5.5 , we first addressed the door detection issue for automatically initializing the tracking process that is required for localisation and navigation purposes. We then defined a low complex solution of automatic door recognition that can be decomposed into three successive steps: line extraction (LSD-based algorithm), vanishing point estimation and door recognition itself by using geometrical cues. As soon as a door is detected and tracked through model-based trackers, the idea is to take into account the position of the wheelchair joystick in order to interpret the intention of the user. First experiments have shown the validity of the proposed approach. This study is conducted in conjunction with the scope of the Inria large-scale initiative action Pal (see Section 8.2.7).

6.3.7. *Obstacle avoidance*

Participants: Fabien Spindler, François Chaumette.

This study was realized in collaboration with Andrea Cherubini who is now Assistant Prof. at Université de Montpellier. It is concerned with our long term researches about visual navigation from a visual memory without any accurate 3D localization [9]. In order to deal with obstacle avoidance while preserving the visibility in the visual memory, we have proposed a control scheme based on tentacles for fusing the data provided by a pan-tilt camera and a laser range sensor [16].

6.4. Medical robotics

6.4.1. *Visual servoing based on dense ultrasound information*

Participants: Caroline Nadeau, Alexandre Krupa.

In the context of the ANR USComp project (see Section 8.2.3), we pursued our works on the development of ultrasound image-based visual servoing methods that directly use pixel intensities of the ultrasound image as control inputs. In opposite with methods based on geometrical visual features, this new approach does not require any image segmentation step that is difficult to robustly perform on ultrasound images. By coupling our method with a predictive control law based on the periodicity of physiological motion, we propose a solution to stabilize the ultrasound image by actively compensating the physiological motions of the patient. The principle consists in automatically synchronizing the 6 DOF motion of a 2D or 3D probe with the rigid motion of a soft tissue target. First ex-vivo results obtained on animal tissues demonstrated the validity of the concept [39].

In collaboration with Prof. Pierre Dupont from Harvard University at Boston, we also addressed the motion tracking of a target that can consist of either the tip of a robot inserted on a beating heart or cardiac tissues. Unlike the previous work, where the motion compensation task was realized physically by moving the probe attached to a robotic arm, we propose here to track the motion of the target using a 3D region of interest (ROI) which is automatically moved within the whole volume observed by a 3D probe thanks to our intensity-based ultrasound visual servoing method. In vivo animal experiments were conducted in Children's Hospital at Boston and validated this tracking approach [38].

6.4.2. *Autonomous control modes for ultrasound probe guidance*
Participants: Tao Li, Alexandre Krupa.

In the context of the ANR Prosit project (see Section 8.2.2), we proposed several autonomous control modes in order to assist a doctor during a robotized and teleoperated ultrasound examination (tele-echography). This year we developed an assistance functionality that automatically maintains the visibility of an anatomic element of interest while the doctor teleoperates the 2D ultrasound probe held by the medical robot. The method is based on a multi-task controller that gradually activates an ultrasound visual servoing in case some geometrical features leave a pre-defined safe area of the image in order to bring them back inside the view [33]. With this approach the DOFs of the robotized probe are not exclusively constrained by the visibility task but also available for the tele-operation. This new assistance functionality was implemented on the ANR Prosit robotic platform and first in vivo results obtained on a human volunteer validated the concept.

6.4.3. *Real-time soft-tissue deformation tracking in 3D ultrasound*
Participant: Alexandre Krupa.

We proposed a dense ultrasound tracking algorithm that estimates in real time both rigid and non-rigid motions of a region of interest observed in a sequence of 3D ultrasound images. The deformation is modeled by 3D thin-plate splines (TPS) whose parameters are estimated online from intensity difference measured in successive volumes. To increase the robustness of this approach to image noise, we proposed two solutions to mechanically constrain the deformable model. The first is based on the addition of a regularization term in the TPS model and the second consists in coupling the TPS with a mass-spring system. These methods were validated on simulated sequences of deformed 3D ultrasound images.

6.4.4. *Needle detection and tracking in 3D ultrasound*
Participant: Alexandre Krupa.

We designed an algorithm able to detect a needle inserted manually in a 3D ultrasound volume from an arbitrary point, and able to robustly track this needle in real-time. We also experimentally demonstrated the possibility to guide the ultrasound probe to keep the needle visible and aligned, using visual servoing. Such a system could assist an operator during manual insertions, which are currently performed under free-hand ultrasound monitoring. In addition, we plan in future works to combine this method to a needle steering robotic system for guiding accurately the needle toward a target while optimizing its visibility.

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

5. New Results

5.1. A Backward-Compatible Protocol for Inter-routing over Heterogeneous Overlay Networks

Participants: Giang Ngo Hoang, Luigi Liquori, Vincenzo Ciancaglini, Petar Maksimovic, Hung Nguyen Chan [HUST, Vietnam].
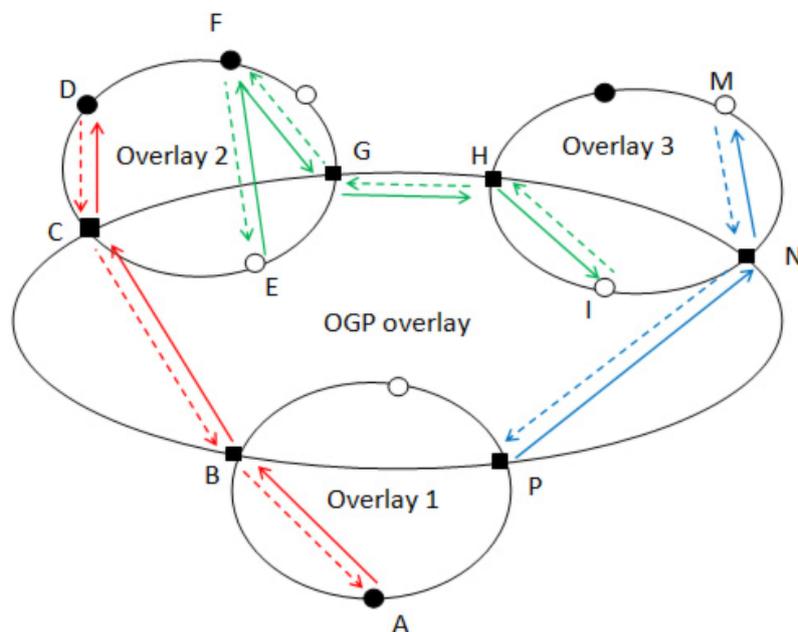
Figure 11. An Overlay Gateway Protocol Topology

Overlay networks are logical networks running on the highest level of the OSI stack: they are applicative networks used by millions of users everyday. In many scenarios, it would be desirable for peers belonging to overlays running different protocols to communicate with each other and exchange certain information. However, due to differences in their respective protocols, this communication is often difficult or even impossible to be achieved efficiently, even if the overlays are sharing common objectives and functionalities. In this paper, we address this problem by presenting a new overlay protocol, called OGP (Overlay Gateway Protocol), allowing different existing networks to route messages between each other in a backward-compatible fashion, by making use of specialized peers joined together into a super-overlay. Experimental results on a large scale Grid5000 infrastructure show that having only a small number of nodes running the OGP protocol is sufficient for achieving efficient routing between heterogeneous overlay networks.

The three scenarios in Figure 11 are shown to illustrate the routing of three lookup queries, in which full OGP peers, lightweight OGP peers and blind peers interact in order to reach across overlays represent requests, while dashed lines represent responses. using the OGP super-overlay. The three smaller ovals represent standard overlays, while the largest oval represents the OGP super-overlay, forwarding messages back and forth between standard overlays. The black squares B; C; G; N and P represent full OGP peers, the black circles A; D and F represent lightweight OGP peers, while the white circles E; H, and M represent blind peers. Solid lines requests, while dashed lines represent responses. The paper is the continuation of the work of HotPost 2011 [7] and it has been accepted to ACM SAC 2013 [33] and a long version will be submitted in a high level conference [34].

5.2. Interconnection of large scale unstructured P2P networks: modeling and analysis

Participants: Rossano Gaeta [Univ. Turin], Riccardo Loti, Luigi Liquori, Vincenzo Ciancaglini [contact].

Interconnection of multiple P2P networks has recently emerged as a viable solution to increase system reliability and fault-tolerance as well as to increase resource availability. In this paper we consider interconnection of large scale unstructured P2P networks by means of special nodes (called *Synapses*) that are co-located in more than one overlay. Synapses act as *trait d'union* by sending/forwarding a query to all the P2P networks they belong to. Modeling and analysis of the resulting interconnected system is crucial to design efficient and effective search algorithms and to control the cost of interconnection. To this end, we develop a generalized random graph based model that is validated against simulations and it is used to investigate the performance of search algorithms for different interconnection costs and to provide some insight in the characteristics of the interconnection of a large number of P2P networks. To overcome this strong limitation, we develop a generalized random graph based model to represent the topology of one unstructured P2P network, the partition of nodes into Synapses, the probabilistic flooding based search algorithms, and the resource popularity. We validate our model against simulations and prove that its predictions are reliable and accurate. We use the model to investigate the performance and the cost of different search strategies in terms of the probability of successfully locating at least one copy of the resource and the number of queries as well as the interconnection cost. We also gain interesting insights on the dependency between interconnection cost and statistical properties of the distribution of Synapses. Finally, we show that thanks to our model we can analyze the performance of a system composed of a large number of P2P networks.

To the best of our knowledge, this is the first paper on model-based analysis of interconnection of large scale unstructured P2P networks [27], [28]

5.3. SIEVE: a distributed, accurate, and robust technique to identify malicious nodes in data dissemination on MANET

Participants: Rossano Gaeta [Univ. Turin], Riccardo Loti [contact], Marco Grangetto [Univ Turin].

We consider the following problem: nodes in a MANET must disseminate data chunks using rateless codes but some nodes are assumed to be malicious, i.e., before transmitting a coded packet they may modify its payload. Nodes receiving corrupted coded packets are prevented from correctly decoding the original chunk. We propose SIEVE, a fully distributed technique to identify malicious nodes.

SIEVE is based on special messages called *checks* that nodes periodically transmit. A check contains the list of nodes identifiers that provided coded packets of a chunk as well as a flag to signal if the chunk has been corrupted. SIEVE operates on top of an otherwise reliable architecture and it is based on the construction of a *factor graph* obtained from the collected checks on which an incremental belief propagation algorithm is run to compute the probability of a node being malicious. Analysis is carried out by detailed simulations using ns-3. We show that SIEVE is very accurate and discuss how nodes speed impacts on its accuracy. We also show SIEVE robustness under several attack scenarios and deceiving actions. The paper has been accepted to [20]

5.4. CCN-TV: a data-centric approach to real-time video services

Participants: Luigi Liquori, Vincenzo Ciancaglini [contact], Riccardo Loti, Giuseppe Piro [Politech Bari], Alfredo Grieco [Politech Bari].

Content Centric Networking is a promising data- centric architecture, based on in-network caching, name-driven routing, and receiver-initiated sessions, which can greatly enhance the way Internet resources are currently used, thus making the support for a broader set of users with increasing traffic demands possible. The CCN vision is, currently, attracting the attention of many researchers across the world, because it has all the potential to become ready to the market, to be gradually deployed in the Internet of today, and to facilitate a graceful transition from a host-centric networking rationale to a more effective data-centric working behavior. At the same time, several issues have to be investigated before CCN can be safely deployed at the Internet scale. They include routing, congestion control, caching operations, name-space planning, and application design. With reference to application-related facets, it is worth to notice that the demand for TV services is growing at an exponential rate over the time, thus requiring a very careful analysis of their performance in CCN architectures. To this end, in the present contribution we deploy a CCN-TV system, able to deliver real-time streaming TV services and we evaluate its performance through a simulation campaign based on real topologies. The paper has been accepted to [19].

5.5. Towards a Trust and Reputation Framework for Social Web Platforms and @-economy

Participants: Thao Nguyen [contact], Bruno Martin [Unice], Luigi Liquori, Karl Hanks.
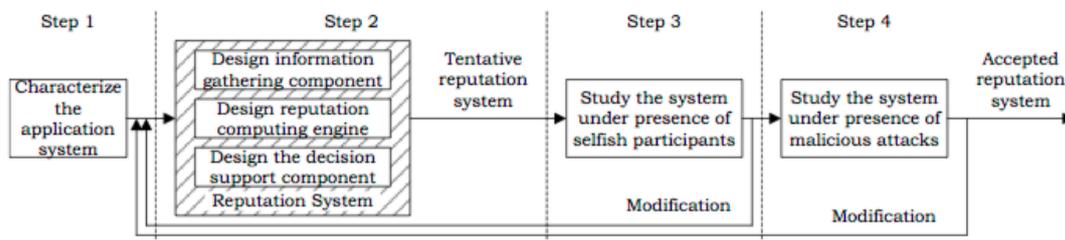
Figure 12. Process of designing a robust trust and reputation system

Trust and reputation systems (TRSs) have recently seen as a vital asset for the safety of online interaction environment. They are present in many practical applications, e.g., e-commerce and social web. A lot of more complicated systems in numerous disciplines also have been studied and proposed in academia. They work as a decision support tool for participants in the system, helping them decide whom to trust and how trustworthy the person is in fulfilling a transaction. They are also an effective mechanism to encourage honesty and cooperation among users, resulting in healthy online markets or communities. The basic idea is to let parties rate each other so that new public knowledge can be created from personal experiences. The greatest challenge in designing a TRS is making it robust against malicious attacks. In this paper, we provide readers an overview on the research topic of TRSs, propose a consistent research agenda in studying and designing a robust TRS, and present an implemented reputation computing engine alongside simulation results, which is our preliminary work to acquire the target of a trust and reputation framework for social web applications.

Information concerning the reputation of individuals has always been spread by word-of-mouth and has been used as an enabler of numerous economic and social activities. Especially now, with the development of technology and, in particular, the Internet, reputation information can be broadcast more easily and faster than ever before. Trust and Reputation Systems (TRSs) have gained the attention of many information and computer scientists since the early 2000s. TRSs have a wide range of applications and are domain specific. The multiple areas where they are applied, include social web platforms, e-commerce, peer-to-peer networks, sensor networks, ad-hoc network routing, and so on. Among these, we are most interested in social web platforms. We observe that trust and reputation is used in many online systems, such as online auction and shopping websites, including eBay, where people buy and sell a broad variety of goods and services, and Amazon, which is a world famous online retailer. Online services with TRSs provide a better safety to their users. A good TRS can also create incentives for good behavior and penalize damaging actions. Markets with the support of TRSs will be healthier, with a variety of prices and quality of service. TRSs are very important for an online community, with respect to the safety of participants, robustness of the network against malicious behavior and for fostering a healthy market.

From a functional point of view, a TRS can be split into three components The first component gathers feedback on participants' past behavior from the transactions that they were involved in. This component includes storing feedback from users after each transaction they take part in. The second component computes reputation scores for participants through a Reputation Computing Engine (RCE), based on the gathered information. The third component processes the reputation scores, implementing appropriate reward and punishment policies if needed, and representing reputation scores in a way which gives as much support as possible to users' decision-making. A TRS can be centralized or distributed. In centralized TRSs, there is a central authority responsible for collecting ratings and computing reputation scores for users. Most of the TRSs currently on the Internet are centralized, for example the feedback system on eBay and customer reviews on Amazon. On the other hand, a distributed TRS has no central authority. Each user has to collect ratings and compute reputation scores for other users himself. Almost all proposed TRSs in the literature are distributed.

Some of the main unwanted behaviors of users that might appear in TRSs are: *free riding* (people are usually not willing to give feedback if they are not given an incentive to do so), *untruthful rating* (users give incorrect feedback either because of malicious intent or because of unintended and uncontrolled variables), *colluding* (a group of users coordinate their behavior to inflate each other's reputation scores or bad-mouth other competitors. Colluding motives are only clear in a specific application), *whitewashing* (a user creates a new identity in the system to replace his old one when the reputation of the old one has gone bad), *milking reputation* (at first, a participant behaves correctly to get a high reputation and then turns bad to make a profit from their high reputation score). The milking reputation behavior is more harmful to social network services and e-commerce than to the others.

This research aims to build on these studies and systematize the process of designing a TRS in general as depicted in Fig. 12 . First, we characterize the application system into which we want to integrate a TRS, and find and identify new elements of information which substitute for traditional signs of trust and reputation in the physical world. Second, based on the characteristics of the application, we find suitable working mechanisms and processes for each component of the TRS. This step should answer the following questions: "What kind of information do we need to collect and how?", "How should the reputation scores be computed using the collected information?", and "How should they be represented and processed to lead users to a correct decision?". To answer the first question, which corresponds to the information gathering component, we should take advantage of information technology to collect the vast amounts of necessary data. An RCE should meet these criteria: *accuracy* for long-term performance (distinguishing a newcomer with unknown quality from a low-quality participant who has stayed in the system for a long time), *weighting* towards recent behavior, *smoothness* (adding any single rating should not change the score significantly), and *robustness* against attacks. Third, we study the tentative design obtained after the second step in the presence of selfish behaviors. During the third step, we can repeatedly return to Step 2 whenever appropriate until the system reaches a desired performance. The fourth step will refine the TRS and make it more robust against malicious attacks. If a modification is made, we should return to Step 2 and check all the conditions in steps 2 and 3 before accepting the modification. The paper has been accepted to [19]

5.6. An Open Logical Framework

Participants: Luigi Liquori [contact], Marina Lenisa [Univ. Udine], Furio Honsell [Univ. Udine], Petar Maksimovic, Ivan Scagnetto [Univ. Udine].

The LFP Framework is an extension of the Harper-Honsell-Plotkin's Edinburgh Logical Framework LF with external predicates, hence the name Open Logical Framework. This is accomplished by defining lock type constructors, which are a sort of "diamond"-modality constructors, releasing their argument under the condition that a possibly external predicate is satisfied on an appropriate typed judgement. Lock types are defined using the standard pattern of constructive type theory, i.e. via introduction, elimination, and equality rules. Using LFP, one can factor out the complexity of encoding specific features of logical systems which would otherwise be awkwardly encoded in LF, e.g. side-conditions in the application of rules in Modal Logics, and sub-structural rules, as in non-commutative Linear Logic. The idea of LFP is that these conditions need only to be specified, while their verification can be delegated to an external proof engine, in the style of the Poincaré Principle or Deduction Modulo. Indeed such paradigms can be adequately formalized in LFP. We investigate and characterize the meta-theoretical properties of the calculus underpinning LFP: strong normalization, confluence, and subject reduction. This latter property holds under the assumption that the predicates are well-behaved, i.e. closed under weakening, permutation, substitution, and reduction in the arguments. Moreover, we provide a canonical presentation of LFP, based on a suitable extension of the notion of $\beta\eta$-long normal form, allowing for smooth formulations of adequacy statements.

LFP is parametric over a potentially unlimited set of (well-behaved) predicates P, which are defined on derivable typing judgements of the form $\Gamma \vdash_\Sigma N : \sigma$, see Fig 13 .

$$\frac{\Gamma \vdash_\Sigma M : \rho \qquad \Gamma \vdash_\Sigma N : \sigma}{\Gamma \vdash_\Sigma \mathcal{L}^\mathcal{P}_{N,\sigma}[M] : \mathcal{L}^\mathcal{P}_{N,\sigma}[\rho]} \qquad \text{(O·Lock)}$$

$$\frac{\Gamma \vdash_\Sigma M : \mathcal{L}^\mathcal{P}_{N,\sigma}[\rho] \qquad \Gamma \vdash_\Sigma N : \sigma \qquad \mathcal{P}(\Gamma \vdash_\Sigma N : \sigma)}{\Gamma \vdash_\Sigma \mathcal{U}^\mathcal{P}_{N,\sigma}[M] : \rho} \qquad \text{(O·Unlock)}$$

Figure 13. Some rule of the Open Logical Framework

The syntax of LFP predicates is not specified, with the main idea being that their truth is to be verified via a call to an external validation tool; one can view this externalization as an oracle call. Thus, LFP allows for the invocation of external "modules" which, in principle, can be executed elsewhere, and whose successful verification can be acknowledged in the system via L-reduction. Pragmatically, lock types allow for the factoring out of the complexity of derivations by delegating the {checking, verification, computation} of such predicates to an external proof engine or tool. The proof terms themselves do not contain explicit evidence for external predicates, but just record that a verification {has to be (lock), has been successfully (unlock)} carried out. In this manner, we combine the reliability of formal proof systems based on constructive type theory with the efficiency of other computer tools, in the style of the Poincaré Principle. In this paper, we develop the meta-theory of LFP. Strong normalization and confluence are proven without any additional assumptions on predicates. For subject reduction, we require the predicates to be well-behaved, i.e. closed under weakening, permutation, substitution, and $\beta\mathcal{L}$-reduction in the arguments. LFP is decidable, if the external predicates are

decidable. We also provide a canonical presentation of LFP, based on a suitable extension of the notion of $\beta\eta$-long normal form. This allows for simple proofs of adequacy of the encodings. In particular, we encode in LFP the call-by-value λ-calculus and discuss a possible extension which supports the design-by-contract paradigm. We provide smooth encodings of side conditions in the rules of Modal Logics, both in Hilbert and Natural Deduction styles. We also encode sub-structural logics, i.e. non-commutative Linear Logic. We also illustrate how LFP can naturally support program correctness systems and Hoare-like logics. In our encodings, we utilize a library of *external predicates*. As far as expressiveness is concerned, LFP is a stepping stone towards a general theory of shallow vs deep encodings, with our encodings being shallow by definition. Clearly, by Church's thesis, all external decidable predicates in LFP can be encoded, possibly with very deep encodings, in standard LF. It would be interesting to state in a precise categorical setting the relationship between such deep internal encodings and the encodings in LFP. LFP can also be viewed as a neat methodology for separating the logical-deductive contents from, on one hand, the verification of structural and syntactical properties, which are often needlessly cumbersome but ultimately computable, or, on the other hand, from more general means of validation.

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

5. New Results

5.1. Network Science

Participants: Eitan Altman, Konstantin Avrachenkov, Mahmoud El Chamie, Philippe Nain, Giovanni Neglia, Marina Sokol.

5.1.1. *Epidemic models of propagation of content*

E. Altman and P. Nain have studied in [96] in collaboration with A. Shwartz (Technion, Israel) and Y. Xu (Univ. Avignon/LIA) the efficiency of the existing methods for reducing availability of non-authorized copyrighted content for free download on the Internet. To model the propagation of the content, they used both branching processes as well as several epidemic models. One of the important finding is that the greatest impact of measures against unauthorized download is obtained whenever some parameter that describes the virality of the content is close to some critical value (which is computed in this work).

5.1.2. *Control and game models for malware attack*

In collaboration with M. H. R. Khouzani (Ohio State Univ., USA) and S. Sarkar (Univ. of Pennsylvania, USA), E. Altman has used in [31],[33], [32], optimal control theory to study malware attack in networks. The structure of optimal policies is obtained by using the Pontryagin maximum principle. In the first two references, optimal defense policies are studies in the goal of protecting the network. In the third work, the worst case behavior of the attack is identified using control theory. The authors then study in [34] the combined problem of identifying the defensive control that achieves the best performance under the worst possible malware attack. This is done through a zero-sum game context.

5.1.3. *Time random walks on time varying graphs*

In collaboration with D. Figueiredo (Federal Univ. of Rio de Janeiro, Brazil), B. Ribeiro and D. Towsley (both from the Univ. of Massachusetts at Amherst, USA), P. Nain has studied the behavior of a continuous time random walk (CTRW) on a stationary and ergodic time varying dynamic graph [57]. Conditions have been established under which the CTRW is a stationary and ergodic process. In general, the stationary distribution of the walker depends on the walker rate and is difficult to characterize. However, the stationary distribution has been characterized in the following cases: i) the walker rate is significantly larger or smaller than the rate in which the graph changes (time-scale separation), ii) the walker rate is proportional to the degree of the node that it resides on (coupled dynamics), and iii) the degrees of nodes belonging to the same connected component are identical (structural constraints). Examples are provided that illustrate these theoretical findings.

5.1.4. *Quick detection of central nodes*

In [50] K. Avrachenkov and M. Sokol, together with N. Litvak (Twente Univ., The Netherlands) and D. Towsley (Univ. of Massachusetts at Amherst, USA) propose a random walk based method to quickly find top k lists of nodes with the largest degrees in large complex networks. The authors show theoretically and by numerical experiments that for large networks the random walk method finds good quality top lists of nodes with high probability and with computational savings of orders of magnitude. They also propose stopping criteria for the random walk method which requires very little knowledge about the structure of the network.

5.1.5. *Graph-based semi-supervised learning*

In [48] K. Avrachenkov and M. Sokol, together with P. Gonçalves (INRIA project-team RESO) and A. Mishenin (St. Petersburg State Univ., Russia) develop a generalized optimization framework for graph-based semi-supervised learning. The framework gives as particular cases the Standard Laplacian, Normalized Laplacian and PageRank based semi-supervised learning methods. The authors provide new probabilistic interpretation based on random walks and characterize the limiting behaviour of the methods. The random walk based interpretation allows one to explain differences between the performances of methods with different smoothing kernels. It appears that the PageRank based method is robust with respect to the choice of the regularization parameter and the labelled data. The theoretical results are illustrated with two realistic datasets, characterizing different challenges: "Les Misérables" characters social network and Wikipedia hyper-link graph. It appears that the PageRank based method can classify the Wikipedia articles with very good precision and perfect recall employing only the information about the hyper-text links.

In [47] K. Avrachenkov and M. Sokol, together with P. Gonçalves (INRIA project-team RESO) and A. Legout (INRIA project-team PLANETE) apply the theoretical results of [48] to classification of content and users in BitTorrent. The general intuition behind the application of the graph based semi-supervised learning methods is that the users with similar interests download similar contents. PageRank based semi-supervised learning method was chosen as it scales well with very large volumes of data. The authors provide recommendations for the choice of parameters in the PageRank based semi-supervised learning method, and show, in particular, that it is advantageous to choose labelled points with large PageRank score.

5.1.6. *Optimal weight selection in average consensus protocols*

In average consensus protocols, nodes in a network perform an iterative weighted average of their estimates and those of their neighbors. The protocol converges to the average of initial estimates of all nodes found in the network. The speed of convergence of average consensus protocols depends on the weights selected on links (to neighbors). In [92] K. Avrachenkov, M. El Chamie and G. Neglia address how to select the weights in a given network in order to have a fast speed of convergence for these protocols. They approximate the problem of optimal weight selection by the minimization of the Schatten p-norm of a matrix with some constraints related to the connectivity of the underlying network. They then provide a totally distributed gradient method to solve the Schatten norm optimization problem. By tuning the parameter p in the proposed minimization, it is possible to simply trade-off the quality of the solution (i.e. the speed of convergence) for communication/computation requirements (in terms of number of messages exchanged and volume of data processed). Simulation results on random graphs and on real networks show that this approach provides very good performance already for values of p that only needs limited information exchange. The weight optimization iterative procedure can also run in parallel with the consensus protocol and form a joint consensus–optimization procedure.

5.1.7. *Reducing communication overhead of average consensus protocols*

The average consensus protocol converges only asymptotically to consensus and implementing a termination algorithm is challenging when nodes are not aware of some global information (e.g. the diameter of the network or the total number of nodes). In [93] K. Avrachenkov, M. El Chamie and G. Neglia propose a totally distributed algorithm for average consensus where nodes send more messages when they have large differences in their estimates, and reduce their message sending rate when the consensus is almost reached. The convergence of the system is guaranteed to be within a predefined margin from the true average and the algorithm gives a trade-off between the precision of consensus and the number of messages send in the network. The proposed algorithm is robust against nodes changing their estimates and can also be applied in dynamic networks with faulty links.

5.2. Wireless Networks

Participants: Eitan Altman, Philippe Nain, Giovanni Neglia.

5.2.1. *Estimation of population sizes in sensor networks*

We have been working on several problems related to the estimation of population sizes. In collaboration with D. Kumar (IBM Research Center, Hawthorne, USA) and T. Başar (Univ. of Illinois at Urbana-Champaign, USA), E. Altman develops in [73] a Wiener filter that allows to estimate the number of sensors that cover the space at some selected points. The authors take advantage of spatial correlations between the number of sensors covering different points in order to derive the filter. We note that causality is not an issue in space, in contrast to filtering at different points in time.

In collaboration with A. Ali, T. Chahed and M. K. Panda (Telecom SudParis, France), D. Fiems (Gent Univ., Belgium), and L. Sassatelli (I3S, Univ. Nice Sophia Antipolis - CNRS, France), E. Altman has used in [37] Kalman filtering theory in order to estimate the number of mobiles in a delay tolerant ad-hoc network which have a copy of a broadcasted message.

5.2.2. *Cellular networks: Small cells*

Analysing performance measures of cellular systems combines tools from queueing theory and stochastic processes, on one hand, and geometric considerations on the other hand. In [72], V. Kavitha (Mymo Wireless, Bangalore, India), S. Ramanath (Lekha Wireless Solutions, Bangalore, India), and E. Altman compute the time it takes to transmit a file taking into account the channel conditions which vary due to mobility of terminals. Mobility considerations play a key role in small cells since handover may occur way before the transmission of the file ends.

5.2.3. *Multi scale fairness concepts for resource allocation in wireless networks*

In many applications that require resources, one needs these resources within some given deadline. These impose constraints when attempting to allocate resources fairly. In [14], E. Altman, K. Avratchenkov and S. Ramanath have extended the α fairness concept by Mo and Walrand so as to include time constraints. They study the question of how to compute such constrained fair allocation, and derive some asymptotic properties of constrained fair assignment.

5.2.4. *Self organization in cellular communications*

Self organization is an approach to design networks so as to allow them to configure in an automatic way. This allows to reduce the complexity in systems containing thousands of mobiles and a huge number of small cells. In cellular networks, self organization can be used for deciding on time or frequency reuse according to the interference in these time and frequency slots from other cells. The impact of self organization on communications are derived in [55] and [21] by R. Combes, and Z. Altman (Orange Labs, Issy les Moulineaux), in collaboration with E. Altman.

5.2.5. *Streaming over wireless*

In [75], E. Altman and M. Haddad study in collaboration with T. Jiménez and R. El-Azouzi (Univ. Avignon/LIA) and S.-E. Elayoubi (Orange Labs, Issy les Moulineaux) streaming service over cellular networks. The purpose is to obtain the exact distribution of the number of buffer starvations within a sequence of N consecutive packet arrivals. This is then applied to optimize the quality of experience (QoE) of media streaming service over cellular networks by exploiting the tradeoff between the start-up delay and the starvation.

5.2.6. *Wireless network security*

The operation of a wireless network relies extensively on exchanging messages over a universally known channel, referred to as the control channel. The network performance can be severely degraded if a jammer launches a denial-of-service (DoS) attack on such a channel.

In [94], P. Nain, M. Krunz, H. Rahbari and M. J. Abdel Rahman (all three from Univ. of Arizona, USA) design frequency hopping (FH) algorithms that mitigate DoS attacks on the control channel of an asynchronous ad hoc network. More specifically, three FH algorithms (called NUDoS, KMDoS, and NCMDoS) are developed for establishing unicast (NUDoS) and multicast (KMDoS and NCMDoS) communications in the presence of multiple jammers. KMDoS and NCMDoS provide different tradeoffs between speed and robustness to node compromise. These algorithms are fully distributed, do not incur any additional message exchange overhead, and can work in the absence of node synchronization. Furthermore, KMDoS and NCMDoS have the attractive feature of maintaining the multicast group consistency. NUDoS exploits the grid quorum system, whereas KMDoS and NCMDoS use the uniform k-arbiter and the Chinese remainder theorem (CRT) quorum systems, respectively. Extensive simulations are used to evaluate these algorithms.

5.3. Network engineering games

Participants: Eitan Altman, Konstantin Avrachenkov, Ilaria Brunetti, Richard Combes, Julien Gaillard, Majed Haddad, Manjesh Kumar Hanawal, Alexandre Reiffers.

5.3.1. Fairness

Anti-trust laws have been introduced by many countries in the last century. This is due to the perception that free competition is better for society. This motivated H. Kameda (Univ. Tsukuba, Japan), C. Touati and A. Legrand (MESCAL, INRIA - CNRS) in cooperation with E. Altman, to define in [28] a fairness concept related to the outcome of competition, which is the Nash equilibrium concept.

5.3.2. Association problem

In [70], E. Altman and M. Haddad study in collaboration with C. Hasan and J.-M. Gorce (SOCRATE, INRIA - INSA) games related to the association problem of mobiles to an access point. It consists of deciding to which access point to connect. Here the choice is between two access points or more, where the access decisions may depend on the number of mobiles connected to each one of the access points. New results were obtained using elementary tools in congestion and crowding games.

5.3.3. Association and placement

The location of a base station has an impact on the throughuput of arriving mobiles that decide to connect to it. Given a cooperative behavior among base stations, E. Altman derives in [54] in collaboration with A. Coluccia (Univ. Salento, Italy) the equilibrium association policy and maximizes its performance by a suitable cooperative positioning of the base stations. The non-cooperative related model was studied in [16] by E. Altman, in collaboration with A. Kumar, C. Singh and R. Sundaresan (all three from IISc, Bangalore, India).

5.3.4. Power control with energy state

In [42] and [64], E. Altman, M. Haddad, J. Gaillard study with D. Fiems (Gent Univ., Belgium) a power control game over a collision channel. Each player has an energy state. When choosing a higher transmission power, the chance of a successful transmission (in the presence of other interference) increases at the cost of a larger decrease in the energy state of the battery. This dynamic game is studied when restricting to simple non-dynamic strategies that consist of choosing a given power level that is maintained during the lifetime of the battery. Surprising paradoxes were identified in the proposed Hawk and Dove game.

5.3.5. Routing games

In [65], M. Haddad, E. Altman and J. Galliard study in collaboration with D. Fiems (Gent Univ., Belgium) a sequential dynamic routing game on a line, where the decision of a user is spatio-temporal control. Each user ships its demand over time on a shared resource. Explicit expressions of the equilibrium of such systems are presented and compared to the global optimum case. The basic idea is taken from a previous paper on this subject by M. K. Hanawal (also with Univ. Avignon/LIA) and E. Altman, in collaboration with R. El-Azouzi (Univ. Avignon/LIA) and B. Prabhu (CNRS - LAAS), who show in [67] that one may transform the time dimension into a spatial component and thus obtain an equivalent standard routing game (where time plays no role) with infinitely many nodes.

5.3.6. *Bayesian games in networking*

We have considered several problems in networks in which decision makers have asymmetrical information. One of these is how one agent may benefit from revealing part of his information? We considered two types of hierarchical scenarios. In the first, we assume that an agent signals some information to another agent who then chooses an action based on that signal. This action determines the utility of both agents. In the second scenario, a player takes an action (such as pricing) and then the second player reacts to it. Both players' utilities depend on the actions of the two players. The action of the first player may reveal to the second player some of his private information. We use the framework of signalling game to solve the first type of problem and that of Bayesian game to solve the second. Other problems include pricing access to the Internet with partial information [52] (by I. Brunetti (Univ. Bologna, Italy), M. Haddad (Univ. Avignon/LIA) and E. Altman). In [45], M. Haddad and E. Altman, in collaboration with P. Wiecek (Wroclaw Univ. of Technology, Poland), apply Bayesian games for the association problem in which users have to decide to which access point to connect.

5.3.7. *Jamming*

We have been working on various models that capture different aspects of jamming (on purpose noise generation). Jamming with partial information is studied in [51] using Bayesian games, by M. Haddad (Univ. Avignon/LIA), E. Altman and S. Azad, as well as [62] and [63] by E. Altman in collaboration with A. Garnaev (St. Petersburg State Univ., Russia) and Y. Hayel (Univ. Avignon/LIA). With K. Avrachenkov, they further consider a dynamic jamming problem in [61]. In all these models the jammer creates interference to the data packets. In [29] V. Kavitha and R. El-Azouzi (Univ. Avignon/LIA), R. Sundaresan (IISc, Bangalore, India), and E. Altman study a different type of jamming game. The jammer attacks the signalling channel and not the data itself. A Bayesian game is formulated and solved there.

5.3.8. *Network neutrality and collusions*

Network neutrality is a key issue in the future Internet. It is related to the question of whether the access to Internet will remain a universal service or whether it would be regulated by market forces according to economic interests of those that control the Internet access. One form of network non-neutrality is when an ISP gives preferential treatment to one content provider over others. We call this "collusion" or "vertical monopoly". In collaboration with T. Jimenez and Y. Hayel (Univ. Avignon/LIA), E. Altman studies this in [71] along with "horizontal monopolies" that may occur when several ISPs merge. They introduce a new concept of "price of collusion" and identify in [44] cases in which not only consumers loose from collusions but also the colliding agents, as also seen in a different model for network non-neutrality given in [69] by M. K. Hanawal (also with Univ. Avignon/LIA) and E. Altman in collaboration with R. Sundaresan (IISc, Bangalore, India). This is related to a special kind of Braess type paradox.

5.3.9. *Competition over popularity in social networks*

We focus on competition of video contents for popularity. We analyze the impact of sharing, embedding, advertisement and other actions by the users for increasing the popularity and visibility. This then allowed E. Altman in [80], [38] and [95] to propose stochastic game models and to fully determine the equilibrium policy. He further proposes a dynamic game for the study of partial information and obtain the equilibrium policies and equilibrium performance. In [39], [79] the results are further extended for the wireless context.

5.3.10. *Stochastic geometry methods for wireless design issues*

Stochastic geometry seems to be the adequate tool in order to model correctly randomness in the location of networks elements such as the mobile terminals and the fixed base stations. Modeling the locations of both as independent spatial processes, In [66] and [25], M. K. Hanawal and E. Altman study in collaboration with F. Baccelli (TREC, INRIA - ENS) properties of Nash equilibria obtained in a multiple access game. They also derive the saddle point obtained in jamming games [68].

5.3.11. In which content to specialize

E. Altman considers in [40] the question of how should a content provider decide in which content to specialize. He shows that the problem is equivalent to the so called "Crowding" games, which allows him to prove the existence of a pure equilibrium. The conclusion is then that there is no gain by diversifying in several contents.

5.3.12. Cognitive radio

In collaboration with J. Elias (Univ. Paris Descartes-Sorbonne) and F. Martignon (LRI-Univ. Paris-Sud), E. Altman study in [56] the question of which priority level to use in a cognitive radio network: higher priority (primary user) or lower one (secondary user). The utilities are function of both the price and the quality of service. After deriving an equilibrium in this game problem, the authors study the question of how to choose prices so as to induce efficient equilibria.

5.3.13. Constrained games

In collaboration with A. Galindo-Serrano and L. Guipponi (CTTC, Spain), E. Altman studies in [60] a game theoretical problem of power control in several base stations with a coupled constraint: the interference at a given point in space should be upper bounded by some constant. The authors establish the existence of a continuum of constrained equilibria to this type of games and show that there is a unique one with some desirable scaling properties (i.e. that consitutes a normalized Nash equilibrium).

5.3.14. Dynamic coalition games

In collaboration with M. K. Panda and T. Chahed (Telecom SudParis, France), E. Altman considers the question of whether to join a multicast session or not. In contrast to many queueing problems, the congestion here is a desirable property, since the cost per user decreases as the number of users connected to the multicast session increases. In [74] the equilibrium policies are derived; these exhibit a surprising structure.

5.3.15. Evolutionary games

The relatively young theory of Evolutionary games considers a large number of interactions between pairs of randomly selected players. It is thus based on a relatively narrow scope in which the one that interacts is the player. In collaboration with Y. Hayel (Univ. Avignon/LIA) and E. V. Belmega (ETIS/ENSEA - Univ. Cergy-Pontoise - CNRS), E. Altman has been developing in [26] an alternative theory of evolutionary game in which a player consists of a group of interacting agents. This is in line with today's understanding of evolution of species (e.g. Dawkins' book "The Selfish Gene" in which the player is the gene of the species). We plan to apply this to energy dependent power control in wireless systems. We also plan to apply these in other areas such as the evolution of languages over social networks, in which some preliminary results (over Twitter) were already obtained in [81] by E. Altman and Y. Portilla (Univ. Avignon/LIA).

5.4. Green networking

Participants: Sara Alouf, Nicaise Choungmo Fofack, Delia Ciullo, Alain Jean-Marie.

5.4.1. Analysis of power saving in cellular networks with continuous connectivity

We have pursued our effort in the analysis of the continuous connectivity mode used in 4G cellular networks. Assuming Poisson traffic at each user, S. Alouf and V. Mancuso (Institute IMDEA Networks, Madrid, Spain) analyze the impact of 3GPP-defined power saving mechanisms on the performance of users with continuous connectivity. Each downlink mobile user's traffic is seen as $M/G/1$ queue, and the base station's downlink traffic as an $M/G/1\,PS$ queue with multiple classes and inhomogeneous vacations. The model is validated through packet-level simulations in [35]; its results show that consistent power saving can be achieved in the wireless access network.

The case of web traffic is investigated in [13] where the same authors, with the participation of N. Choungmo Fofack, perform in addition a sensitivity analysis to assess the impact of model parameters on the performance and cost metrics. It is found that significant power save can be achieved while users are guaranteed to experience high performance. Important outcomes of this work include the need to limit the number of active users in a cell (to less than 350 users – reasonable for 3GPP LTE, 802.16 and HSPA networks) in order to limit the web page download time, and the need to limit the web page size as large pages can dramatically decrease the energy saving. A *green attitude* would be to design web sites with short pages having few embedded objects, enabling cellular operators to use reasonable power save parameters and yet achieve a dramatic cost economy at both base station and mobile user sides, without any quality degradation.

5.4.2. *Analysis of base station sleep modes in cellular networks*

D. Ciullo, L. Chiaraviglio (INRIA project-team MASCOTTE), M. Ajmone Marsan (Politecnico di Torino, Italy and Institute IMDEA Networks, Spain), M. Mellia and M. Meo (Politecnico di Torino, Italy) study in [78] base station sleep modes. Putting into sleep mode some base stations in periods of low traffic improves the energy efficiency of cellular access networks. Two schemes are considered whether the sleep mode is activated once per day or multiple times per day having progressively fewer active base stations. For both schemes, the optimal base station sleep times are identified according to the traffic. Considering real traffic traces, the study reveals that significant energy saving can be achieved, the actual value strongly depending on the traffic pattern. An important result is that most of the potential savings can be attained with a single daily sleep mode, avoiding the increased complexity coming from the use of multiple sleep modes per day.

5.4.3. *Analysis of sleep modes in backbone networks*

The case of backbone networks is considered in [86] where L. Chiaraviglio (INRIA project-team MASCOTTE)), D. Ciullo, M. Mellia and M. Meo (Politecnico di Torino, Italy) formulate a theoretical model based on random graph theory. This model allows to estimate the potential gains achievable by adopting sleep modes in fixed networks where some devices consume energy proportionally to the handled traffic. Putting a given fraction of devices in sleep mode reduces the energy these consume but increases the energy consumed by the devices still active due to the additional load these have to handle. The model of [86] allows to predict how much energy can be saved in different scenarios. The results show that sleep modes can be successfully combined with load proportional solutions. However, if the static power consumption component is one order of magnitude less than the load proportional component, then sleep modes are no longer convenient. Thanks to random graph theory, this model gauges the impact of different properties of the network topology.

5.5. Content-oriented systems

Participants: Konstantin Avrachenkov, Nicaise Choungmo Fofack, Delia Ciullo, Philippe Nain, Giovanni Neglia, Marina Sokol.

5.5.1. *Performance analysis of peer-assisted Video-on-Demand (VoD) systems*

In [88] and [97], D. Ciullo, V. Martina and E. Leonardi (Politecnico di Torino, Italy), M. Garetto (Università di Torino, Italy), and G. L. Torrisi (CNR, Italy) consider peer-assisted Video-on-Demand systems. Some of the essential aspects of such systems are peer churn, bandwidth heterogeneity, and Zipf-like video popularity. The authors propose an analytical framework to tightly characterize the scaling laws for the additional bandwidth that servers must supply to guarantee perfect service, taking into account these essential aspects.

The results in [88] and [97] reveal that the catalog size and the content popularity distribution have a huge effect on the system performance. Also, users' cooperation can effectively reduce the servers' burden for a wide range of system parameters, confirming it as an attractive solution to limit the costs incurred by content providers as the system scales to large populations of users. Moreover, in [89] the same authors provide important hints for the design of efficient peer-assisted VoD systems under server capacity constraints.

5.5.2. *Analysis of TTL-based cache networks*

N. Choungmo Fofack, P. Nain and G. Neglia, together with D. Towsley (Univ. of Massachusetts at Amherst, USA) introduced in [87] a novel Time-To-Live (TTL) replacement policy to manage a set of documents buffering routers in information-centric networks. The TTL policy assigns a timer to each content stored in the cache and redraws the timer at each content request. In [53] they have showed that this TTL policy is more general than other policies like least frequently used (LRU), first-in-first-out (FIFO) or random (RND) as it mimics their behavior under an appropriate choice of its parameters. While exact formulas for the performance metrics of interest (hit/miss processes) are derived for a linear network and a tree network with one root cache and N leaf caches, for more general networks, an approximate solution is found with relative errors smaller than 10^{-3} and 10^{-2} for exponentially distributed and constant TTLs respectively. It is demonstrated in [53] that the TTL model can be implemented and used to optimize a multi-content cache network under realistic constraints such as the cache size limitation.

5.5.3. *CCN interest routing as multi-armed bandit problem*

In [49] K. Avrachenkov and P. Jacko (BCAM, Spain) consider Content Centric Network (CCN) interest forwarding problem as a Multi-Armed Bandit (MAB) problem with delays. The authors investigate the transient behaviour of the ϵ-greedy, tuned ϵ-greedy and Upper Confidence Bound (UCB) interest forwarding policies. Surprisingly, for all the three policies very short initial exploratory phase is needed. It is demonstrated that the tuned ϵ-greedy algorithm is nearly as good as the UCB algorithm, commonly reported as the best currently available algorithm. The uniform logarithmic bound for the tuned ϵ-greedy algorithm in the presence of delays is proved. In addition to its immediate application to CCN interest forwarding, the new theoretical results for MAB problem with delays represent significant theoretical advances in machine learning discipline.

In [46] K. Avrachenkov together with L. Cottatellucci and L. Maggi (both from Eurecom, France) consider the choice of CCN Access Points (APs) when CCN APs are wireless base stations. It is assumed that the slow fading channel attenuations follow an autoregressive model. In the single user case, the authors formulate this selection problem as a restless multi-armed bandit problem and propose two strategies to dynamically select a band at each time slot. The objective is to maximize the SNR in the long run. Each of these strategies is close to the optimal strategy in different regimes. In the general case with several users, the authors formulate the problem as a stochastic game with uncountable state space, where the objective is the SINR. Then the authors propose two strategies to approximate the best response policy for one user when the other users' strategy is fixed.

5.6. Advances in methodological tools

Participants: Eitan Altman, Konstantin Avrachenkov, Alain Jean-Marie, Philippe Nain.

5.6.1. *Perturbation analysis*

In [17] K. Avrachenkov, together with R. Burachik, J. Filar V. Gaitsgory (Univ. of South Australia, Australia), study a linear programming problem with a linear perturbation introduced through a parameter $\epsilon > 0$. The authors identify and analyze an unusual asymptotic phenomenon in such a linear program. Namely, discontinuous limiting behavior of the optimal objective function value of such a linear program may occur even when the rank of the coefficient matrix of the constraints is unchanged by the perturbation. The authors show that, under mild conditions, this phenomenon is a result of the classical Slater constraint qualification being violated at the limit and propose an iterative, constraint augmentation approach for resolving this problem.

5.6.2. *Zero-sum games*

In [18] K. Avrachenkov, together with L. Cottatellucci and L. Maggi (both from Eurecom, France), study zero-sum two-player stochastic games with perfect information. The authors propose two algorithms to find the uniform optimal strategies and one method to compute the optimality range of discount factors. The convergence in finite time for one algorithm is proved. In particular, the uniform optimal strategies are also optimal for the long run average criterion and, in transient games, for the undiscounted criterion as well.

5.6.3. *Approximations in semi-Markov zero-sum games*

In conjunction with E. Della Vecchia and S. Di Marco (both from National Univ. Rosario, Argentina), A. Jean-Marie has pursued the studies on the Rolling Horizon procedure and other approximations in stochastic control problems. Their first study on convergence conditions for average-cost MDPs has been published in [23].

They have then turned to the case of discounted semi-Markov zero-sum games. Generalizing previous contributions of the literature, they have established existence conditions and geometric convergence results when action spaces are compact and rewards possibly unbounded. The bounds they obtain hold for the Rolling Horizon procedure as well as for variants called Approximate Rolling Horizon [91]. In the same semi-Markovian context, they have also performed a sensitivity analysis of the model with respect to its parameters: cost function, discount factor, transition probabilities and state space [90].

5.6.4. *Retrial queues*

In [84] K. Avrachenkov and P. Nain, in collaboration with U. Yechiali (Tel Aviv Univ.), consider a retrial system with two input streams and two orbit queues. More specifically, there are two independent Poisson streams of jobs feeding a single-server service system having a limited common buffer that can hold at most one job. If a type-i job (i=1,2) finds the server busy, it is blocked and routed to a separate type-i retrial (orbit) queue that attempts to re-dispatch its jobs at its specific Poisson rate. This creates a system with three dependent queues. Such a queueing system serves as a model for two competing job streams in a carrier sensing multiple access system. The authors study the queueing system using multi-dimensional probability generating functions, and derive its necessary and sufficient stability conditions while solving a boundary value problem. Various performance measures are calculated and numerical results are presented.

5.6.5. *Branching processes*

In collaboration with D. Fiems (Gent Univ., Belgium), E. Altman introduces in [41] non-standard new branching processes and applies them to evaluate queueing processes. The processes are characterized by replacing the standard Algebra involved in the definition of branching processes by the max-plus algebra. Among the applications introduced are (i) polling systems with infinite server, and (2) new Cruz type bounds for systems with feedback.

Standard branching have been used in the past to study polling systems. In [30] V. Kavitha (LIA/Univ. Avignon) and E. Altman have revisited this method and applied it to spatial sensors, that receive or send data via a mobile relay or base stations. They derive conservation laws for this continuous state space polling system which allows them to compute optimal polling strategies.

D. Fiems (Gent Univ., Belgium) and E. Altman have further used in [24] semi-linear processes, which extend branching processes, to compute expected waiting times in polling systems with generally distributed walking times (the standard i.i.d. assumption is replaced with the assumption that the walking times are stationary ergodic).

In [22], the problem of parallel TCP connections is studied by O. Czerniak and U. Yechiali (Tel Aviv Univ., Israel), in collaboration with E. Altman, for a model in which, when the sum of throughputs reaches some value, there is a loss. It is assumed that the connection to suffer the loss is chosen according to a round robin policy. The expected throughputs of the connections are computed using an approach based on multitype branching processes.

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

5. New Results

5.1. Coq and SMT provers

Participants: Michaël Armand, Benjamin Grégoire, Laurent Théry.

Continuing the work of previous years, we added an extra theory to the interface between Coq and Satisfiability Modulo Theory (SMT) provers: instantiation. It is the last really needed piece to make our tactic based on SMT provers really useful to Coq users. Part of the work was to make the proof work on statements existing in the Propositional type instead of the boolean type. This requires a change in the correctness proof.

5.2. Formal proofs on Pi

Participant: Yves Bertot.

We studied the chain of definitions and proofs necessary to show that

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} \cdots$$

and removed the axiom that was left on this topic in Coq's standard library. This part re-used a past contribution of Guillaume Allais during an internship from Ecole Normale Supérieure de Lyon. We then added a study of Machin's formula to compute decimals of π.

5.3. Formal proofs on linear algebra

Participants: Guillaume Cano, Maxime Dénès, Anders Mörtberg [University of Chalmers, Sweden], Vincent Siles [University of Chalmers, Sweden], Yves Bertot.

This year we completed a work on matrix canonical forms, providing formal proofs for the following results:

- Smith normal forms of matrices on principal ideal domains are unique,
- Every matrix on a field is similar to its Frobenius normal form
- Every matrix on an algebraically closed field is similar to its Jordan normal form

We also studied techniques to combine high-level mathematical descriptions and proofs of algorithms with executable implementations. This work led to a publication at ITP'12 [10]. We are still working on extending this work to rational numbers and real algebraic numbers.

We then worked on tools to automate proofs. In the ring tactic, all elements considered must belong to the same type. We worked on extending this tactic to dependent families of types, like the type of matrices where each dimension gives rise to a different type in the family and multiplications typically concern matrices of different types, while remaining associative.

5.4. Formal proof of the Feit-Thompson theorem

Participants: Laurence Rideau, Laurent Théry.

The Feit-Thompson theorem, established in the beginning of the 1960s, states that every odd-order finite group is solvable. The proof of this result was initially published in an article with around 250 pages. This proof was cleaned by a team of mathematicians and re-published in the form of two books, totaling approximately the same number of pages. But these books also rested on some general knowledge about groups and various areas of algebras.

All this knowledge is now formally described in the Mathematical Components library. The proof of the theorem has been completed in September 2012. The team that achieved this result includes members of the Marelle project-team, along with members of the Typical project-team at Inria Saclay-Ile de France, members of the Microsoft Research Cambridge laboratory, and guests from other institutions.

This year, the members of the Marelle team concentrated on the following topics:

- General character theory: chapters 5 and 6 of the book by Isaacs,
- Character theory for the odd order theorem: chapters 1 to 4 of the book of the book by Peterfalvi.

More information at http://www.msr-inria.inria.fr/Projects/math-components/feit-thompson.

5.5. Native execution for the Coq system

Participants: Maxime Dénès, Benjamin Grégoire, Yves Bertot.

We have continued our work on the native execution of dependently typed terms, aiming at the integration of this work in the main branch of the Coq system.

5.6. Provably correct approximations of elementary functions

Participants: Erik Martin-Dorel, Laurence Rideau, Laurent Théry.

The elementary functions are general purpose mathematical functions that are often implemented in the hardware of modern micro-processors: exponential and trigonometric functions, and inverse functions like arctan or square-root. We participate in a nationally funded project (ANR-TaMaDi) where precise approximations of these functions and their combinations must be computed. A first approach is to use Taylor models. We implemented such an approach and proved its correctness in the Coq system. This led to the publication [9].

We are now working on applying Bernstein polynomials to the problem of approximating transcendental functions.

5.7. Geometric algebras

Participant: Laurent Théry.

We translated our library to the ssreflect setting and provided a very concise certified implementation of geometric algebras based on binary trees.

5.8. Bourbaki in Coq

Participant: José Grimm.

In previous years, we developed a formal library describing the part of the Bourbaki books on set theory, cardinals, and ordinals. The whole development now runs under Coq 8.4, ssreflect 1.4. The main contribution this year is the study of some families of numbers (Stirling numbers of the second kind, Euler numbers, Bell numbers), and their relations to cardinalities (number of partitions of a set, number of partition with p parts, number of surjections $I_n \to I_p$). We have some explicit formulas for $\sum_{i<n} i^k$ as sums of binomial coefficients.

5.9. Reasoning on polynomial expressions

Participants: José Grimm, Julianna Zsido, Yves Bertot.

Continuing previous work by Bertot, we showed that if p is a polynomial on any ordered ring, that has n positive roots, the list of its coefficients has at least n sign changes. If there is exactly one sign change, and the ring is an Archimedian field, there is a number a such that the polynomial is negative on $[0, a]$ and strictly increasing after a; thus it has at most one positive root, and there is a Cauchy sequence x_i such that $p(x_i) < 0$ but $p(x_i + c/2^n) > 0$.

The publication by Bertot, Mahboubi, and Guilhot in 2011 on Bernstein polynomials describes a procedure that works only for polynomials with simple roots. We added the proofs that describe how to obtain such polynomials, starting from arbitrary ones. In other words, we proved the following statement: *for every polynomial p, p divided by the greatest common divisor of p and its derivative has the same roots as p and all the roots are simple.*

We started working on a proof that the dichotomy process based on Bernstein polynomials is bound to terminate, concentrating on a theorem known as *the theorem of three circles*.

5.10. Higher-Order Abstract Syntax

Participant: Julianna Zsido.

With Martin Hyland from the University of Cambridge, we worked on an approach to reconcile the points of view of Fiore, Plotkin, and Turi on the one hand and Hirschowitz and Maggesi on the other hand. This approach relies on a large monad that abstracts over the two approaches.

5.11. Proofs in cryptography

Participants: Gilles Barthe [IMDEA Software Institute], Juan Manuel Crespo [IMDEA Software Institute], Benjamin Grégoire, Sylvain Heraud [Prove&Run], César Kunz [IMDEA Software Institute], Yassine Lakhnech [University of Grenoble], Pierre-Yves Strub [IMDEA Software Institute], Santiago Zanella Béguelin [IMDEA Software Institute].

We are continuing our work on providing a user-friendly tool for cryptographers who want to develop formal proofs of correctness, based on Certicrypt and SMT provers. There were invited talks at ITP, CPP, MPP, SAS, and JFLA. There was also an article in ERCIM news, whose contents is more oriented towards the open public. See also the web page http://easycrypt.gforge.inria.fr/.

As an illustrative example, we proposed a machine-checked proof of a construction of a hash function based on elliptic curves, where the correctness proof uses the Random Oracle Model. The proof is based on an extension of CertiCrypt for reasoning about approximate forms of observational equivalence and uses mathematical results from group theory and elliptic curves.

Thanks to our language-based approach to describing cryptographic constructions and our automatic approach to proving them correct, we can now explore systematically the space of possible designs. Using this approach, we have been able to explore over 1.3 million schemes, including more than 100 variants of OAEP studied in the literature and to prove the correctness of 250,000 schemes for one kind of model and 17,000 for another kind.

MASCOTTE Project-Team

6. New Results

6.1. Network Design and Management

Participants: Gianlorenzo D'Angelo, Jean-Claude Bermond, Khoa Phan, David Coudert, Frédéric Giroire, Joanna Moulierac, Nicolas Nisse, Ronan Pardo Soares, Stéphane Pérennes, Issam Tahiri.

6.1.1. Network Design

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. 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. All-Optical Label Switching, AOLS

All-Optical Label Switching (AOLS) is a promising technology that performs packet forwarding without any optical-electrical-optical conversions, thus speeding up the forwarding. However, the cost of this technology requires limiting the number of labels needed to ensure the forwarding when routing a set of requests using GMPLS technology. In particular, this prevents the usage of label swapping techniques.

We have studied the routing problem in this context using label stacking techniques. We have formalized the problem by associating to each routing strategy a logical hypergraph, called a hypergraph layout, whose hyperarcs are dipaths of the physical graph, called tunnels in GMPLS terminology. We defined a cost function for the hypergraph layout, depending on its total length plus its total hop count. Minimizing the cost of the design of an AOLS network can then be expressed as finding a minimum cost hypergraph layout. In [24], we prove hardness results for the problem. On the other hand, we provide approximation algorithms, in particular an $O(\log n)$-approximation for symmetric directed networks. We focused on the case where the physical network is a directed path, providing a polynomial-time dynamic programming algorithm first for one source, and then for a fixed number k of sources running in time $O(n^{k+2})$.

6.1.1.2. Protocols

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, 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 [59], [60], a hybrid model of Application Layer Multicast (ALM) and Xcast6 that can work for large multicast groups. We have shown the feasibility of our model by simulation and comparison with IP multicast and NICE protocols.

Congestion control is a distributed algorithm to share network bandwidth among competing users on the Internet. In the common case, quick response time for mice traffic (http traffic) is desired when mixed with elephant traffic (ftp traffic). The current approach using loss-based with Additive Increase, Multiplicative Decrease (AIMD) is too greedy and eventually, most of the network bandwidth would be consumed by elephant traffic. As a result, it causes longer response time for mice traffic because there is no room left at the routers. MaxNet is a new TCP congestion control architecture using an explicit signal to control transmission rate at the source node. In [60] , we show that MaxNet can control well the queue length at routers and therefore the response time to http traffic is several times faster than with TCP Reno/RED.

6.1.1.3. Shared Risk Link Group

The notion of *Shared Risk Link Group*, SRLG has been introduced to capture multiple correlated failures in a network. A SRLG is a set of links that fail simultaneously if a given event (risk) occurs. In such multiple failures scenario, the problem of Diverse Routing consists in finding two SRLG-disjoint paths between a pair of nodes. We consider in [42], [66] such problem for localized failures, when all the links of a SRLG verify the star property i.e. when they are incident to the same node. We prove that in this case the problem is in general NP-complete and determine some polynomial cases.

6.1.1.4. Data Gathering in Radio Networks

We study the problem of gathering information from the nodes of a radio network into a central node. We model the network of possible transmissions by a graph and consider a binary model of interference in which two transmissions interfere if the distance in the graph from the sender of one transmission to the receiver of the other is d_I or less. A *round* is a set of non-interfering transmissions. In [25], we determine the exact number of rounds required to gather one piece of information from each node of a square two-dimensional grid into the central node. The even case uses a method based on linear programming duality to prove the lower bound, and sophisticated algorithms using the symmetry of the grid and non-shortest paths to establish the matching upper bound. We then generalize our results to hexagonal grids.

Other results on multi-interface networks were obtained outside of MASCOTTE [30], [31], [55].

6.1.2. Routing

The problem of finding and updating shortest paths in distributed networks is considered crucial in today's practical applications. In the recent past, there has been a renewed interest in designing new efficient distance-vector algorithms (e.g., the distributed Bellman-Ford method implemented in the routing information protocol, RIP) as an alternative to link-state solutions (e.g., open shortest path first, OSPF) for large-scale distributed networks such as the autonomous systems topology of the Internet.

This year, we have proposed a new loop-free distance-vector routing algorithm, called LFR (Loop Free Routing), which is able to update the shortest paths of a distributed network with n nodes in fully dynamic scenarios [47]. We compared experimentally this new algorithm with DUAL, one of the most popular loop-free distance vector algorithms which is part of CISCO's EIGRP protocol. Our experiments on CAIDA IPv4 routed /24 topology dataset show that LFR out-performs DUAL in terms of memory requirements and number of messages.

We then proposed a new technique, called Distributed Computation Pruning (DCP) [48], for reducing the total number of messages sent and the space occupancy per node of every distance-vector routing algorithm based on shortest paths. We have evaluated experimentally the combination of DCP with DUAL and with LFR. We have observed that these combinations lead to a significant gain both in terms of number of messages sent and memory requirements per node.

We have also considered routing problems arising in road networs. In particular, we have conducted a theoretical study of the graph-augmentation problem of adding shortcuts in order to speedup route planning techniques [23]. We studied the algorithmic complexity of the problem and proposed approximation algorithms for a special graph class. We have also investigated ILP-based exact approaches and show how to stochastically evaluate a given shortcut assignment on graphs that are too large to do so exactly.

6.1.2.1. Compact routing

With the constant increase of the number of routing entries in the Internet, the size of the routing tables stored at router nodes increases drastically. Routing schemes such as BGP are showing their limits in terms of update time, search time, cost of signaling, etc. and alternatives have to be proposed. In particular, compact routing schemes propose interesting trade-offs between the size of the routing tables and the quality of the routes. They also take advantage of the particular properties arising in large scale networks such as low (logarithmic) diameter and high clustering coefficient.

High clustering coefficient implies the existence of few large induced cycles. Considering this fact, we proposed in [37] a routing scheme that computes short routes in the class of k-chordal graphs, i.e., graphs with no induced cycles of length more than k. Our routing scheme achieves an additive stretch of at most $k - 1$, and the routing tables are computed with a distributed algorithm which uses messages of size $O(\log n)$ and takes $O(D)$ time, where D is the diameter of the network.

We also used *cops-and-robber* games (See Section 6.2.1.2) to propose the first compact routing scheme for k-chordal graphs using routing tables, addresses and headers of size $O(\log n)$ bits and achieving an additive stretch of $O(k \log \Delta)$ [58], [57], [77]. This scheme is based on a new structural decomposition for a graph class including k-chordal graphs: we proposed a quadratic algorithm that, given a graph G and $k \geq 3$, either returns an induced cycle larger than k in G, or computes a *tree-decomposition* of G, each *bag* of which contains a dominating path with at most $k - 1$ vertices. We thus proved that any k-chordal graph with maximum degree Δ has treewidth at most $(k - 1)(\Delta - 1) + 2$, improving the $O(\Delta(\Delta - 1)^{k-3})$ bound of Bodlaender and Thilikos (1997). Moreover, any graph admitting such a tree-decomposition has small hyperbolicity.

In addition, we have pursued our investigation of the kind of structural graph properties that can or cannot be deduced from local (partial) views of the network. Such knowledge is crucial for the design of routing schemes. To this end, we have exhibited a hierarchy of problems and distributed models of computation [40].

6.1.2.2. Routing models evaluation

The evaluation of new routing models asks for large-scale and intensive simulations. However, existing routing models simulators such as DRMSim are limited in terms of the number of routing table entries it can dynamically process and control on a single computer. Therefore, we have conducted a feasibility study of the extension of DRMSim so as to support the Distributed Parallel Discrete Event paradigm [46]. We have studied several distribution models and their associated communication overhead. We have in particular evaluated the expected additional time (in hours) required by a distributed simulation of BGP (border gate protocol), the current interdomain routing protocol of the Internet, compared to its sequential simulation. We show that such a distributed simulation of BGP is possible with a reasonable time overhead.

6.1.2.3. Reconfiguration

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 consists of scheduling 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 and established NP-completeness and inapproximability results.

This year, we proved the monotonicity of the *process strategy* game [78], the graph searching game modeling the routing reconfiguration problem. Then, we have investigated on the influence of physical layer impairment constraints on the reconfiguration problem [41]. Setting up a new wavelength in a fiber of a WDM network requires recalibrating the other wavelengths passing through this fiber. This induces a cost (e.g., time, energy, degradation of QoS) that depends nonlinearly on the number of wavelengths using the fiber. Therefore, the order in which requests are switched affects the total cost of the operation. We have studied the corresponding optimization problem by modeling the cost of switching a request as a non-linear function depending on the load of the links used by the new lightpath. We have proved that determining the optimal rerouting order is NP-complete for a 2-nodes network, established general lower and upper bounds, identified classes of instances where the problem can be solved in polynomial time, and proposed a heuristic algorithm.

6.1.3. Energy efficiency

Recently, energy-aware routing has gained increasing popularity in the networking research community. The idea is that traffic demands are aggregated over a subset of the network links, allowing other links to be turned off to save energy. We develop several methods to improve routing protocols for backbone, wireless and content delivery networks. 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,... The goal thus is to find a routing that minimizes the (weighted) number of active network elements used when routing. In [62], we exhibit that the power consumption can be reduced of approximately 33 MWh for a medium-sized backbone network.

In [54], we propose GreenRE - a new energy-aware routing model with the support of the new technique of data redundancy elimination (RE). Based on real experiments on Orange Labs platform and on simulations on several network topologies, we show that GreenRE can gain further 30% energy savings in comparison with the traditional energy-aware routing model.

One of the new challenges facing research in wireless networks is the design of algorithms and protocols that are energy aware. In [33], we use for the first time the evolving graph combinatorial model as a tool to prove an NP-Completeness result, namely that computing a Minimum Spanning Tree of a planar network in the presence of mobility is actually NP-Complete.

Recently, there is a trend to introduce content caches as an inherent capacity of network equipment, with the objective of improving the efficiency of content distribution and reducing network congestion. In [63], we study the impact of using in-network caches and CDN cooperation on an energy-efficient routing: up to 23% of power can be saved in the backbone this way.

In [32], we study the energy efficiency of the networking part of data centers, accounting for between 10-20% of the total power consumption. We proposed a novel approach, called VMPlanner, for power reduction in the virtualization-based data centers. The idea of VMPlanner is to optimize both virtual machine placement and traffic flow routing so as to turn off as many unneeded network elements as possible for power saving.

Finally, in [56], [38], we summarize the main research results of the last years for energy efficiency for backbone, wireless, cellular and content distribution networks and highlight the main challenges of the field. Results are given for two operator networks, considering power and traffic forecasts for 2020.

6.2. Graph Theory

Participants: Julio Araújo, Jean-Claude Bermond, Frédéric Giroire, Frédéric Havet, František Kardoš, Ana Karolinna Maia, Remigiusz Modrzejewski, Leonardo Sampaio, Michel Syska.

6.2.1. Algorithms in graphs

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.

6.2.1.1. 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.

Degree Constraint Subgraphs. A natural question in current social networks is *How do one find a small community (subgraph) in which anyone as at least d friends (neighbors)?* This problem can be modelled as degree-constrained subgraph problems where the objective is to find an optimal weighted subgraph, subject to certain degree constraints (in which each node has degree at most d), in a weighted graph. When $d = 2$, the problem is easy to solve since one simply needs to compute the girth of the graph. In [16], we proved that the problem is not in Apx for any $d \geq 3$. The proof is obtained by a reduction from Vertex Cover in regular graphs, followed by the use of an error amplification technique. On the positive side, we give an $\frac{n}{\log n}$-approximation

algorithm for the class of graphs excluding a fixed graph H as a minor (including planar or bounded genus graphs), using dynamic programming.

Hyperbolicity in Large graphs. Hyperbolicity is a geometric notion that measure how the various shortest paths connecting two vertices can diverge in a graph. Knowing its value provides information on the geometry of the network, moreover it has practical implications for shortest path routing. Hyperbolicity can be computed in polynomial time algorithm ($\Theta(n^4)$). This is far from being practical for large graphs. So, in [69] we proposed a scalable algorithm for this problem. We also led some computational experiments of our algorithms on large-scale graphs.

Hull Number of graphs. In [64], we study the (geodesic) hull number of graphs. For any two vertices $u, v \in V$ of a connected undirected graph $G = (V, E)$, the closed interval $I[u, v]$ of u and v is the set of vertices that belong to some shortest (u, v)-path. For any $S \subseteq V$, let $I[S] = \bigcup_{u,v \in S} I[u, v]$. A subset $S \subseteq V$ is (geodesically) convex if $I[S] = S$. Given a subset $S \subseteq V$, the convex hull $I_h[S]$ of S is the smallest convex set that contains S. We say that S is a hull set of G if $I_h[S] = V$. The size of a minimum hull set of G is the hull number of G, denoted by $hn(G)$. First, we show a polynomial-time algorithm to compute the hull number of any P_5-free triangle-free graph. Then, we present four reduction rules based on vertices with the same neighborhood. We use these reduction rules to propose a fixed parameter tractable algorithm to compute the hull number of any graph G, where the parameter can be the size of a vertex cover of G or, more generally, its neighborhood diversity, and we also use these reductions to characterize the hull number of the lexicographic product of any two graphs. More on the hull number of graphs may be found in Araujo's thesis [13].

6.2.1.2. 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.

Web Caching & the surfer Game. A surprising application of some variant of pursuit-evasion games (namely Cops and Robber games) is the problem for a web-browser to download documents in advance while an internaut is surfing on the Web. In [53], [52], 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.

Connected Graph Searching. Another variant of pursuit-evasion games is graph searching which is mainly related to graph decompositions. For instance, the minimum number of searchers needed to capture an invisble fugitive in a graph is equal to its pathwidth plus one. In [21], we investigated the connected variant of this game. A strategy is called connected if the clear part (the part where the fugitive cannot stand) always induces a connected subgraph. The main motivation for studying connected graph searching is the design of distributed protocols allowing searchers to compute a capture strategy (see also Section 6.2.1.3). [21] gathers most of the results of the last decade concerning connected graph searching, mainly focussing on the cost of connectivity in terms of number of searchers.

6.2.1.3. 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 6.2.1.2) to adress some models of distributed computing. We considered distributed or even self-stabilizing algoritms for gathering and graph searching problems.

Graph Searching and Routing Reconfiguration. In [29], we developed a generic distributed algorithm for computing and updating various parameters on trees including the process number (see Section 6.1.2.3), and other related graph searching parameters (see Section 6.2.1.2). We also proposed an incremental version of the algorithm allowing to update these parameters after addition or deletion of any tree edge.

Robots in anonymous networks. Motivated by the understanding of the limits of distributed computing, 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 various problems such as graph exploration, graph searching and gathering in various graph classes. 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. More precisely, we design algorithms for the gathering in rings [51], [70], grid [50] and trees [61]. We also proposed a general approach [71] to solve the three problems in rings. Finally, in [67], [44], [43], algorithms are designed to solve the graph searching problem in trees.

6.2.2. *Structural graph theory*

6.2.2.1. *Directed graphs*

Graph theory can be roughly partitioned into two branches: the areas of undirected graphs and directed graphs (digraphs). Even though both areas have numerous important applications, for various reasons, undirected graphs have been studied much more extensively than directed graphs. One of the reasons is that many problems for digraphs are much more difficult than their analogues for directed graphs. For example, one of the cornerstones of modern (undirected) graph theory is Minor Theory of Robertson and Seymour. Unfortunately, we cannot expect an equivalent for directed graphs. Minor Theory implies in particular that, for any fixed H, detecting a subdivision of H in an input graph G can be performed in polynomial time by the Robertson and Seymour linkage algorithm. In contrast, the analogous subdivision problem for digraph can be either polynomial-time solvable or NP-complete, depending on the fixed digraph H. In [65], we give a number of examples of polynomial instances, several NP-completeness proofs as well as a number of conjectures and open problems. We also investigated the related problem in which we want to detect an *induced* subdivision of H. Already, for undirected graphs the complexity of this problem depends on H. In [20], we show that for digraphs the complexity of this problem depends on H and on whether the input digraph G 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.

In a directed graph, a *star* is an arborescence with at least one arc, in which the root dominates all the other vertices. A *galaxy* is a vertex-disjoint union of stars. In [34], we consider the Spanning Galaxy problem of deciding whether a digraph D has a spanning galaxy or not. We show that although this problem is NP-complete (even when restricted to acyclic digraphs), it becomes polynomial-time solvable when restricted to strong digraphs. In fact, we prove that restricted to this class, the Spanning Galaxy problem is equivalent to the problem of deciding if a strong digraph has a strong digraph with an even number of vertices. We then show a polynomial-time algorithm to solve this problem. We also consider some parameterized versions of the Spanning Galaxy problem. Finally, we improve some results concerning the notion of *directed star arboricity* of a digraph D, denoted $dst(D)$, which is the minimum number of galaxies needed to cover all the arcs of D. We show in particular that $dst(D) \leq \Delta(D) + 1$ for every digraph D and that $dst(D) \leq \Delta(D)$ for every acyclic digraph D.

Hypergraphs are a generalization of graphs, in which every edge is incident to a set of vertices of any size (not necessarily 2). Like for digraphs, a lot fewer is known about them than about graphs. The two notions of eulerian and hamoltinians cycles have been extensively studied for graphs and digraphs. The analogue notion of eulerian cycle in a hypergraph was only introduced in 2010 by Lonc and Naroski. In [72], we introduce the notions of eulerian and hamiltonian circuits in directed hypergraphs. We show that both associated decision problems are NP-complete. Some necessary conditions for a dihypergraph to be have an eulerian circuit are presented. We exhibit some families of hypergraphs for which those are sufficient conditions. We also generalize a part of the properties of eulerian digraphs to the uniform and regular directed hypergraphs. Finally, we show that the de Bruijn and Kautz dihypergraphs are eulerian and hamiltonian in most cases.

6.2.2.2. *Graph colouring*

We mainly study graph colouring problems that model channel assignment problems.

A well-known such general problem is the following: we are given a graph G, whose vertices correspond to transmitters, together with an edge-weghting w. The weight of an edge corresponds to the minimum separation between the channels on its endvertices to avoid interferences. (If there is no edge, no separation is required, the transmitters do not interfere.) We need to assign positive integers (corresponding to channels) to the vertices so that for every edge e the channels assigned to its endvertices differ by at least $w(e)$. The goal is to minimize the largest integer used, which corresponds to minimizing the *span* of the used bandwidth.

We mainly studied a particular, yet quite general, case, called *backbone colouring*, in which there are only two levels of interference. So we are given a graph G and a subgraph H, called *the backone*. Two adjacent vertices in H must get integers at least q apart, while adjacent vertices in G must get integers at distance at least 1. The minimum span is this case is called the q-backbone chromatic number and is denoted $BBC_q(G, H)$. Backbone forests in planar graphs are of particular interests. In [74], we give a series of NP-hardness results as well as upper bounds for $BBC_q(G, H)$, depending on the type of the forest (matching, galaxy, spanning tree). Eventually, we discuss a circular version of the problem. In [73], we also consider a list version of the problem in which every vertex must be assigned an integer in its own list of available ones. We provide bounds using the Combinatorial Nullstellensatz for the list version on the channel assignment problem. Through this result and through structural approaches, we obtain good upper bounds for forests and matching backbone in planar graphs. In [68], we give an evidence to a conjecture of Broersma et al. stating that $BBC_2(G, T) \leq 6$, for every planar graph G and spanning tree T. We prove this conjecture in the particular case when T has diameter at most 4.

Another meaningful and very well-studied particular case of backbone colouring is $L(p, 1)$-labelling, which is p-backbone colouring of (G^2, G), where G^2 is the square of G (the graph with same vertex set as G, in which two vertices are adjacents if they are at distance at most 2 in G). Griggs and Yeh conjecture in 1992, that for every graph with maximum degree $\Delta \geq 2$, $BBC_2(G^2, G) \leq \Delta^2 + 1$. In [36], we prove this conjecture when Δ is large. In fact, we prove a more general statement. We prove for any q and sufficiently large Δ, if $\Delta(H) \leq \Delta^2$ and $\Delta(G) \leq \Delta$, then $BBC_q(H, G) \leq \Delta^2 + 1$. Our result also holds for the list version.

In [17], we studied another colouring problem motivated by a practical frequency assignment problem and, up to our best knowledge, new. In wireless networks, a node interferes with other nodes, the level of interference depending on numerous parameters: distance between the nodes, geographical topography, obstacles,... We model this with a weighted graph (G, w) where the weight function w on the edges of G represents the noise (interference) between the two end-vertices. The total interference in a node is the sum of all the noises of the nodes emitting on the same frequency. A *weighted t-improper k-colouring* of (G, w) is a k-colouring of the nodes of G (assignment of k frequencies) such that the interference at each node does not exceed the threshold t. We consider the Weighted Improper Colouring problem which consists in determining the *weighted t-improper chromatic number* defined as the minimum integer k such that (G, w) admits a weighted t-improper k-colouring. We also consider the dual problem, denoted the Threshold Improper Colouring problem, where, given a number k of colours, we want to determine the minimum real t such that (G, w) admits a weighted t-improper k-colouring. We show that both problems are NP-hard and present general upper bounds for both problems; in particular we show a generalisation of Lovász's Theorem for the weighted t-improper chromatic number. 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 neighbours and a noise of intensity $1/2$ for the nodes at distance two. We derive the weighted t-improper chromatic number for all values of t.

Since some of the channel assignment problems must be done on-line, we are interested in some on-line graph colouring heuristics. We only studied such heuristics for the classical proper colouring. The easiest one, and the most widespread one, is the greedy algorithm, which colours the vertices one after another, giving to each vertex the smallest possible positive integer that is not already used by one of its neighbours. The *Grundy number* of a graph G is the largest number of colours used by any execution of the greedy algorithm to colour G. In [27], we give new bounds on the Grundy number of the different product of two graphs. The problem of determining the Grundy number of G is polynomial-time solvable if G is a P_4-free graph and NP-hard if G is a P_5-free graph. In [19], we define a new class of graphs, the *fat-extended P_4-laden graphs*, and we show a polynomial-time algorithm to determine the Grundy number of any graph in this class. Our class intersects the

class of P_5-free graphs and strictly contains the class of P_4-free graphs. More precisely, our result implies that the Grundy number can be computed in polynomial time for any graph of the following classes: P_4-reducible, extended P_4-reducible, P_4-sparse, extended P_4-sparse, P_4-extendible, P_4-lite, P_4-tidy, P_4-laden and extended P_4-laden, which are all strictly contained in the fat-extended P_4-laden class.

A colouring c of a graph $G = (V, E)$ is a *b-colouring* if in every colour class there is a vertex whose neighborhood intersects every other colour classes. Such a colouring appears, when we try to optimize on-line the colouring of a graph, by changing the colour of all vertices of a colour class if it is possible. The *b-chromatic number* of G, denoted $\chi_b(G)$, is the greatest integer k such that G admits a b-coloring with k colours. A graph G is *tight* if it has exactly $m(G)$ vertices of degree $m(G) - 1$, where $m(G)$ is the largest integer m such that G has at least m vertices of degree at least $m - 1$. Determining the b-chromatic number of a tight graph had been shown to be NP-hard even for a connected bipartite graph. In [35], we show that it is also NP-hard for a tight chordal graph, and that the b-chromatic number of a split graph can be computed in polynomial time. Then we define the b-closure and the partial b-closure of a tight graph, and use these concepts to give a characterization of tight graphs whose b-chromatic number is equal to $m(G)$. This characterization is used to develop polynomial-time algorithms for deciding whether $\chi_b(G) = m(G)$, for tight graphs that are complement of bipartite graphs, P_4-sparse and block graphs. We generalize the concept of pivoted tree introduced by Irving and Manlove and show its relation with the b-chromatic number of tight graphs.

Many more results on greedy colourings and b-colourings have been proved in Sampaio's thesis [14].

We studied other variations of graph colouring. 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 (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 $\{C_3, K_{2,3}\}$-free graphs.

For a connected graph G of order at least 3 and a k-labelling $c : E(G) \to \{1, 2, \cdots, k\}$ of the edges of G, the *code* of a vertex v of G is the ordered k-tuple $(n_1, \cdots, n_k)$, where n_i is the number of edges incident with v that are labelled i. The k-labelling c is detectable if every two adjacent vertices of G have distinct codes. The minimum positive integer k for which G has a detectable k-labelling is the *detection number* of G. In [76], we show that it is NP-complete to decide if the detection number of a cubic graph is 2. We also show that the detection number of every bipartite graph of minimum degree at least 3 is at most 2. Finally, we give some sufficient condition for a cubic graph to have detection number 3.

MCTAO Team

5. New Results

5.1. Optimal control for quantum systems: the contrast problem in NMR

These studies aim at optimizing the contrast in Nuclear Magnetic Resonance imaging using advanced optimal control.

5.1.1. Theoretical aspects

Participants: Bernard Bonnard, John Marriott, Monique Chyba [University of Hawaii], Gautier Picot [University of Hawaii], Olivier Cots, Jean-Baptiste Caillau.

This is done in collaboration with University of Hawaii, and deals with many theoretical aspects of the contrast problem in NMR: analysis of the optimal flow, feedback classification in relation with the relaxation times of the species. This activity has been the object of two publications [5], [4], and a conference talk [14] on feedback classification in the contrast problem, that will be followed by a journal article.

John Marriott will defend his Phd thesis on this topic, august 28, 2013; This will be followed by a two day conference on quantum control systems with applications, supported by a NSF grant and by the Engineering Department (P.E. Crouch).

5.1.2. Experimental aspects

Participants: Bernard Bonnard, Olivier Cots, Dominique Sugny [Univ. de Bourgogne], Steffan Glaser [TU München].

As said in section 4.2 , our work on this problem is based on experiments conducted in Prof. S. Glaser in Munich. Experiments using our techniques and measuring the improvement between materials that have an importance in medicine, like oxygenated and de-oxygenated blood have been conducted successfully, see [7], [9].

5.1.3. Numerical aspects

Participants: Bernard Bonnard, Olivier Cots, Jean-Baptiste Caillau.

In december, Pierre Martinon and Mathieu Caeys visited our group. This launhes a collaboration whose objective is to compare the direct and indirect methods in the contrast problem (implemented in the Bocop and Hampath sofwares) and use LMI techniques to get a global bound on the problem (in the contrast problem there are many local optima and the global optimality is a complicated issue)-also O. Cots visited R. Zidani (COMMANDS team) to investigate the use of numerical HJB techniques in the problem. This collaboration will allow to compare in a physical important problem the various available numerical methods in optimal control.

5.2. Conjugate and cut loci computations and applications

Participants: Bernard Bonnard, Olivier Cots, Jean-Baptiste Caillau.

One of the most important results obtained by B. Bonnard and his collaborators concern the explicit computations of conjugate and cut loci on surfaces. This has applications in optimal control to compute the global optimum and in optimal transport where regularity properties of the map in the Monge problem is related to convexity properties of the tangent injectivity domains. This shows also the transverse part of the team: [3] complete the previous results obtained with Rifford [33] ; the paper [20] analyses the conjugate and cut loci in Serret-Andoyer metrics and dynamics of spin particles with Ising coupling, and is a first step towards the computation of conjugate and cut loci on left invariant Riemannian and SR- metrics in S0(3) with applications for instance to the attitude control problem of a spacecraft. The submitted paper [19] concerns the analysis of singular metrics on surfaces in relation with the average orbital transfer problem.

5.3. Averaging in control

Participants: Bernard Bonnard, Helen-Clare Henninger, Jean-Baptiste Pomet.

A reference paper on the construction and properties of an "average control system" [2] is to be published; it is based on Alex Bombrun's doctoral work (2007). It connects solutions of highly oscillating control systems to those of an average control system, when the frequency of oscillation goes high.

This average system in the case of minimum time for low thrust orbit transfer in the two body problem is currently being explored, in particular the study of its inherent singularities. Helen Henninger's PhD aims at going much further in this direction and then apply this local study to real missions, possibly in a three-body environment.

5.4. Optimal transport

Participants: Ludovic Rifford, Alice Erlinger, Ahed Hindawi, Alessio Figalli, Bernard Bonnard, Jean-Baptiste Caillau, Lionel Jassionesse, Robert Mc Cann [U. of Toronto].

This year has seen new results or starting directions in many areas of optimal cotrol.

- The very general condition for continuity of the transport map given in [47] motivated exploration of conditions for convexity of the tangent injectivity domain [10], [3] on. Lionel Jassionnesse's PhD is in part devoted to Ma-Tudinger-Wang tensor that also plays an important role in this matter.

- In Ahed Hindawi's PhD [1], defended this year, results in optimal transport for sub-Riemannian costs (see the survey [16]) are generalized to costs coming from optimal control problems with quadratic cost and a drift.

- Alice Erlinger's PhD, joint with University of Toronto is exploring Optimal Transport's application to modeling in economics

5.5. Applications of control methods to lDynamical systems

Participants: Ludovic Rifford, Ayadi Lazrag, Riccardo Ruggiero, Alessio Figalli, Rafael Ruggiero [PUC, Rio de Janeiro].

Ludovic Rifford and collaborators have been applying, with success, techniques from geometric control theory to open problems in dynamical systems. Mostly on genericity properties and using controllability methods to build suitable perturbations See [11], [13], [21].

Ayadi Lazrag's PhD also deals with such problems

MODEMIC Project-Team

6. New Results

6.1. Theoretical results

6.1.1. Models resource/consumer

The team maintains a significant activity about the theory of the chemostat model, proposing and studying extensions of the classical models.

6.1.1.1. Theory of competition and coexistence
Participants: Jérôme Harmand, Claude Lobry, Tewfik Sari.

In the papers [41], [50] we consider deterministic models of competition. We study the persistence of species. In [25] we study a syntrophic relation between microbial species. In [26], 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.

6.1.1.2. Study of interconnected chemostats
Participants: Jérôme Harmand, Alain Rapaport.

We have shown how a particular spatial structure with a buffer globally stabilizes the chemostat dynamics with non-monotonic response function, while this is not possible with single, serial or parallel chemostats of the same total volume and input flow. We give a characterization of the set of such configurations that enjoy this property, as well as the configuration that ensures the best nutrient conversion. Furthermore, we characterize the minimal buffer volume to add to a single chemostat for obtaining the global stability. These results are illustrated with the Haldane function that models inhibition in micro-organisms growth [67].

In industrial applications, the attraction of the wash-out equilibrium is undesired because it presents a risk that may ruin the culture in case of disturbance, temporarily pump breakdown or presence of toxic material that could drive the state in the attracting basin of the wash-out equilibrium. This approach has led to a patent deposit by INRA [59] during the PhD of H. Haidar, a former PhD student of the team [80].

6.1.1.3. Aggregation models in the chemostat
Participants: Radhouene Fekih-Salem, Jérôme Harmand, Claude Lobry, Alain Rapaport, Tewfik Sari.

We have studied a model of the chemostat where the species are present in two forms, isolated and aggregated individuals, such as attached bacteria in biofilm or bacteria in flocks. We show that our general model contains a lot of models that were previously considered in the literature. Assuming that flocculation and deflocculation dynamics are fast compared to the growth of the species, we construct a reduced chemostat-like model in which both the growth functions and the apparent dilution rate depend on the density of the species. We also show that such a model involving monotonic growth rates may exhibit bi-stability, while it may occur in the classical chemostat model, but when the growth rate is non monotonic [21], [54]. This work is part of the PhD of R. Fekih-Salem co-supervised by A. Rapaport and T. Sari.

This research subject has been mainly motivated by the DISCO project (see Section 7.3).

6.1.1.4. Overyielding in continuous bioprocesses
Participants: Denis Dochain, Alain Rapaport.

We have shown that for certain configurations of two chemostats fed in parallel, the presence of two different species in each tank can improve the yield of the whole process, compared to the same configuration having the same species in each volume. This leads to a (so-called) "transgressive over-yielding" due to spatialization [35].

This work has been achieved during the stay of Prof. P. de Leenheer (Univ. Florida).

6.1.2. *Measuring taxonomic diversity of microbial communities*

Participant: Bart Haegeman.

Diversity is considered to be a main determinant of the behavior of microbial communities. However, measuring microbial diversity is challenging. Although metagenomic techniques allow us to sample microbial communities at unprecedented depths, the disparity between community (e.g., 10^{15} organisms) and sample (e.g., 10^5 organisms) remains large. We have studied what the diversity observed in a sample tells us about the real diversity of the community.

For a given empirical sample the aim is to construct the community from which this sample was taken. It turns out that a large set of community structures are consistent with the sample data. Some diversity metrics vary widely over this set of consistent communities, and are therefore difficult to infer from the sample data. Other diversity metrics are approximately constant over the set of consistent communities, and are therefore much easier to infer from the sample data.

The analysis of the set of consistent communities has yielded the following insights. First, it is impossible to robustly estimate the number of species from sample data. This is easy to understand. Microbial communities typically contain a large number of rare species, and these rare species are unlikely to be present in the sample. Hence, sample data are lacking crucial information to estimate species richness. Second, other diversity metrics, in particular Shannon and Simpson diversity, can be robustly estimated from sample data. We have constructed lower and upper estimates for a general class of diversity metrics, and we have shown that the difference between the extremal estimators, that is, the estimation uncertainty, is small for Shannon and Simpson diversity.

6.1.3. *A theory of genetic diversity within bacterial species*

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.

In [23] we propose a theory for the frequency distribution of genes within a set of genomes. The model describes the genetic diversity as a balance between two forces. Demographic processes such as division and death tend to reduce the genetic diversity; horizontal gene transfer from other species can increase the genetic diversity. Our model predicts that the gene frequency distribution is U-shaped, meaning that there are a large number of genes present in only a few genomes, a small number of genes present in about half of the genomes, and a large number of genes present in almost all genomes. This prediction is consistent with the gene frequency distributions of six bacterial species we have analyzed (about 100 sequenced genomes in total). Importantly, the model does not assume any functional difference between the genes, that is, genes are considered to be selectively neutral. By showing that empirical gene frequency distributions can be reproduced by a neutral genome model, we contend that the frequency of a gene should not be interpreted as an indication of its essentiality or importance.

6.1.4. *Individual-based modelling for bacterial ecosystems*

Participants: Fabien Campillo, Chloé Deygout, Coralie Fritsch, Jérôme Harmand, 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 [78]. 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 many studies in the field of IBMs. The first is part of the ANR MODECOL on the modelling of clonal plant growth (see Section 7.4); the second is part of the ANR DISCO on modelling of biofilms (see Section 7.3); the third is also part of the ANR DISCO is dedicated to the modelling of biofilms in plug-flow reactors (see Section 6.2.2); the last one is the ongoing thesis of Coralie Fritsch at the École Doctorale I2E of the University of Montpellier 2, under the supervision of Fabien Campillo and Jérôme Harmand. The thesis aims at developing and analyzing "simple" individual-based microbial ecosystems models.

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 6.1.5 .

6.1.5. *Stochastic/discrete and stochastic/continuous modelling for biotechnology and population dynamics*
Participants: Fabien Campillo, Marc Joannides, Claude Lobry.

In [14], we consider a stochastic model of the two-dimensional chemostat as a diffusion process for the concentration of substrate and the concentration of biomass. The model allows for the washout phenomenon: the disappearance of the biomass inside the chemostat. We establish the Fokker-Planck equation associated with this diffusion process, in particular we describe the boundary conditions that modelize the washout. We propose an adapted finite difference scheme for the approximation of the solution of the Fokker-Planck equation.

In [15], we consider a hybrid version of the classical predator-prey differential equation model. The proposed model is hybrid: continuous/discrete and deterministic/stochastic. It 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 discrete model – 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. This means that we must be very cautious when we use continuous variables in place of jump processes in dynamic population.

6.1.6. *Optimal control of continuous bioprocesses*
Participants: Walid Bouhafs, Amel Ghouali, Jérôme Harmand, Alain Rapaport.

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 proven 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 [39]. Another problem arises for alternating aerobic-anoxic systems. Revisiting the results obtained several years ago within the framework of Djalel Mazouni's thesis, we aims at proposing optimal time control policies for sequencing batch reactors in which simultaneous nitrification and denitrification take place. The solution for the original problem is a difficult task but we have already proposed a solution for a slightly modified model [30]. These last results have been obtained within the framework of the PhD thesis of W. Bouhafs.

Reference points in batch processes can be mimicked by a series of continuously stirred bioreactors in series at steady state (see applications 6.2.4 and 7.1). 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.

One important issue in bioprocesses is to find optimal feedback control laws in order to steer a system describing a perfectly-mixed bioreactor to a given target value in a minimal amount of time. Finding adequate feeding strategies can significantly improve the performance of the system. A typical target (for fed-batch bioreactors) is to consider the substrate concentration to be lower than a given reference value at the end of the process. Other criterium can be studied such as the maximization of the production of biomass in a given time period. Singular strategies (in reference to the theory of singular arcs in optimal control theory) can be defined in this context and are natural due to the constraints on the system. One objective of our research is to characterize optimal feedback control laws using mathematical tools from optimal control laws, and also to develop numerical methods that can handle problems with a large number of parameters.

6.1.7. *Modelling the functioning of soil ecosystems*

Participants: Céline Casenave, Jérôme Harmand, Alain Rapaport.

In ecology, one of the important challenges is the understanding of the biodiversity observed in the natural ecosystems. Several models have been proposed to explain this biodiversity, and in particular the coexistence of different species. In these models, it is often assumed that, when they die, the micro-organisms are directly converted in an assimilable resource, that is a resource that alive organisms can consume. However, we know that it is not the case in reality. Indeed, the organic matter stemmed from the dead organisms has to be transformed in assimilable resource before it can be consumed. This transformation is performed by some micro-organisms which have this special ability.

We have proposed a new model of soil ecosystems, of chemostat type. This model is rather simple, but also original because it takes into account several mechanisms:

- the growth, the mortality and the respiration,
- the ability of some organisms to transform the non assimilable resources in assimilable ones,
- the advantage that an organism can gain from this ability of transformation.

For the moment, we have considered the case where only one or two types of organisms are present. The model is finally composed of 3 (or 4) nonlinear ordinary differential equations: one per type of organisms, one for the assimilable resource and one for the non assimilable one. The study of the equilibrium points has first shown the possibility of coexistence, at equilibrium, of the two organisms. Then, in numerical simulations, we have observed the possible existence of limit cycles, which can also explain the coexistence of organisms observed in the nature.

This problem is still under study; a working group (in particular with researchers of the UMR Eco & Sols, conducted by B. Jaillard) meet regularly to discuss about the problems of modelling in ecology.

We have also investigated the *sampling effect* that occurs when micro-biologists select randomly species in a natural ecosystem for studying reconstituted ecosystems in a controlled environment. We have proposed a very simple probabilistic model, that shows that observing average increases or decreases on the performances of these reconstitute ecosystems when modifying the size of the sampling, allows to infer kinds and proportions of the interactions among species present in the original ecosystem [65]. This research is conducted with the UMR Eco & Sols, Montpellier, (B. Jaillard) and the UMR BIOEMCO, Grignon (N. Nunan).

6.1.8. *Equivalence between models of fractured porous media*

Participants: Céline Casenave, Jérôme Harmand, Alain Rapaport, Alejandro Rojas-Palma.

In geosciences, models of fractured porous media are often described as a *mobile* zone driven by advection, and one or several *immobile zones* directly or indirectly connected to the mobile zone by diffusion terms. We believe that these models are also relevant to describe flows in soil or in porous media such as biofilms. They are very close from the spatial representations used in Section 6.2.6 . We have shown, using a transfer function approach, that two schemes often used in the literature: the MINC (Multiple INteractive Continua) where diffusive compartments are connected in series, and the MRMT (Multiple Rate Mass Transfer) where diffusive compartments are connected in star around the mobile zone, are equivalent input-output representations [56], and providing formulas (up to three compartments) to pass from one representation to another. This result means that one can simply choose the most convenient approach when dealing with control or optimization without any loss of generality. We are currently working on the general case of n compartments with n larger than three, and looking for equivalent classes of configurations that could be half way between MINC and MRMT and fit better the spatial representations of real world.

This work is performed with the UMR GéoSciences Rennes (J.R. de Dreuzy), and has led to the internship of a Chilean MsC student (A. Rojas-Palma).

6.1.9. *Non-linear filtering for the chemostat*

Participants: Boumediene Benyahia, Amine Boutoub, Fabien Campillo, Jérôme Harmand.

Monte Carlo-based inference methods like particle filtering are bound to develop in the context of biotechnology. In contrast with the classical observer approach, inference through Monte Carlo methods can handle measurements in discrete time in low frequency and with low signal-to- noise ratio. Based on the stochastic modeling of the chemostat, these approaches may also be used for model selection and hypothesis testing.

In a preliminary work [28] we consider the bootstrap particle filter applied to a 2-dimensional chemostat model. The internship of Amine Boutoub dedicated to the study of particle filtering for more realistic chemostat models has recently started.

6.1.10. *Functional identification of growth functions in the chemostat*

Participant: Alain Rapaport.

We have proposed an adaptive control law that allows one to identify unstable steady states of the open-loop system in the single-species chemostat model without the knowledge of the growth function. We have then shown how to use a continuation technique to reconstruct the whole graph of the growth function, providing a new method for identifying non-monotonic growths [42], [27]. Two variants, in continuous and discrete time, have been studied. An analysis of the case of two species in competition has shown the ability of the method to detect a mixed culture for which dominance depends on the dilution rate, due to a property of stability loss in slow-fast dynamics. This method leaves open future extensions for extremum seeking problems.

This work has been conducted in cooperation with Universities of Exeter (J. Sieber) and Plymouth (S. Rodrigues), and the EPI SISYPHE (M. Desroches).

6.1.11. *Diffusive representation of integro-differential models*

Participant: Céline Casenave.

This work is done in collaboration with Emmanuel Montseny (LAAS/CNRS), Gérard Montseny (LAAS/CNRS), and Christophe Prieur (LIAFA/CNRS).

In lots of dynamic systems of Physics or others scientific fields such as Biology (Volterra models), dynamic integral operators, often of convolution type, are involved. Problems relating to integro-differential models are often difficult to solve, especially because these models are not time-local. In this context, the methodology called "diffusive representation" presents some interests: an integral operator is represented by its gamma-symbol, directly deduced from its transfer function. It can be formulated by means of a state realization whose dimension is numerically reasonable whatever the size of the system may be. In addition to this interesting practical side, the diffusive representation offers a unified mathematical framework, well adapted to analysis of integral convolution operators.

Several dynamic problems can be tackled in an original and quite simple way by using the diffusive representation. In fact, all the operatorial problems of modeling, simulation, control, model identification, model reduction, etc. can be formulated in such a way that the object of the problem is the gamma-symbol of the operator solution.

Several problems are under study:
- the identification of integro-differential models [66],
- the controllability of some SISO Volterra models [63],
- the simulation and the analysis of a model of porous media[64].

These works follow up on the ones developed during the PhD thesis of Céline Casenave, which deals with the problem of the operator inversion for the application to non local dynamic problems.

6.2. Applications

6.2.1. *Modelling and control of Anaerobic Digestion processes*

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

We consider the AM2 or AMOCO model developed in [72] 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 [12]. Following this study, it was shown that the overloading tolerance, a parameter proposed in [81] 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) [16], [29].
- 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) [36], [37]; (ii) we try to find links between complex models such as the ADM1 model and simple models such as the AM2b or the AM3 [40].

Apart from this work on the modelling of anaerobic digesters and membrane bioreactors, we have developed a number of specific simple models for control design accounting for the coupling of such processes with membrane modules in the chemostat (PhD thesis of G. Pimentel). This work aims at contributing to the efficient treatment of wastewaters produced in fish production farms. The work of G. Pimentel aims at studying the coupling of simple fouling models with the model of the chemostat in order to propose new simple models for control design.

6.2.2. *Hybrid modelling of biofilms in plug-flow reactors*

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

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.

Such a "hybrid" approach allows for modelling and understanding in plug-flow reactors 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 and time) of the substrate density.

This work has been conducted within the DISCO project (see Section 7.3) and the postdoctoral stay of C. Deygout hired by the project, in close collaboration with A. Lesne (LPTMC, Univ. Paris VI). A first paper on the simulation model has been published [17].

Within the DISCO project, experiments on real tubular plug-flow reactors have been simultaneously driven at IRSTEA Antony with the perspective of comparison with numerical simulations.

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

6.2.3. *Individual-based models for the bacterial degradation of the cellulose*
Participants: Fabien Campillo, Chloé Deygout.

We propose an individual-based model for the degradation of one cellulose bead (dozens of micrometers in diameter) by cellulolytic bacteria. This model accounts for biofilm formation with minimal hypotheses: soluble substrate diffusion combined with bacterial chemotaxis-like movement in the liquid phase, lack of bacterial movement in the solid phase. The IBM results are qualitatively different from the main macroscopic degradation models previously used for cellulose degradation. It suggests that random and discrete processes could significantly impact the cellulose degradation dynamics by their effect on the spatial structuration of the colonized cellulose particles [44].

6.2.4. *Modelling and control of cascade biosystems to mimic batch wine making processes*
Participants: Térence Bayen, Céline Casenave, Jérôme Harmand, Alain Rapaport, Matthieu Sebbah.

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". The input constraint $Q_i \geq Q_{i-1} \geq 0$ is unusual in optimal control problems that are linear w.r.t. to the control. The optimality of candidate singular arcs is still open for this problem.

This summer, some experiments have been made to test the first feedback law on the experimental setup. The control law seems to perform work, but other experiments should be made next year with more reliable input flow pumps.

This work was conducted as a part of the European CAFE project (Computer-Aided Food processes for control Engineering) described in Section 7.1 .

6.2.5. *Modelling and control of an ice cream crystallization process*
Participants: Céline Casenave, Denis Dochain.

In the ice cream industry, the type of final desired product (large cartons or ice creams on a stick) determine the viscosity at which the ice cream has to be produced. The control the viscosity of the ice cream at the outlet of a continuous crystallizer is therefore an important challenge. The problem has been studied in two steps. First, we have completed and validated on experimental data the reduced order model of the system. This model has been obtained by application of the method of moments on a population balance equation describing the evolution of the crystal size distribution. Then, we have proposed a nonlinear control strategy to control of the viscosity of the ice cream with the temperature of the refrigerant fluid of the crystallizer. This control strategy is based on a linearizing control law coupled with a Smith predictor to account for the measurement delay. The control has been validated on an experimental pilot plant located at IRSTEA (Antony, France).

This work was conducted as a part of the European CAFE project (Computer-Aided Food processes for control Engineering) described in Section 7.1 .

6.2.6. *Bioremediation of natural resources*
Participants: Sébastien Barbier, Jérôme Harmand, Alain Rapaport, Antoine Rousseau.

The objective of this work is to provide efficient strategies for the bioremediation of natural water resources. The originality of the approach is to couple minimal time strategies that are determined on a simplified model with a faithful numerical model for the hydrodynamics. This work has been carried out in close cooperation with A. Rousseau. Based on a previous paper that deals with an implicit representation of the spatial inhomogeneity of the resource with a small number of homogeneous compartments (with a system of ODEs), we have implemented a coupled ODE-PDE system that accounts for the spatial non-homogeneity of pollution in natural resources. The main idea is to implement a Navier-Stokes model in the resource (such as a lake), with boundary conditions that correspond to the output feedback that has been determined to be optimal for the simple ODEs model of a (small) bioreactor. A first mathematical model has been introduced and numerical simulations have been performed in academic situations. During the internship of S. Barbier (co-advised by A. Rousseau and A. Rapaport) we built a reduced model that approximates the reference PDE model thanks to a set of ODEs with parameters. Numerical optimization is performed on these parameters in order to better fit the reference model. This will lead to a publication.

The study of the minimal time strategies on the system of ODEs has been mainly achieved in cooperation with Chilean researchers (P. Gajardo, Universidad Tecnica Federico Santa Maria, and H. Ramirez, Centro de Modelamiento Matemático) and a Chilean PhD student (V. Riquelme, Depto. Ingenieria Matematica, Universidad de Chile) within the associated team DYMECOS [57].

6.2.7. *Modelling and simulating terrestrial plant ecological dynamics*
Participant: Fabien Campillo.

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. The first semester of 2012 was the last part of the project. We propose a very original individual-based model for clonal plant dynamics in continuous time and space that focuses on the effects of the network structure of the plants on the reproductive strategy of ramets. The model is coupled with a PDE dynamics for resources. The basic structure of the IBM encompass a population of "ramets" (the individuals) connected by "stolons or rhizomes" (the network) [13], [22]. See http://www-sop.inria.fr/members/Fabien.Campillo/software/ibm-clonal/ for more details.

6.2.8. *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 [32] a simple Markov model has been successfully confronted to a first data set with the help of maximum likelihood and Bayesian approaches. On a new data set developed by IRD, the Markov hypothesis has been rejected and we proposed to use semi-Makov models: for this new dataset the law of the sojourn time on certain states will depend on the next state visited, which is incompatible with the Markov hypothesis and which is consistent with the semi-Markov hypothesis.

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

5. New Results

5.1. Imaging

5.1.1. *ML estimation of wavelet regularization hyperparameters in inverse problems*
Participant: Laure Blanc-Féraud.

This work was made in collaboration with Caroline Chaux from LATP (Marseille) and Roberto Cavicchioli and Luca Zanni from University of Modena (Italy).

Parameter estimation, Maximum likelihood estimation, Wavelet transforms, Deconvolution, Gradi- ent methods

We are interested in regularizing hyperparameter estimation by maximum likelihood in inverse problems with wavelet regularization. One parameter per subband is estimated by gradient ascent algorithm.We have to face with two main difficulties: i) sampling the a posteriori image distribution to compute the gradient of the objective function; ii) choosing a suited step-size to ensure good convergence properties of the gradient ascent algorithm. We first show that introducing an auxiliary variable makes the sampling feasible using classical Metropolis-Hastings algorithm and Gibbs sampler. Secondly, we propose an adaptive step-size selection and a line-search strategy to improve the gradient-based method. Good performances of the proposed approach are demonstrated on both synthetic and real data.
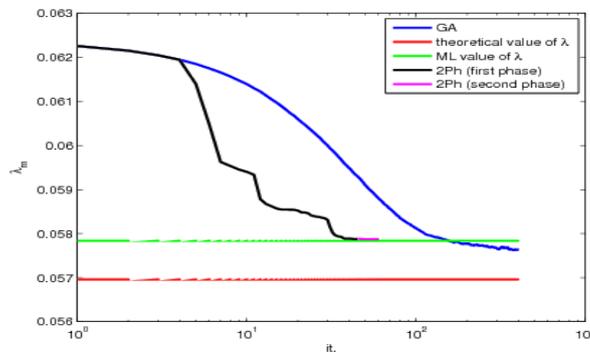
Figure 1. λ_m behavior over iterations of ascent algorithm for a sub band at first level of wavelet decomposition

5.1.2. *Joint optimization of noisy image coding and denoising*
Participants: Mikael Carlavan, Laure Blanc-Féraud.

this work was made in collaboration with Marc Antonini (I3S), Roberto Camarero and Christophe Latry (CNES) and Yves Bobichon (TAS).

coding, denoising, wavelet transform, global rate-distortion optimization

This work concerns the study of optimal noisy source coding/denoising. A global optimization of the problem is usually difficult to perform as the global fidelity criterion needs to be optimized in the same time over the sets of both coding and denoising parameters. Most of the bibliography in this domain is based on the fact that, for a specific criterion, the global optimization problem can be simply separated into two independent optimization problems: The noisy image should be first optimally denoised and this denoised image should then be optimally coded. In many applications however, the layout of the acquisition imaging chain is fixed and can not be changed, that is a denoising step can not be inserted before coding. For this reason, we are concerned here with the problem of global joint optimization in the case the denoising step is performed, as usual, after coding/decoding. In this configuration, we showed on a simple case how to express the global distortion as a function of the coding and denoising parameters. We presented an algorithm to minimize this distortion to get the optimal values of these parameters. Figure 2 shows results of this joint optimization algorithm, on the classical test image *Barbara*, in comparison to the usual disjoint optimization technique, which consists in selecting the coding and the denoising parameters such that the coding and the denoising errors are independently minimized. On the range of validity of the proposed model, we see that the joint optimized distortion slightly outperforms the disjoint optimized distortion (in the presented example, the PSNR of the reconstructed image increases of $0.4dB$ at 1.85 bits/pixels). The interesting point of the proposed method is that it allows to reach the same global error than the disjoint optimized technique but for a lower coding rate. For example, on this image, the joint optimization technique reaches at 1.42 bits/pixel the same distortion than the one obtained at 2.04 bits/pixels for the disjoint optimization technique. The benefit in terms of compression performances of the joint optimization appears then to be very significant.
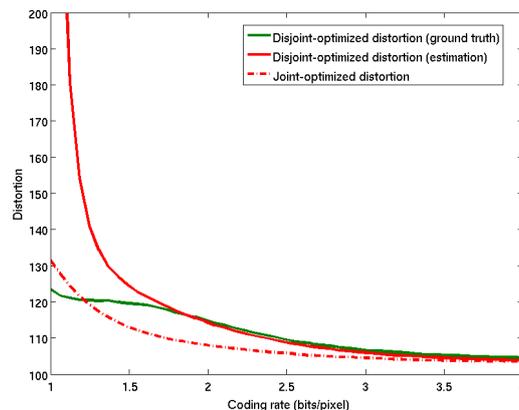
Figure 2. Comparison of the disjoint optimized distortion (ground truth and estimation) to the joint optimized distortion Barbara.

5.1.3. *Blind deconvolution*

Participants: Saima Ben Hadj, Laure Blanc-Féraud.

This research takes place within the ANR DIAMOND. This work was made in collaboration with Gilles Aubert, Laboratoire J. Dieudonné (CNRS,UNS).

One of our tasks within the ANR Diamond project is the blind restoration of images coming from Confocal laser scanning microscopy (CLSM). CLSM is a powerful technique for studying biological specimens in three dimensions by optical sectioning. Nevertheless, it suffers from some artifacts. First, CLSM images are affected by a depth-variant (DV) blur due to spherical aberrations induced by refractive index mismatch between the different media composing the system as well as the specimen. Second, CLSM images are corrupted with a

Poisson noise due to low illumination. Because of these intrinsic optical limitations, it is essential to remove both DV blur and noise from these images by digital processing.

In this context, we first study space-variant (SV) blur models and prove that a model where the SV point spread function (PSF) is approximated by a convex combination of a set of space-invariant (SI) PSFs is efficient and adequate to the inversion problem [30] [10]. Afterwards, we focus on the non-bind restoration problem and we fit a fast restoration method based on a domain decomposition technique [33] to our DV blur model [10], [9].

Recently, we focus on the blind case. In fact, in practice it is difficult to obtain the DV PSF in spite of the existence of theoretical PSF models [34], because these models are dependent on some unknown acquisition parameters (e.g. the refractive index (RI) of the specimen). Therefore a blind or semi-blind restoration algorithm is needed for this system. We propose two methods for this problem : In the first method, we define a criterion to be jointly minimized w.r.t to the image and the PSF set. In this method, the intensities of each SI PSF are estimated at every voxel. Although the big number of parameters to be estimated, the method allows more freedom on the shape of the PSF which could be more or less deformed according to spherical aberration level. We provide a theoretical proof of the existence of a minimizer of the considered problem [23]. Then, we perform the minimization by following an alternate minimization scheme, each elementary minimization is performed using the recently proposed scaled gradient projection (SGP) algorithm that has shown a fast convergence rate [29]. Results on simulated CLSM images and comparison with another alternate scheme based on a regularized version of the Richardson–Lucy algorithm [31] are shown in Fig. 3 . In the second blind method, we use a Gaussian approximation of each of the SI PSFs. This presents the advantage of significantly reducing the number of parameters to be estimated but constraints the PSF shape. We prove on simulated data that the first method provides more accurate restoration result than the second one.

| (a) | (b) | (c) | (d) |
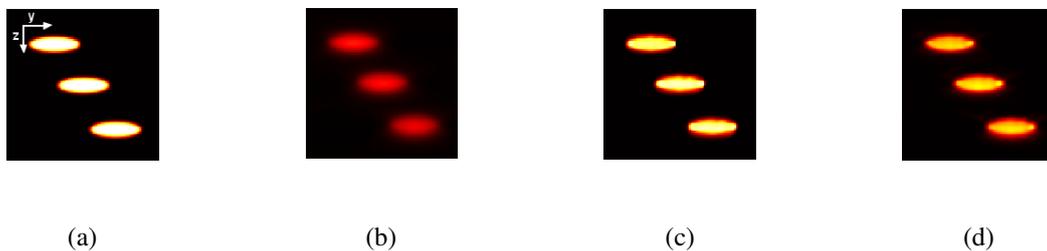
Figure 3. (Y, Z) sections of blind restoration results on a simulated CLSM image. (a) original image, (b) simulated observation, (c) restoration using our blind restoration method using SGP algorithm, (d) restoration using a regularized Richardson–Lucy algorithm embedded in an alternate scheme.

5.1.4. Morphogenesis of living organisms

Participant: Grégoire Malandain.

This research takes place within the Inria Large-scale initiative Morphogenetics.

This work was made in collaboration with Christophe Godin and Léo Guignard from Virtual Plants.

super-resolution, SPIM, morphogenesis

We extended a previous work [32] for the reconstruction of microscopic images. In particular, we extended the super-resolution image reconstruction (where several images, acquired from different viewpoints, are fused) to the lightsheet (or SPIM for Selective Plane Illumination Microscope) microscope modality. This modality offers a high acquisition speed, allowing imaging an organism frequently. As an exemple, Phallusia mammillata and Ciona intestinalis embryos can be imaged from 32 cells to around 1000 cells. The organism is captured from four different angles every 2 minutes during 2 hours (collaboration with CRBM Montpellier and EMBL Heidelberg).
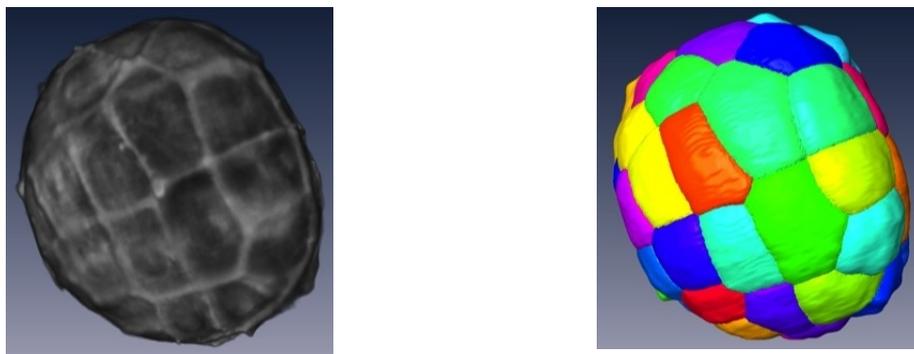
Figure 4. Left: 3D rendering of a reconstructed image of a Phallusia mammillata embryo; right: segmentation of cells.

5.2. Features Extraction

5.2.1. *Axon extraction from fluorescent confocal microscopy images*

Participants: Alejandro Mottini, Xavier Descombes, Florence Besse.

It it known that the analysis of axonal topologies allows biologists to study the causes of neurological diseases such as Fragile X Syndrome and Spinal Muscular Atrophy. In order to perform the morphological analysis of axons, it is first necessary to segment them. Therefore, the automatic extraction of axons is a key problem in the field of neuron axon analysis.

For this purpose, biologists label single neurons within intact adult Drosophila fly brains and acquire 3D fluorescent confocal microscopy images of their axonal trees. These images need to be segmented.

In our work presented in [16], we propose a new approach for the automatic extraction of axons from fluorescent confocal microscopy images which combines algorithms for filament enhancement, binarization, skeletonization and gap filling in a pipeline capable of extracting the axons containing a single labeled neuron. Unlike other segmentation methods found in the literature, the proposed is fully automatic and designed to work on 3D image stacks. This allows us to analyze large image databases.

The method performance was tested on 12 real 3D images and the results quantitatively evaluated by calculating the RMSE between the tracing done by an experienced biologist and the automatic tracing obtained by our method. The good results obtained in the validation show the potential use of this technique in helping biologists for extracting axonal trees from confocal microscope images (see figures 5 and 6).

Figure 5. Results obtained on each step of the algorithm for one image stack (maximum intensity projections). From left to right: original image, filament enhancement, binarization and final result.

5.2.2. *Dendrite spine detection from X-ray tomographic volumes*

Participants: Anny Hank, Xavier Descombes, Grégoire Malandain.
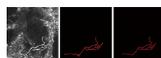
Figure 6. Comparison between original image (left), our result (middle) and ground truth (right) for two images (maximum intensity projections).

We have developed an automated algorithm for detecting dendritic spines from XRMT data. XRMT data allows imaging a large volume of tissue, and therefore a higher number of spines than laser scanning microscopy. We have shown that despite the lower image quality compared to microscopic data, we were able to extract dendritic spines. The main idea of the proposed approach is to define a mask for performing the spine detection without facing the false alarms problem as we introduce some information on spines localization. We therefore first extract the dendrites themselves and then compute the spine mask based on prior knowledge on their distance to dendrites. To extract dendrite we first compute the medial axis thanks to a multi-scale Hessian-based method. Then, we extract segments by a 3D Hough transform and reconstruct the dendrites using a conditional dilation. The spine mask is defined nerby the detected dendrites using anatomical parameters described in the literature. A point process defined on this mask provides the spine detection.

To exemplify the proposed approach, a subvolume ($220 \times 180 \times 100$) has been extracted from a XRMT volume that is given on figure 7 . As expected, the spines appear as small objects, whose size is close to the image resolution, along the tubular structures representing dendrites. Using the localization information to detect spine is essential to prevent false alarms due to noise or to the deviation of dendrites from a cylinder model. Figure 7 shows the detected dendrite medial axis and the obtained spine detection. The obtained results are promising and correspond to a visual inspection of the data. Forthcoming validation study will allow to better assess the quality of the detection by providing a quantitative evaluation.
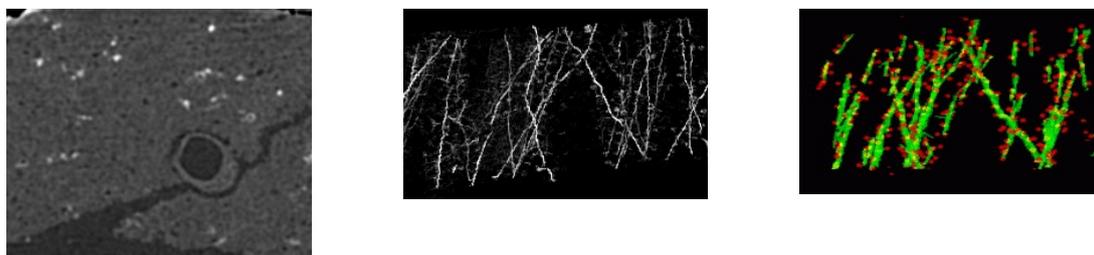
Figure 7. XMRT slice (left), dendrites medial axis (middle) and spine detection (in red) (right)

5.2.3. Cell detection

Participant: Xavier Descombes.

This work was done in collaboration with Emmanuel Soubies and Pierre Weiss from ITAV (Toulouse)

We have proposed some improvements of the Multiple Birth and Cut algorithm (MBC) in order to extract nuclei in 2D and 3D images. We have introduced a new contrast invariant energy that is robust to degradations encountered in fluorescence microscopy (e.g. local radiometry attenuations). Another contribution of this work is a fast algorithm to determine whether two ellipses (2D) or ellipsoids (3D) intersect. Finally, we propose a new heuristic that strongly improves the convergence rates. The algorithm alternates between two birth steps. The first one consists in generating objects uniformly at random and the second one consists in perturbing the

current configuration locally. Performance of this modified birth step is evaluated and examples on various image types show the wide applicability of the method in the field of bio-imaging.

Figure 8 left shows the segmentation result on a Drosophila embryo obtained using SPIM imaging. This is a rather easy case, since nuclei shapes vary little. The images are impaired by various defects: blur, stripes and attenuation. Despite this relatively poor image quality, the segmentation results are almost perfect. The computing time is 5 minutes using a C++ implementation. The image size is 700 × 350. Figure 8 right presents a more difficult case, where the image is highly deteriorated. Nuclei cannot be identified in the image center. Moreover, nuclei variability is important meaning that the state space size χ is large. Some nuclei are in mitosis (see e.g. top-left). In spite of these difficulties, the MBC algorithm provides acceptable results. They would allow to make statistics on the cell location and orientation, which is a major problem in biology. The computing times for this example is 30 minutes.
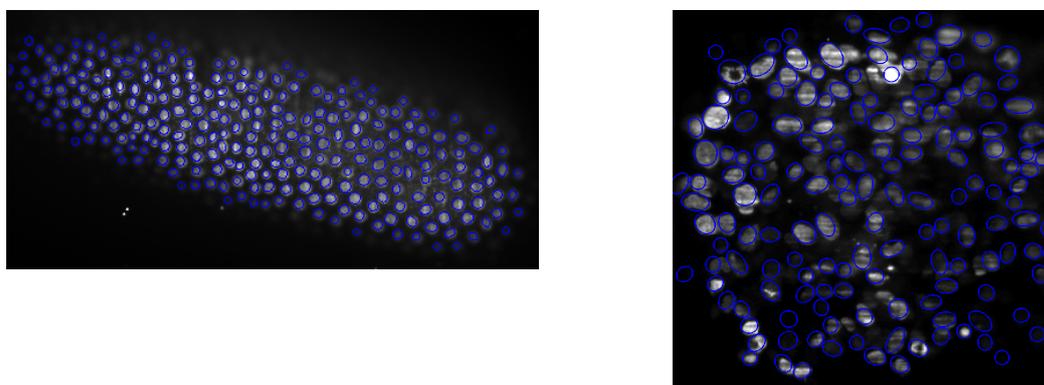
Figure 8. 2D segmentations of a nuclei of Drosophila embryo (left) and a multicellular tumor spheroid (right).

5.2.4. *Spermatozoid tracking*

Participants: Clarens Caraccio, Xavier Descombes.

In this work, we have proposed an algorithm for tracking spermatozoid in a sequence of confocal images. We first detect the spermatozoids by thresholding the result of a top hat operator. The thresold is automatically estimated using Otsu's method. We then analyse the different connected components to detect overlaps between adjacent spermatozoids. Temporal neighbors are selected based on the spatial consistency of the object sets between two consecutive time. A first result is given on figure 9 .

5.3. Classification

5.3.1. *Axon morphology comparison using elastic shape analysis*

Participants: Alejandro Mottini, Xavier Descombes, Florence Besse.

It is known that neuronal morphology impacts network connectivity, thus providing information on its functioning. Moreover, it allows the characterization of pathological states. Therefore, the analysis of the morphological differences between normal and pathological structures is of paramount importance.

We present a new method for comparing reconstructions of axonal trees (obtained, for example, by applying our segmentation method on confocal microscopy images of normal and mutant axonal trees) which takes into account both topological and geometrical information and is based on the Elastic Shape Analysis Framework. The method computes the geodesic between two axons in a space of tree like shapes, and the distance between the two is defined as the length of the geodesic. Moreover, our method is capable of showing how one axon transforms into the other by taking intermediate points in the geodesic.
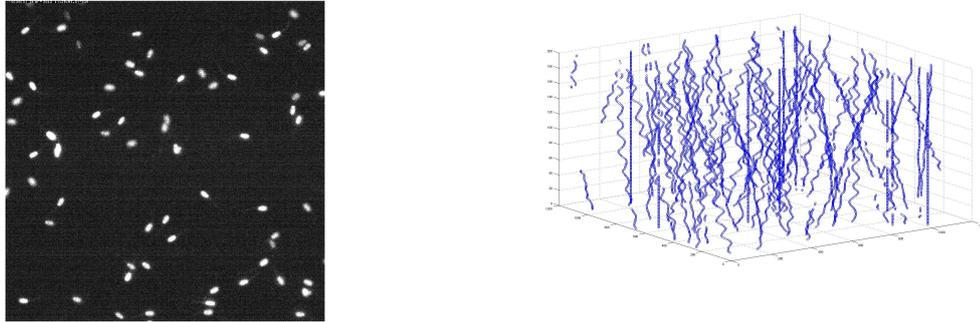
Figure 9. Original confocal image and estimated spermatozoid trajectories.

We consider two axonal trees T_1 and T_2, each consisting of an axon and several branches (and possibly sub branches). All are represented by 3D open curves in $\mathbb{R}^3$ (see Figure 10). We start by defining the matching function M such that $M : (0, 1, 2, ...n) \times (0, 1, ..., m)$, where n and m are the number of branches in T_1 and T_2 respectively. The matching function matches the branches of the two trees, for example, by assigning branch $i = 1$ of T_1 to branch $j = 3$ of T_2. We then define a branch function C which indicates, for a given time t_c, how many branches remain after $\beta(t_c)$ (see Figure 10). We only take into account branches which have a match in the other axonal tree. Finally, we define the distance between two axonal trees T_1, T_2 as:

$$D(T_1, T_2) = \min_M \, d((\beta_1(t), C1(t, M)), (\beta_2(t), C2(t, M))) + \sum_{(i,j)} \alpha_{i,j} M(i, j) D(T_1(i), T_2(j)) \qquad (3)$$

where β_k is the main curve (axon) of tree k, C_k its branch function, M the matching function, $\alpha_{i,j}$ a weight parameter and $D(T_1(i), T_2(j))$ the distance between the matched branches of the two trees. All distances between simple curves are calculated using the elastic shape analysis framework.

The method performance was tested on a group of 22 (11 normal and 11 mutant) 3D images, each containing one axonal tree manually segmented by an experienced biologist from a set of real confocal microscopy images. The mean and standard deviation of the inter and intra class distances between the neurons were calculated and results suggest that the proposed method is able to distinguish between the two populations (an average interpopulation to intrapopulation distance ratio of 1:21 and 1:28 were obtained). In addition, we computed the optimum transformations between axons. An example is shown in figure 11 . This result was obtained by taking intermediate points along the geodesic between the two trees.

Figure 10. Axonal tree diagrams (a) and their corresponding C functions for a given M (b).

5.3.2. *Vascular network segmentation from X-ray tomographic volumes*
Participant: Xavier Descombes.
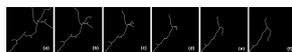
Figure 11. Optimum transformation between two axonal trees (transformation starts in (a) and finishes in (f), maximum intensity projections).

Micro-tomography produces high resolution images of biological structures such as vascular networks. We have defined 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 territory, defined as Local Vascular Territory (*a Local Vascular Territory (LVT) is a connected region corresponding to the catchment bassin associated with a vascular element. It can be obtained through the watershed computation on the opposite distance map from the vessels and is not connected to the sample border.*), is characterized by its volume and the local vascular density. The volume and density maps are first regularized by minimizing the total variation, within a Markov Random Field framework, using a graph cut algorithm . Then, a new approach is proposed to segment the volume from the two previous restored images using an iterative algorithm based on hypothesis testing. We consider the variables density and volume for each LVT and the populations constituted by the different classes obtained by the segmentation at a given step. Classes which are not statistically significantly different are merged using a MANOVA. This blind segmentation provides different regions which have been interpreted by expert as tumor, necrosis, tumor periphery and sane tissue 12 .

5.3.3. *Statistical analysis of skin pigmentation under treatment*

Participants: Sylvain Prigent, Xavier Descombes.

multispectral imaging, skin, hyperpigmentation, hypothesis tests, statistical inferences

One of the steps to evaluate the efficacy of a therapeutic solution is to test it on a clinical trial involving several populations of patients. Each population receives a studied treatment and a reference treatment for the disease. For facial hyper-pigmentation, a group of N_e patients receives the treatment on one cheek and a comparator on the other. The comparator can be a reference treatment or a placebo. 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 time t along the treatment period.

We propose a methodology to assess the efficacy a treatment by calculating three differential criteria: the darkness, the area and the homogeneity. The darkness measure the average intensity of the disease on a gray scaled image I obtained by a linear combination of the spectral bands of the original multi-spectral image. A differential darkness is then obtained by measuring the deviation between the initial measurement at time t_0, and the measurement at time t_k. The differential area criterion is calculated by analyzing the histogram of $I_{diff} = I_{t_0} - I_{t_k}$ a difference gray scale image between two measurements in a time series. The differential homogeneity criterion is obtaining with a multi-scale analysis of I_{diff} adapted from the Statistical Parametric Mapping (SPM) methodology. Indeed, statistical inferences allow to assign a probability of change to each region of I_{diff} above a set of thresholds. These probabilities are calculated with respect to the maximum intensity and the spatial extend of each region. An integration of the obtained statistical map denoted SM, allows to get a homogeneity criterion.
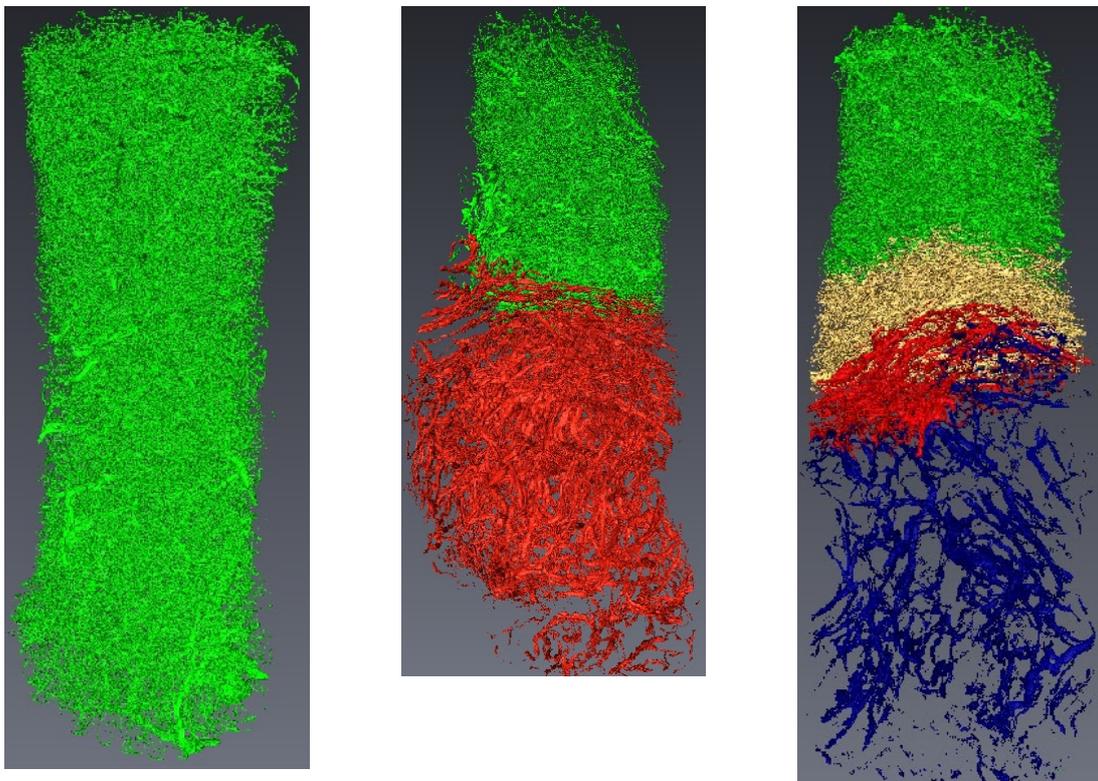
Figure 12. Examples of segmentation: tumor (red), necrosis (blue), tumor periphery (yellow) and sane (green)

The figure 13 illustrates the differential score calculated on a patient whose pathology decreases during the clinical trial. The proposed differential score have been tested in a full clinical study and provided results that agreed with the clinical analysis. This work have been patented and published in Inria research reports [25], [26].

(a) $I_{diff}^{t_1}$ equalised   (b) $I_{diff}^{t_2}$ equalised   (c) $I_{diff}^{t_3}$ equalised

(d) SM_{t_1} (e) SM_{t_2} (f) SM_{t_3}

(g) score t_1 (h) score t_2 (i) score t_3
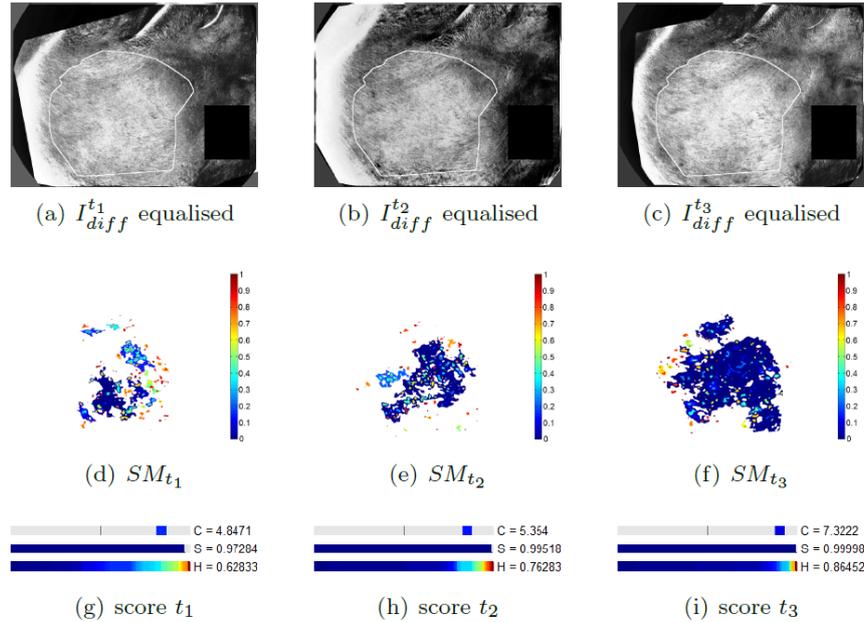
Figure 13. I_{diff}, SM and differential score for the three measurements t_1, t_2, t_3 calculated for a patient whose disease decreases.

5.3.4. A Recursive Approach For Multiclass Support Vector Machine: Application to automatic classification of endomicroscopic videos

Participants: Alexis Zubiolo, Eric Debreuve.

This work is made in collaboration with Barbara André (Mauna Kea Technologies)

The problem of automatic image (or video, or object) classification is to find a function that maps an image to a class or category among a number of predefined classes. An image can be viewed as a vector of high-dimension. In practice, it is preferable to deal with a synthetic signature of lower dimension. Therefore, the two classical steps of image classification are: image signature extraction and signature-based image classification. The classification rule can be learned from a set of training sample images manually classified by experts. This is known as supervised statistical learning where *statistical* refers to the use of samples and *supervised* refers to the sample classes being provided. We are interested in the learning aspect of the multiclass [1] problem when using a binary classification approach as a building block. We chose the Support Vector Machine (SVM), a well-known binary classifier.

Among the proposed extensions of binary classification methods to multiclass (three classes or more), the one-versus-one and one-versus-all approaches are the most popular ones. Let us suppose that there are $p \geq 3$ classes. The idea of the one-versus-all strategy is to oppose to any of the classes the union of the remaining $p - 1$ classes. Then, p SVM classifiers are determined, each one scoring, say, positively for one of the classes.

---

[1] Traditionally in classification, *multiclass* means "three classes or more" while the two-class case is referred to as binary classification.

The one-versus-one strategy opposes the classes by pair for all possible pairs. Therefore, $\frac{p(p-1)}{2}$ SVMs are determined and classification is performed by a majority vote.

As an alternative to these aforementioned strategies (as well as to other, less popular ones), we developed a recursive learning strategy. A tree of SVMs is built, achieving three goals: a fair balance in the number of samples used in each binary SVM learnings, a logarithmic complexity for classification ($\log_2 (p)$ compared to the linear or quadratic complexities of one-versus-all or one-versus-one, respectively), and a coherent, incremental classification procedure (as opposed to selecting the final class based on possibly competing partial decisions). During learning, at each node of the tree, a combinatorial search is performed to determine an optimal separation of the current classification problem into two sub-problems. The proposed method was applied to automatic classification of endomicroscopic videos.
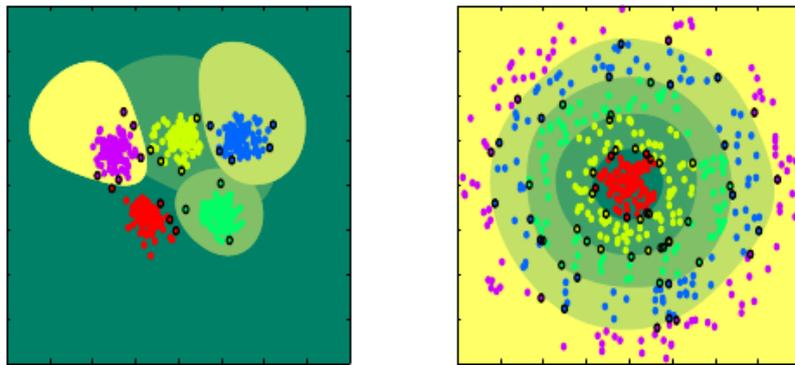
Figure 14. Illustration of the proposed recursive approach for multiclass Support Vector Machine. Colored dots: learning feature samples; Encircled dots: computed support vectors; Colored areas: computed class regions. Left: classical example; Right: concentric example.

5.4. Modeling

5.4.1. *Tracking Growing Axons in Fluorescent Microscopy Images*

Participants: Huei Fang Yang, Florence Besse, Xavier Descombes.

This work has been done in collaboration with Caroline Medioni from iBV.

Analyzing how growing axons correctly reach their target neurons is essential for biologists to better understand the development of a nervous system. Analysis of the properties of axon growth requires detecting axonal tips and tracking their trajectories within complex and large data sets. When performed manually, the tracking task is arduous and time-consuming. To this end, we proposed a tracking method, based on the particle filtering technique, to follow the traces of axonal tips that appear as small bright spots in the $3D + t$ fluorescent two-photon microscopy images exhibiting low signal-to-noise ratios (SNR) and complex background. Our tracking method uses multiple dynamic models in the proposal distribution to predict the positions of the growing axons. Moreover, it incorporates object appearance, motion characteristics of the growing axons, and filament information in the computation of the observation model. The integration of these three sources results in improved accuracy of recovered trajectories. The experimental results obtained from the microscopy images, presented in Figure 15 , showed that the proposed method can successfully estimate trajectories of growing axons, demonstrating its effectiveness even under the presence of noise and complex background.

5.4.2. *Trajectory Simulation of Growing Axons:*

Participants: Huei Fang Yang, Florence Besse, Xavier Descombes.

(a)   (b)   (c)
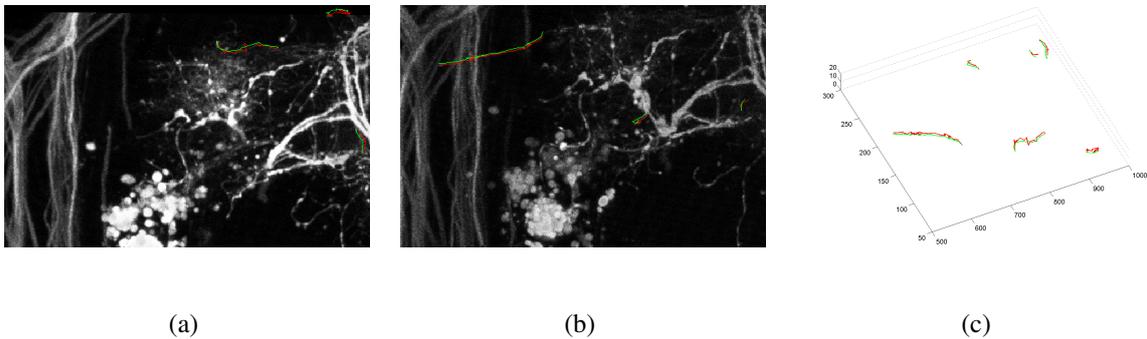
Figure 15. Visual comparison between the tracking results of the proposed method and the manually created ground truth in 2D and 3D. The red trajectories are produced by the proposed method, and the green are the ground truth manually created by the expert; both are overlaid on the MIPs (a) and (b) and visualized in 3D (c). The computer generated tracks are consistent with the ground truth in general, with minor differences between the estimated positions and the ground truth positions. The differences are caused by the noise and by the effect of complex background.

This work has been done in collaboration with Caroline Medioni from iBV.

It is established in biology that axons reach their target cells in the developing nervous system by the guidance of molecular gradients. To better understand how growing axons react to the molecular cues, either attractant or repellent, we simulated the trajectories of growing axons using a mathematical model that investigates the effect of molecular gradients on the axon's growth angle. Figure 16 shows the simulated trajectories of 50 growing axons. The initial position of axons is $(0, 0)$, and the red point on the right denotes the target cell that secretes the attractant cue.
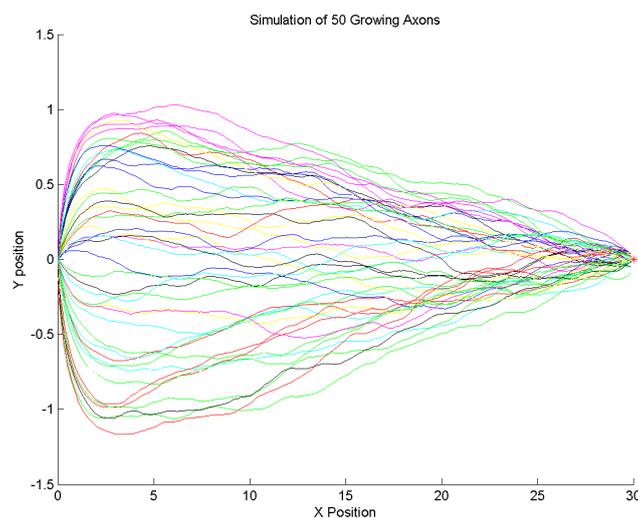
Figure 16. Simulated trajectories of 50 growing axons. The initial position of 50 axons is set to $(0, 0)$, and the red point on the right represents the target cell that secretes the attractant cue.

<center>**NACHOS Project-Team**</center>

6. New Results

6.1. Discontinuous Galerkin methods for Maxwell's equations

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

Participants: Loula Fezoui, 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-$\mathbb{P}_p$ 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. [52]) and have thus been designed with global conformity requirements (i.e. H_1, $H(\mathrm{rot})$ or (div)) whose role in the context of a discontinuous Galerkin formulation has to be clarified. This represents one of the objectives of this study.

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

Participants: Clément Durochat, Stéphane Lanteri, Raphael Léger, 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) discretization of the regular zones of the computational domain with an unstructured 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. With this objective in mind, we have designed and analyzed a DGTD-$\mathbb{P}_p\mathbb{Q}_k$ method formulated on non-conforming hybrid quadrangular/triangular meshes (2D case) or non-conforming hexahedral/tetrahedral meshes (3D case) for the solution of the time domain Maxwell's equations.

6.1.3. DGTD-$\mathbb{P}_p$ method for Debye media and applications to biolectromagnetics

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). The stability and a priori convergence analysis of the resulting DGTD-$\mathbb{P}_p$ method has been recently studied [25], and its application to the simulation of the propagation in realistic geometrical models of head tissues is underway.
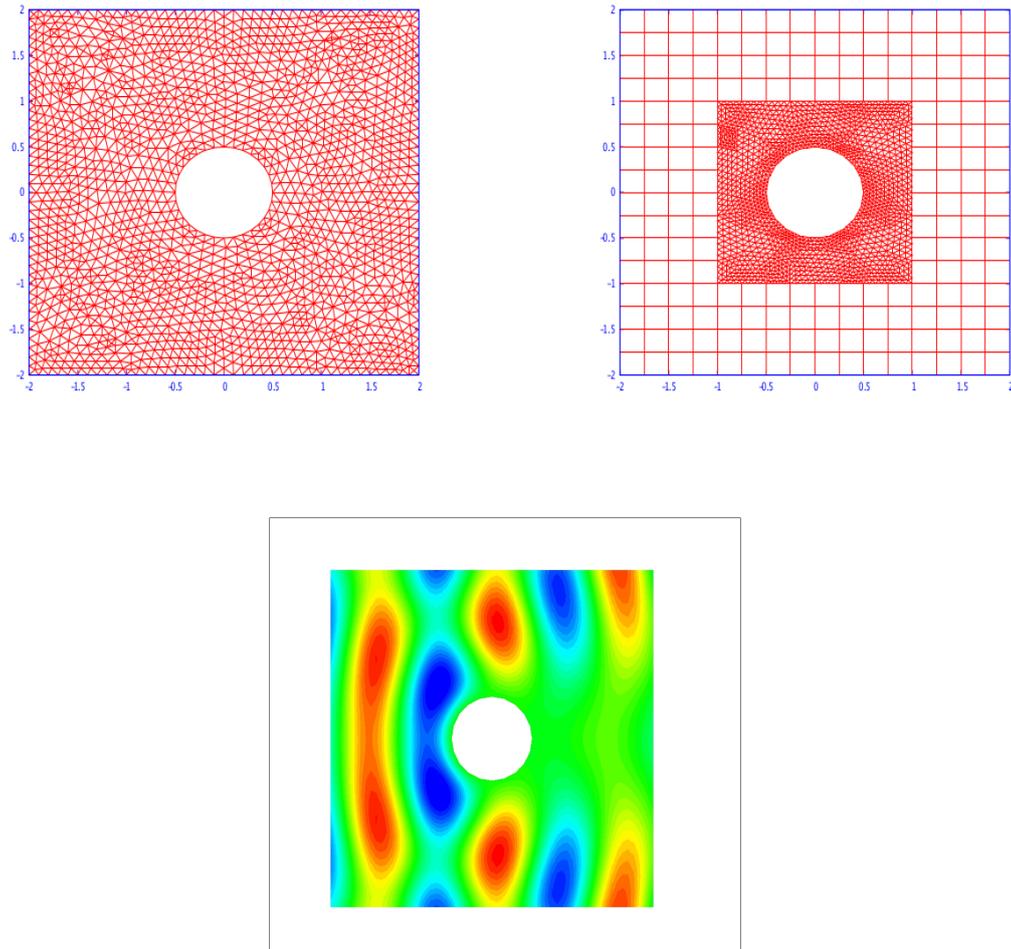
Figure 2. Scattering of a plane wave by an disk. Conforming triangular mesh (top left) and non-conforming quadrangular/triangular mesh (top right). Contour lines of electrical field component E_z from a simulation with a DGTD-$\mathbb{P}_2\mathbb{Q}_4$ method (bottom).

6.1.4. DGTD-$\mathbb{P}_p$ method for Drude media and applications to nanophotonics

Participants: Claire Scheid, Maciej Klemm [Communication Systems & Networks Laboratory, Centre for Communications Research, University of Bristol, UK], Stéphane Lanteri, Jonathan Viquerat.

Nanostructuring of materials has opened up a number of new possibilities for manipulating and enhancing light-matter interactions, thereby improving fundamental device properties. Low-dimensional semiconductors, like quantum dots, enable one to catch the electrons and control the electronic properties of a material, while photonic crystal structures allow to synthesize the electromagnetic properties. These technologies may, e.g., be employed to make smaller and better lasers, sources that generate only one photon at a time, for applications in quantum information technology, or miniature sensors with high sensitivity. The incorporation of metallic structures into the medium allows one to exploit plasmonic effects and adds further possibilities for manipulating the propagation of electromagnetic waves. In particular, this allows subwavelength localisation of the electromagnetic field and, by subwavelength structuring of the material, novel effects like negative refraction, e.g. enabling super lenses, may be realized. Nanophotonics is the recently emerged, but already well defined, field of science and technology aimed at establishing and using the peculiar properties of light and light-matter interaction in various nanostructures. Because of its numerous scientific and technological applications (e.g. in relation to telecommunication, energy production and biomedicine), nanophotonics represents an active field of research increasingly relying on numerical modeling beside experimental studies. We have started this year a new research direction aiming at the numerical modeling of electromagnetic wave interaction with nanoscale metallic structures. In this context, one has to take into account the dispersive characteristics of cartain metals in the frequency range of interest to nanophotonics. As a first step in this direction, we have considered an auxiliary differential equation approach for the numerical treatment of a Drude dispersion model in the framework of a DGFD-$\mathbb{P}_p$ method.

6.1.5. Frequency domain hybridized DGFD-$\mathbb{P}_p$ methods

Participants: Stéphane Lanteri, Liang Li [Faculty Member, School of Mathematical Sciences, Institute of Computational Science, University of Electronic Science and Technology of China Chengdu, China], 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-$\mathbb{P}_p$) method formulated on unstructured meshes for solving the 2D and 3D time harmonic Maxwell equations. However, 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.* [47] 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. In particular, we have recently designed a HDGFD-$\mathbb{P}_p$ method for the solution of the 2D frequency domain Maxwell equations [22] and, based on the very promising results obtained in this study, the extension to the more challenging 3D case has been initiated.

6.1.6. Exact transparent condition in a DGFD-$\mathbb{P}_p$ 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 [26].

6.2. Discontinuous Galerkin methods for the elastodynamic equations

6.2.1. DGTD-$\mathbb{P}_p$ method for viscoelastic media

Participants: Nathalie Glinsky, Stéphane Lanteri, 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-$\mathbb{P}_p$ 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. The resulting DGTD-$\mathbb{P}_p$ methods have been validated and evaluated in detail in the context of propagation problems in both homogeneous and heterogeneous media including problems for which analytical solutions can be computed. Particular attention was given to the study of the mathematical properties of these schemes such as stability, convergence and numerical dispersion.

A recent novel contribution is the extension of the DGTD method to include viscoelastic attenuation. For this, the velocity-stress first-order hyperbolic system is completed by additional equations for the anelastic functions including 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, we need solving 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 by comparison to results obtained by a finite-difference method, in particular for wave propagation in a realistic basin of the area of Nice (south of France) [40]-[35].

6.2.2. DGTD-$\mathbb{P}_p$ method for the assessment of topographic effects

Participants: Etienne Bertrand [CETE Méditerranée], Nathalie Glinsky.

This study addresses 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 occured 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 [38].

6.2.3. DGTD-$\mathbb{P}_p$ method for arbitrary heterogeneous media

Participants: Nathalie Glinsky, Diego Mercerat [CETE Méditerranée].

We have recently devised an extension of the DGTD method for elastic wave propagation in arbitrary heterogeneous media. In realistic geological media (sedimentary basins for example), one has to include strong variations in the material properties. Then, the classical hypothesis that these properties are constant within each element of the mesh can be a severe limitation of the method, since we need to discretize the medium with very fine meshes resulting in very small time steps. For these reasons, we propose an improvement of the DGTD method allowing non-constant material properties within the mesh elements. A change of variables on the stress components allows writing the elastodynamic system in a pseudo-conservative form. Then, the introduction of non-constant material properties inside an element is simply treated by the calculation, via convenient quadrature formulae, of a modified local mass matrix depending on these properties. This new extension has been validated for a smoothly varying medium or a strong jump between two media, which can be accurately approximated by the method, independently of the mesh [39].

6.2.4. DGFD-$\mathbb{P}_p$ method for frequency domain elastodynamics

Participants: Hélène Barucq [MAGIQUE3D project-team, Inria Bordeaux - Sud-Ouest], Marie Bonnasse, Julien Diaz [MAGIQUE3D project-team, Inria Bordeaux - Sud-Ouest], Stéphane Lanteri.

We have started this year a research direction aiming at the development of high order discontinuous Galerkin methods on unstructured meshes for the simulation of frequency domain elastodynamic and viscelastic wave propagation. This study is part of the Depth Imaging Partnership (DIP) between Inria and TOTAL. The PhD thesis of Marie Bonnasse is at the heart of this study which is funded by TOTAL.

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 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. In [6], we conducted a preliminary numerical study of a hyrbid explicit-implicit DGTD-$\mathbb{P}_p$ method, combining a leap-frog scheme and a Crank-Nicolson scheme, and obtained promising results. More recently, we further investigated two such strategies, both theoretically (especially, convergence in the ODE and PDE senses) [24] and numerically in the 2D case [23]. A last topic is to propose higher order time integration techniques based on the second-order locally implicit method to fully exploit the attractive features of this approach combined with a DG discretisation which allows to easily increase the spatial convergence order. Promising results in 2D reaching high order in time, between $3, 5$ and 4, have been obtained in [33] by applying Richardson extrapolation and composition methods.
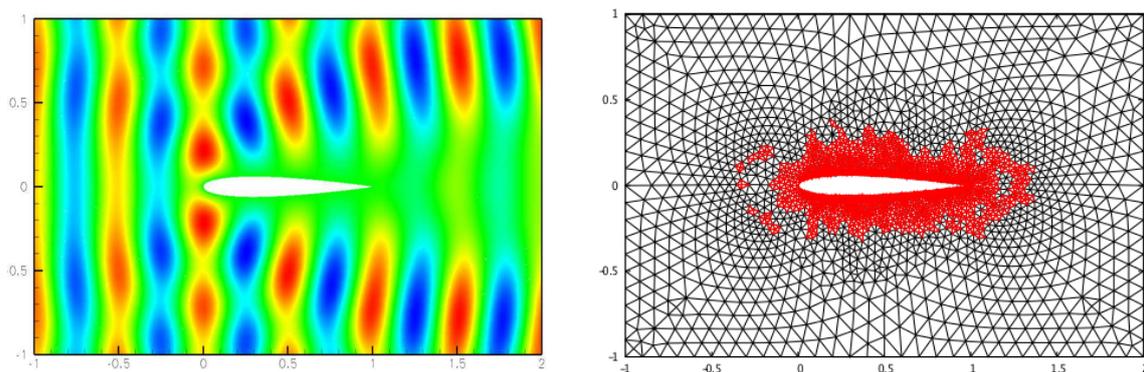
Figure 3. Scattering of a plane wave by an airfoil profile. Contour lines of electrical field component E_z (left) and locally refined triangular mesh with partitioning in explicit/implicit zones (right).

6.3.2. *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].

Even if they have been introduced for the first time two centuries ago, over the last two decades, classical Schwarz methods have regained a lot of popularity with the developement of parallel computers. First developed for the elliptic problems, they have been recently extended to systems of hyperbolic partial differential equations, and it was observed that the classical Schwartz method can be convergent even without overlap in certain cases. This is in strong contrast to the behavior of classical Schwarz methods applied to elliptic problems, for which overlap is essential for convergence. Over the last decade, optimized versions of Schwarz methods have been developed for elliptic partial differential equations. These methods use more effective transmission conditions between subdomains, and are also convergent without overlap for elliptic problems. The extension of such methods to systems of equations and more precisely to Maxwell's system (time harmonic and time discretized equations) has been studied in [8]. The optimized interface conditions proposed in [8] were devised for the case of non-conducting propagation media. We have recently studied the formulation of such conditions for conducting media [17]. Besides, we have also proposed an appropriate discretization strategy of these optimized Schwarz algorithms in the context of a high order DGFD-$\mathbb{P}_p$ method formulated on unstructured triangular meshes for the solution of the 2D frequency domain Maxwell equations [28].
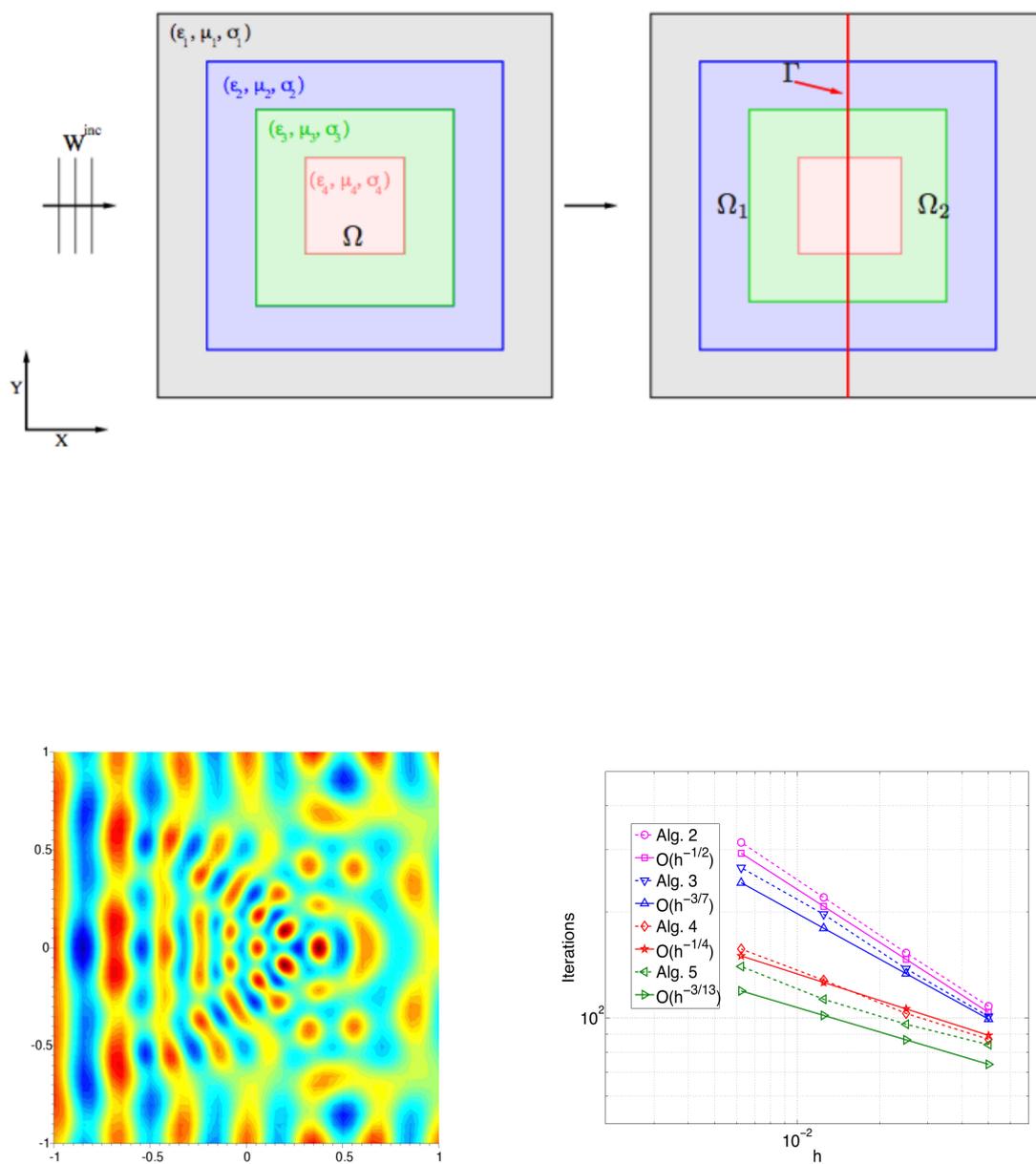
Figure 4. Propagation of a plane wave in a multilayered heterogeneous medium. Problem setting and two-subdomain decompositin (top). Contour lines of the real part of the E_z component of the electrical field (bottom left) and asymptotic convergence of the optimized Schwarz algorithms (bottom right).

NEUROMATHCOMP Project-Team

5. New Results

5.1. Neural Networks as dynamical systems

5.1.1. Dynamics and spike trains statistics in conductance-based Integrate-and-Fire neural networks with chemical and electric synapses

Participants: Rodrigo Cofré, Bruno Cessac [correspondent].

We investigate the effect of electric synapses (gap junctions) on collective neuronal dynamics and spike statistics in a conductance-based Integrate-and-Fire neural network, driven by a Brownian noise, where conductances depend upon spike history. We compute explicitly the time evolution operator and show that, given the spike-history of the network and the membrane potentials at a given time, the further dynamical evolution can be written in a closed form. We show that spike train statistics is described by a Gibbs distribution whose potential can be approximated with an explicit formula, when the noise is weak. This potential form encompasses existing models for spike trains statistics analysis such as maximum entropy models or Generalized Linear Models (GLM). We also discuss the different types of correlations: those induced by a shared stimulus and those induced by neurons interactions. This work has been presented in several conferences [43], [45], [46], [47], [31], [48] and submitted to Chaos, Solitons and Fractals [13].

5.1.2. Parameter estimation in spiking neural networks: a reverse-engineering approach

Participants: Horacio Rostro-Gonzalez [Holistic Electronics Research Lab, University of Cyprus], Bruno Cessac [correspondent], Thierry Viéville [Inria Mnemosyne].

This work presents a reverse engineering approach for parameter estimation in spiking neural networks (SNNs). We consider the deterministic evolution of a time-discretized network with spiking neurons, where synaptic transmission has delays, modeled as a neural network of the generalized integrate and fire type. Our approach aims at by-passing the fact that the parameter estimation in SNN results in a non-deterministic polynomial-time hard problem when delays are to be considered. Here, this assumption has been reformulated as a linear programming (LP) problem in order to perform the solution in a polynomial time. Besides, the LP problem formulation makes explicit the fact that the reverse engineering of a neural network can be performed from the observation of the spike times. Furthermore, we point out how the LP adjustment mechanism is local to each neuron and has the same structure as a 'Hebbian' rule. Finally, we present a generalization of this approach to the design of input–output (I/O) transformations as 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. Numerical verifications and illustrations are provided. This work has been published in Journal of Neural Engineering [24].

5.2. Mean field approaches

5.2.1. Noise-induced behaviors in neural mean field dynamics

Participants: Olivier Faugeras [correspondent], Geoffroy Hermann, Jonathan Touboul [Inria Bang].

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 published in the SIAM Journal on Applied dynamical Systems [25].

5.2.2. *Mean-field description and propagation of chaos in networks of Hodgkin-Huxley neurons*

Participants: Javier Baladron, Diego Fasoli, Olivier Faugeras [correspondent], Jonathan Touboul [Inria Bang].

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 place, 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 a 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 the McKean-Vlasov equations, i.e. the existence and uniqueness of a solution. We also show the results of some numerical experiments that indicate that the mean-field equations are a good representation of the mean activity of a finite size network, even for modest sizes. These experiments 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 appeared in the Journal of Mathematical Neuroscience [11].

5.3. Neural fields theory

5.3.1. *Localized radial bumps of a neural field equation on the Euclidean plane and the Poincaré disk*

Participants: Grégory Faye [correspondent], James Rankin, David, J.B. Lloyd [Department of Mathematics at the University of Surrey].

We analyze radially symmetric localized bump solutions of an integro-differential neural field equation posed in Euclidean and hyperbolic geometry. The connectivity function and the nonlinear firing rate function are chosen such that radial spatial dynamics can be considered. Using integral transforms, we derive a PDE for the neural field equation in both geometries and then prove the existence of small amplitude radially symmetric

spots bifurcating from the trivial state. Numerical continuation is then used to path-follow the spots and their bifurcations away from onset in parameter space. It is found that the radial bumps in Euclidean geometry are linearly stable in a larger parameter region than bumps in the hyperbolic geometry. We also find and path-follow localized structures that bifurcate from branches of radially symmetric solutions with D6-symmetry and D8-symmetry in the Euclidean and hyperbolic cases, respectively. Finally, we discuss the applications of our results in the context of neural field models of short term memory and edges and textures selectivity in a hypercolumn of the visual cortex. This work has been accepted for publication in Nonlinearity [57].

5.3.2. *Center manifold for delayed neural fields equations*

Participants: Olivier Faugeras [correspondent], Romain Veltz [Salk Institute,San Diego,USA].

We develop a framework for the study of delayed neural fields equations and prove a center manifold theorem for these equations. Specific properties of delayed neural fields equations make it impossible to apply existing methods from the literature concerning center manifold results for functional differential equations. Our approach for the proof of the center manifold theorem uses the original combination of results from Vanderbauwhede and colleagues together with a theory of linear functional differential equations in a history space larger than the commonly used set of time-continuous functions. This work has been submitted to the SIAM Journal on Applied Mathematics and is under revision [27].

5.3.3. *Reduction method for localized solutions*

Participant: Grégory Faye [correspondent].

We present a reduction method to study localized solutions of an integrodifferential equation defined on the Poincaré disk. This equation arises in a problem of texture perception modeling in the visual cortex. We first derive a partial differential equation which is equivalent to the initial integrodifferential equation and then deduce that localized solutions which are radially symmetric satisfy a fourth order ordinary differential equation. This work has appeared in the "Comptes Rendus Mathematique" [15].

5.3.4. *Spatially localized solutions*

Participants: Pascal Chossat, Grégory Faye [School of Mathematics,University of Minnesota, correspondent], James Rankin.

The existence of spatially localized solutions in neural networks is an important topic in neuroscience as these solutions are considered to characterize working (short-term) memory. We work with an unbounded neural network represented by the neural field equation with smooth firing rate function and a wizard hat spatial connectivity. Noting that stationary solutions of our neural field equation are equivalent to homoclinic orbits in a related fourth order ordinary differential equation, we apply normal form theory for a reversible Hopf bifurcation to prove the existence of localized solutions; further, we present results concerning their stability. Numerical continuation is used to compute branches of localized solution that exhibit snaking-type behaviour. We describe in terms of three parameters the exact regions for which localized solutions persist. This work has appeared in the Journal of Mathematical Biology [16].

5.3.5. *Bumps in the Poincaré disk*

Participants: Grégory Faye [School of Mathematics,University of Minnesota, correspondent], David, J.B. Loyd, James Rankin.

We analyze radially symmetric localized bump solutions of an integro-differential neural field equation posed in Euclidean and hyperbolic geometry. The connectivity function and the nonlinear firing rate function are chosen such that radial spatial dynamics can be applied. Using integral transforms, we derive a PDE of the neural field equation in both geometries and then prove the existence of small amplitude radially symmetric spots bifurcating from the trivial state. Numerical continuation is then used to path-follow the spots and their bifurcations away from onset in parameter space. It is found that the radial bumps in Euclidean geometry are linearly stable in a larger parameter region than bumps in the hyperbolic geometry. We also find and path follow localized structures that bifurcate from branches of radially symmetric solutions with D6-symmetry and D8-symmetry in the Euclidean and hyperbolic cases, respectively. Finally, we discuss the applications of our results in the context of neural field models of short term memory and edges and textures selectivity in a hypercolumn of the visual cortex. This work has been submitted to Nonlinearity.

5.4. Spike trains statistics

5.4.1. Natural image identification from spike train analysis

Participants: Geoffrey Portelli, Olivier Marre [Institution de la Vision, Paris, France], Marc Antonini [Laboratoire I3S, UMR CNRS, Universite´ Nice Sophia Antipolis, France], Michael Berry II [Princeton Neuroscience Institute, Department of Molecular Biology, Princeton University, Princeton, NJ 08544, USA], Pierre Kornprobst [correspondent].

We started a new activity to analyse how natural images are encoded in retinal output. This work is related to [67], [72] where synthetic stimuli are used. Here, we recorded a population of 100-200 ganglion cells of a salamander retina, while flashing 720 natural images from the Torralba database [71] plus one control image, each repeated 10 times. We characterized the response of each cell by two parameters : the latency of the first spike after the stimulus onset, and the firing rate. A distribution of these two features was then estimated for each neuron and natural image. Pooling the information across all the neurons, a discriminability coefficient between pairs of image is proposed, using either the rate or the latency, or both. We also provide a way to identify a given image among others based on the rate–latency distributions. Preliminary results have been presented in [40]. Results showed that, on average, the discriminability was better based on the latency than on the rate. The most discriminable pairs were different using the rate or the latency, so these two features conveyed complementary information. In addition, we observe a similar evolution of the identification performance when the rate, or the latency, or both are used.

5.4.2. Spike train statistics from empirical facts to theory: the case of the retina

Participants: Bruno Cessac [correspondent], Adrian Palacios [CINV-Centro Interdisciplinario de Neurociencia de Valparaiso, Universidad de Valparaiso].

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 has been published in Mathematical Problems in Computational Biology and Biomedicine, F. Cazals and P. Kornprobst, Springer [55].

5.4.3. Gibbs distribution analysis of temporal correlations structure in retina ganglion cells

Participants: Juan-Carlos Vasquez, Olivier Marre [Institution de la Vision, Paris, France], Adrian Palacios [CINV-Centro Interdisciplinario de Neurociencia de Valparaiso, Universidad de Valparaiso], Michael Berry II [Princeton Neuroscience Institute, Department of Molecular Biology, Princeton University, Princeton, NJ 08544, USA], Bruno Cessac [correspondent].

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 presented in several conferences [29], [30], [28] and pusblished in J. Physiol. Paris [26].

5.4.4. Spatio-temporal spike trains analysis for large scale networks using maximum entropy principle and Monte-Carlo method

Participants: Hassan Nasser, Olivier Marre [Institut de la Vision, Paris, France], Bruno Cessac [correspondent].

Understanding the dynamics of neural networks is a major challenge in experimental neuroscience. For that purpose, a modelling of the recorded activity that reproduces the main statistics of the data is required. We present a review on recent results dealing with spike train statistics analysis using maximum entropy models (MaxEnt). Most of these studies have been focusing on modelling synchronous spike patterns, leaving aside the temporal dynamics of the neural activity. However, the maximum entropy principle can be generalized to the temporal case, leading to Markovian models where memory effects and time correlations in the dynamics are properly taken into account. We also present a new method based on Monte-Carlo sampling which is suited for the fitting of large-scale spatio-temporal MaxEnt models. The formalism and the tools presented here will be essential to fit MaxEnt spatio-temporal models to large neural ensembles. This work has been presented in several conferences [54], [51], [53], [52] and accepted in Journal of Statistical Mechanics [22].

5.4.5. *Spike train statistics and Gibbs distributions*

Participants: Rodrigo Cofré, Bruno Cessac [correspondent].

We introduce Gibbs distribution in a general setting, including non stationary dynamics, and present then three examples of such Gibbs distributions, in the context of neural networks spike train statistics: (i) Maximum entropy model with spatio-temporal constraints; (ii) Generalized Linear Models; (iii) Conductance based Inte- grate and Fire model with chemical synapses and gap junctions. This leads us to argue that Gibbs distributions might be canonical models for spike train statistics analysis. This work has been presented in several conferences [43], [31] and submitted to J. Physiol. Paris [12].

5.5. Visual Neuroscience

5.5.1. *Neural fields models for motion integration: Characterising the dynamics of multi-stable visual motion stimuli*

Participants: Olivier Faugeras, Pierre Kornprobst, Guillaume S. 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.

We are investigating 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 study the temporal dynamics of this phenomenon with a neural fields, population-level representation of activity in MT, a cortical area dedicated to motion estimation. Numerical tools from bifurcation analysis are used to investigate the model's behaviour in the presence of different types of input; this general approach could be applied to a range of neural fields models that are typically studied only in terms of their spontaneous activity. The model reproduces known multistable behaviour in terms of the predominant interpretations (percepts) of the barber pole stimulus.

We probe the early processing from stimulus presentation to initial perceived direction (before perceptual reversals). The basic dynamic properties of the early transition from D to H/V are well predicted by the model. This work has been presented in the European Conference on Vision Perception (ECVP) [38], [41] and it has been published in [23].

We are extending this work to investigate the longer term dynamics for which perceptual reversals are known to occur, due to competition between 1D motion cues aligned with the grating's motion direction and 2D motion cues aligned with aperture edges. This work has been presented in the Vision Sciences Society 12th Annual Meeting (VSS) [39], [42].

5.6. Neuromorphic Vision

Participants: Khaled Masmoudi [Laboratoire I3S, UMR CNRS, Université Nice Sophia Antipolis, France], Marc Antonini [Laboratoire I3S, UMR CNRS, Université Nice Sophia Antipolis, France], Pierre Kornprobst.

In the scope of Khaled Masmoudi's PhD [9], we have developed bio-inspired schemes for image coding. This is a new area of research on which very few teams are committed. We have proposed schemes for encoding/decoding images directly using the functional architecture of the retina and the properties of its spiking output (e.g., using Laplacian pyramids model and the rank-order coding [69], [68][20], [21] and the Virtual Retina simulator [37] from Adrien Wohrer during his PhD [79], [78]).

<h1 style="text-align:center; color:red">OASIS Project-Team</h1>

6. New Results

6.1. Programming and Composition Models for Large-Scale Distributed Computing

6.1.1. *Multi-active Objects*
Participants: L. Henrio, F. Huet, A. Bourdin.

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 to extend 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 we realized the following:

- improvement on the model and its formalisation
- use of the new model in our CAN P2P network (see below); this was also the opportunity to improve our implementation.

This year, we also spent considerable efforts to publish this work; a conference paper is currently under review.

6.1.2. *Events for Algorithmic skeletons*
Participant: L. Henrio.

In the context of the SCADA associated team, we worked on the algorithmic skeleton programming model. The structured parallelism approach (skeletons) takes advantage of common patterns used in parallel and distributed applications. The skeleton paradigm separates concerns: the distribution aspect can be considered separately from the functional aspect of an application.

- This year we focused on the handling of events in algorithmic skeletons: adding the possibility for a skeleton to output an event should increase the control and monitoring capabilities of algorithmic skeletons. The ultimate goal is to improve autonomicity for algorithmic skeletons.

6.1.3. *Behavioural models for Distributed Components*
Participants: E. Madelaine, N. Gaspar, A. Savu, L. Henrio.

In the past [3], we defined the behavioural semantics of active objects and components. This year we extended this work to address 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 the CADP model-checker, that also supports distributed generation of state-space. This year our main achievements are the following:

- We entirely formalised the specification of the behavioural model generation for component systems. This should provide us both a stronger formal background for our works in this area, and a specification for the automatic generation of behavioural models for our component systems.
- We additionally have put considerable efforts on the improvement of the Vercors platform and its integration with the Papyrus framework (see Section 5.2).

The formal work has been published as a research report [40]. A journal version is under submission. This work was done in collaboration with Rabéa Ameur-Boulifa from Télécom-Paristech.

In parallel with core developments of the behavioural specification environment, our collaborations led us to the study of the following application domain. In the context of the Spinnaker project, we are interested in developing a component-based distributed application to manage and monitor some pre-existing component-based distributed application - and hence, we called it The HyperManager. Our in-house component model (GCM) provides all the means to define, compose and dynamically reconfigure such applications. However, special care must be taken for this kind of undertaking. To this end, this year:

- We made the first steps towards a platform for the mechanized specification and verification, in the Coq Proof Assistant, of GCM applications. This work was published in [33], and is progressively being updated [1] to cope with behavioural specification, and to seamlessly combine deductive and model-checking techniques.

- We studied a real-life application scenario for our HyperManager prototype using distributed model-checking techniques in order to cope with the huge space state generated from reconfigurable applications.

6.1.4. *Autonomic Monitoring and Management of Components*

Participants: F. Baude, C. Ruz, B. Sauvan.

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. For this purpose, we finalized the implementation of a factory which, from any GCM ADL description can instantiate the requested non functional components of a GCM application.

The framework for autonomic computing 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. We have demonstrated how such a control loop can be relevant to drive the dynamic reconfiguration of services part of a SOA application, considering as in the SCA standard, that services are components [15].

Our objective now is to exemplify such autonomic and structured approach in the management of any distributed middleware or application, e.g. in the Spinnaker industrial context.

6.1.5. *Optimization of data transfer in SOA and EDA models*

Participants: I. Alshabani, F. Baude, L. Pellegrino, B. Sauvan, Q. Zagarese.

Traditional client-server interactions rely upon method invocations with copy of the parameters. This can be useless in particular if the receiver does not effectively uses them. On the contrary, copying and transferring parameters lazily so to allow the receiver to proceed even without all of them is a meaningful idea that we proved to be effective for active objects in the past [56]. This idea wasn't so far realized in the context of the web services technology, the most popular one used today for client-server SOAP-based interactions.

- To such an aim, we contributed to the offloading of objects representing parameters of the web service Java Apache CXF API [29]. It is innovative notably in the way the offloading of parameters for on-demand access can be delegated from services to services, which resembles the concept of first-class futures.

- Relying upon such an effective approach, we have applied a similar idea of "lazy copying and transfer" to the data parts of events in the context of event-driven architecture applications [28]. The middleware dynamically off-loads data (generally of huge size) attached to an event, according to some user-level policy expressed as annotation in the Java code at the subscriber side. The event itself, without its attachments, gets forwarded into the publish/subscribe brokering system (in our case, the event cloud middleware, that is the subject of section 6.2.1) and its attachements are transferred to the subscriber only on-demand. Compared to some existing propositions geared towards a data centric publish-subscribe pattern (e.g. the DDS OMG standard), ours is more user-friendly as it does not require the user code to explicitly program when to get the data attached to notified events.

---

[1] http://www-sop.inria.fr/members/Nuno.Gaspar/Mefresa.php

Overall, this work opens the way towards a strong convergence between service oriented and event-driven technologies.

6.1.6. *Multi-layer component architectures*

Participant: O. Dalle.

Since a few years, we have been investigating the decomposition of a simulation application into multiple layers corresponding to the various concerns commonly found in a simulation: in addition to the various modeling domains that may be found in a single simulation application (e.g. telecommunications networks, road-networks, power-grids, and so on), a typical simulation includes various orthogonal concerns such as system modelling, simulation scenario, instrumentation and observation, distribution, and so on. This large number of concerns has put in light some limits of the traditional hierarchical component-based architectures and their associated ADL, as found in the FCM and GCM. In order order to cope with these limitations, we started a new component architecture model called Binding Layers centered on the binding rather than the component, with no hierarchy but advanced layering capabilities, and offering advanced support for dynamic structures[32].

6.2. Middleware for Grid and Cloud computing

6.2.1. *Publish-Subscribe in Distributed Environments*

Participants: F. Baude, F. Huet, F. Bongiovanni, L. Pellegrino, B. Sauvan, I. Alshabani, A. Bourdin, M. Antoine, A. Alshabani.

In the context of the SOA4ALL FP7-IP project, we designed and implemented a hierarchical Semantic Space infrastructure based on Structured Overlay Networks (SONS) [62], [63]. It originally aimed at the storage and the retrieval of the semantic description of services at the Web scale [57]. This infrastructure combines the strengths of both the P2P paradigm at the architectural level and the Resource Description Framework (RDF) data model at the knowledge representation level. The achievements of this year are the following:

- In the context of the FP7 Strep PLAY and French ANR SocEDA research projects, we have been extending the aforementioned work with a content-based Publish/Subscribe abstraction in order to support asynchronous queries for RDF-based events in large scale settings, which raises some interesting challenges [26]. The goal is to build a platform for large scale distributed reasoning[25]. Such an integrated working platform [39], [38] has been presented in two tutorials [27], [54].
- We have also investigated the Publish/Subscribe paradigm in the MapReduce programming model. We have proposed the concept of continuous job which allows MapReduce jobs to be re-executed when new data are added to the system. To maintain the correctness of the execution, we have introduced the notion of carried data, i.e. data which are kept between subsequent executions. An implementation has been written on top of Hadoop and a paper submitted.

6.2.2. *Distributed algorithms for CAN-like P2P networks*

Participants: L. Henrio, F. Bongiovanni, F. Huet.

The nature of some large-scale applications, such as content delivery systems or publish/subscribe systems, built on top of SONs, demands application-level dissemination primitives which do not overwhelm the overlay, i.e. efficient, 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. Our objective is to design and prove an efficient and reliable broadcast algorithm for CAN-like P2P networks. To this aim, in 2012 we:

- Improved the formalisation in Isabelle/HOL of a CAN-like P2P system, devised formalised tools to reason on CAN topologies, and on communication protocols on top of CANs. We designed and proved the efficiency of a first naive algorithm.
- Sketched on paper the proof of completeness and efficiency for the algorithm we designed and implemented last year.

Part of this work was done in the PhD thesis of F. Bongiovanni [10]

We are also investigating the new algorithms to efficiently build a SONs in the presence of existing data. Most of the work on SONs assume that new peers joining the network will arrive without data or fail to take into account the cost of distributing these data. Indeed, depending on the key subspace given to the new peer, some or all its data will have to be distributed in the network. In 2012:

- We proposed a first version of new join algorithms which try to allocate key sub-spaces to peers so that the amount of data that needs to be moved is minimal. An expected benefit of this work is that it should allow for fast and efficient reconstruction of a SON in case of a crash, without having to use distributed snapshots.

6.2.3. *Network Aware Cloud Computing*
Participants: S. Malik, F. Huet.

We have worked on the Resource Aware Cloud Computing project. Its primary purpose is to address different issues which can help the scheduler to make more efficient scheduling decisions. These issues are related to the resource characteristics.

- We introduce a framework, which increases the performance of the application and ensures high level of reliability during the scheduling of application onto the cloud. It is a cloud scheduler module named as Resource Aware Cloud Scheduling (RACS) module. It helps the scheduler in making the scheduling decisions on the basis of different characteristics of cloud resources. These characteristics are reliability, network latency, and monetary cost. RACS consists of multiple sub modules, which are responsible for their corresponding tasks. In RACS, we have done the implementation for the different issues.
- We worked on a model for the reliability assessment of the cloud's computing nodes. This reliability assessment mechanism helps to do the scheduling on cloud infrastructure and perform fault tolerance on the basis of the reliability values acquired during reliability assessment. The model has different algorithms for different types of applications. Thus it has multiple reliability values for each computing node. For real time applications, the model has time based reliability assessment algorithms.

This work is part of S. Malik's PhD thesis [12]

6.2.4. *Testbed Designs from Experimenters Requirements*
Participant: F. Hermenier.

The physical design of the Emulab facility, and many other testbeds like it, has been based on the facility operators' expectations regarding user needs and behavior. If operators' assumptions are incorrect, the resulting facility can exhibit inefficient use patterns and sub-optimal resource allocation.

- We have collaborated with Robert Ricci from the University of Utah on the study of the Utah' Emulab facility to provide better testbed designs. Our study gained insight into the needs and behaviors of networking researchers by analyzing more than 500,000 topologies from 13,000 experiments submitted to Emulab.
- Using this dataset, we re-visited the assumptions that went into the physical design of the Emulab facility and considered improvements to it. Through extensive simulations with real workloads, we evaluated alternative testbeds designs for their ability to improve testbed utilization and reduce hardware costs.

The results have been published to TridentCom [22], the reference conference related to testbeds and research infrastructures, and the article received the best paper award.

6.2.5. *Energy Efficient Virtual Machines Placement in Data Centers*
Participant: F. Hermenier.

Data centres are powerful ICT facilities which constantly evolve in size, complexity, and energy consumption. At the same time, tenants' and operators' requirements become more and more complex. The data centre operators may target different energy-related objectives while the workload volatility may alter the data centre capacity at supporting load spikes. Finally, clients of data centres are looking for dependable infrastructures that can comply with their SLA requirements.

To stay attractive, a data centre should then support these expectations. These constraints are however very specific to each of the tenants but also to the infrastructure. They also cover a large range of concerns (hardware requirements, performance, security ...) that are continuously evolving according to new trends and new technologies. Existing solutions are however ad-hoc and can not be updated easily to fit the data centres and the workload specificities.

We proposed a flexible energy-aware framework to address the multiple facets of an energy-aware consolidation of VMs in a cloud data centre.[21] This framework extended BtrPlace to make it able to address specific energy concerns. We integrated a fine grain energy model reducing either gas emissions or power consumption. We also proposed constraints to control the aggressiveness of these objectives to let the data centre reactive when a load spike occurs. We finally proposed various constraints to satisfy the hardware and the resource requirements of the tenants. The evaluation on a testbed running an industrial workload validated the practical benefits provided by the usage of our framework.

6.2.6. *GPU-based High Performance Cloud Computing*
Participants: M. Benguigui, F. Baude, F. Huet.

To address HPC, GPU devices are now considered as unavoidable cheap, energy efficient and very efficient alternative computing units. The barrier to handle such devices is the programming model: it is both very fine grained and synchronous.

Our long term goal is to devise some generic solutions in order to incorporate GPU-specific code whenever relevant into a parallel and distributed computation. The first step towards this objective is to gain some insight on how to efficiently program a non trivial but well known algorithm. We selected the American basked option pricing non embarrassingly parallel problem that was previously parallelized and distributed using ProActive master-slave approach [60], achieving an almost linear speedup and good performances (64 CPUs based computation allowed us to solve the problem in about 8 hours). The same algorithm has been reorganized for running on **a single GPU** [17] and achieved the same option pricing computation in about 9 hours. The current work is to succeed to take advantage of GPUs, even if non homogeneous, hired from a Cloud or a federation of clouds at once, orchestrated by an active object acting as a GPU task delegator. The goal is to drastically lower the overall computation time for such highly time consuming stochastic simulation problems.

6.3. Large-scale Simulation Platform: Techniques and methodologies
Participants: O. Dalle, E. Mancini.

In the domain of simulation techniques and methodologies, this year, we conducted research in the three following areas:

Simulation in the Cloud In recent years, numerous applications have been deployed into mobile devices. However, until now, there have been no attempts to run simulations on handheld devices. In the context of the DISSIMINET Associate Team, we work in collaboration with our partners at the Carleton University to investigate different architectures for running and managing simulations on handheld devices, and putting the simulation services in the Cloud[24]. We propose a hybrid simulation and visualization approach, where a dedicated mobile application is running on the client side and the RISE simulation server is hosted in the Cloud.

Simulation Methodology In the context of the ANR INFRA SONGS project, we are involved (as coordinators) in a Work-Package called "Open Science" whose aim is to investigate and contribute means to ensure the long term visibility and reproductibility of simulation results obtained using the SimGrid simulation platform. Our preliminary work in this direction consisted in identifying the issues, trends and potential solutions to ensure the long-term reproducibility of simulations[16].

Peer-to-peer Simulation In order to evaluate the performance and estimate the resource usage of peer-to-peer backup systems, it is important to analyze the time they spend in storing, retrieving and keeping the redundancy of the stored files. The analysis of such systems is difficult due to the random behavior of the peers and the variations of network conditions. In the context of the ANR USS-SIMGRID and INFRA-SONGS projects, we investigated means for reproducing such varying conditions in a controlled way. We worked on the design of a general simulation meta-model for peer-to-peer backup systems and a tool-chain, based on SimGrid, to help in their analysis[20]. We validated the meta-model and tool-chain through the analysis of a common scenario, and verified that they can be used, for example, for retrieving the relations between the storage size, the saved data fragment sizes and the induced network workload. We also started to investigate a new simulation technique for very-large scale distributed simulation of peer-to-peer systems based on the decomposition of a simulation into many micro-simulation steps[31] in order to optimize the overlap between communications and computations.

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

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], Alexandre Bayen [UC Berkeley, CA, USA].

The activity in traffic flow modeling has being reinforced by the creation of the Associated Team ORESTE between OPALE and the UC Berkeley teams Mobile Millennium and Integrated Corridor Management (ICM) lead by Prof. A. Bayen (see http://www-sop.inria.fr/members/Paola.Goatin/ORESTE/index.html). In this framework, three PhD students from US visited Inria during August and September, and M.L. Delle Monache spent two and half months at UC Berkeley.

During this first year of common research we proposed a new junction model for ramp-metering in the continuous and discrete settings. We focused on a junction consisting in a mainline, an on-ramp and an off-ramp. In particular, we introduced a coupled PDE-ODE model, in which the PDE describes the evolution of the cars flow on the mainline and the ODE describes the evolution of the queue length on the on-ramp, modeled by a buffer, which ensures that boundary conditions are satisfied in strong sense. At the junction we imposed the maximization of the outgoing flux together with a fixed priority parameter for incoming roads. We were able to prove existence and uniqueness of the solution of the corresponding Riemann problem. This approach has then been extended to networks and discretized using the Godunov scheme. The corresponding discrete optimization problem has been solved using the Adjoint Method and it is now being implemented into a MATLAB code. This model will serve as starting point for a subsequent model for optimal rerouting, which includes multi-commodity flow and partial control.

Besides that, we studied a a coupled PDE-ODE system modeling the interaction of a large slow moving vehicle with the surrounding traffic flow. The model consists in a scalar conservation law with moving density constraint describing traffic evolution coupled with an ODE for the slow vehicle trajectory. The constraint location moves due to the surrounding traffic conditions, which in turn are affected by the presence of the slower vehicle, thus resulting in a strong non-trivial coupling. The existence result is given in [60].

The paper [41] 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.

6.1.2. *Crowd motion*

Participants: Nora Aïssiouene, Christophe Chalons [LJLL, UP7], Régis Duvigneau, Paola Goatin, Matthias Mimault, Massimiliano D. Rosini [ICM, Warsaw University, Poland], Nicolas Seguin [LJLL, UPMC], Monika Twarogowska.

The activity on in pedestrian flow modeling is reinforced by the doctoral thesis of M. Mimault, started in October, and the enrollment of M. Twagorowska on a post-doctoral position.

Concerning crowd 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 first and second order macroscopic models for crowd movements consisting in one or two scalar conservation law accounting for mass conservation and momentum balance, 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 [40]). We are now studying existence of solutions of the corresponding initial boundary value problem, using the wave-front tracking approach. In this framework, M. Mimault's internship was devoted to develop a MATLAB code based on wave-front tracking to compute the solutions of Hughes' model of pedestrian motion with generalized running cost. This model displays a non-classical dynamic at the splitting point between the two directions of motion. The wave-front tracking scheme provides us with reference solutions to test numerically the convergence of classical finite volume schemes, which do not treat explicitly the dynamics at the turning point (see [66]). The code can be downloaded at the following URL: http://www-sop.inria.fr/members/Paola.Goatin/wft.html

From the numerical point of view, we are implementing some macroscopic models 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 is now being completed by M. Twarogowska, with the support of N. 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.

Finally, in collaboration with C. Chalons and N. Seguin, we have generalized the results on conservation laws with local flux constraint obtained in [3], [5] to general flux functions and nonclassical solutions arising for example in pedestrian flow modeling. We first define the constrained Riemann solver and the entropy condition, which singles out the unique admissible solution. We provide a well posedness result based on wave-front tracking approximations and Kruzhkov doubling of variable technique. We then provide the framework to deal with nonclassical solutions and we propose a "front-tracking" finite volume scheme allowing to sharply capture classical and nonclassical discontinuities. Numerical simulations illustrating the Braess paradox are presented as validation of the method. The results are collected in [65].

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

6.2. Optimum design and control 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.

The framework of our research aims to contribute to numerical strategies for PDE-constrained multi-objective 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 test-case of aero-structural aircraft wing shape optimization, in which the eigensplit-based optimization reveals clearly superior [38].

While the two-discipline method is currently being applied to various complex physical multi-objective situations (see in particular 6.2.4 , 6.2.5 , 6.2.6), 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 [4], see Fig. 2 .
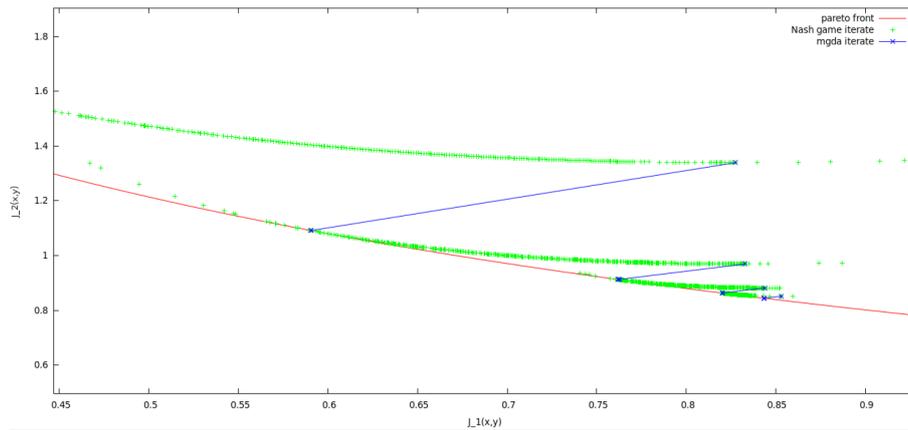
Figure 2. Combination of cooperative and competitive optimization algorithms: in red the Pareto set, in blue MGDA steps directed to the Pareto set, in green steps by Nash games with split of territory tangent to the Pareto set.

Our approach to *cooperative optimization* is based on a result of convex analysis established for a general unconstrained mult-iobjective problem in which all the gradients are assumed to be known. The theorem [39] states that in the convex hull of the gradients, there exists a unique vector of minimal norm, ω; if it is nonzero, the vector ω 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 ω 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 6.2.1 .

6.2.1. Multiple-Gradient Descent Algorithm (MGDA)

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

6.2.1.1. Theory and numerical experimentation of the MGDA construction

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 the original report [14], and in [39], we have introduced the notion of *Pareto-stationarity*, and given a first proof that it was the natural necessary condition for Pareto-optimality when the objective-functions are locally smooth in some open domain about the design-point. This report has been revised to provide a more rigorous,

and extended proof. In particular, in the revised version [14] (version 3, 2012), the number of objective-functions n and the dimension of the design space compare arbitrarily. The objective-functions are assumed to be locally convex.

Additionally, we had established that MGDA converges to Pareto-stationary design-points. This had been confirmed by numerical experiments in which MGDA had been tested over a number of classical multi-objective optimization test-cases, and found successful to converge to Pareto-optimal solutions in situations of either convex or concave Pareto sets. Additionally, MGDA [57] and PAES [69] were found to have complementary merits, making a hybrid method promising.

The method was tested successfully in a domain partition model problem in which the sub-solutions to the Poisson equation are matched at the interfaces by minimization of the integral along the interface of the squared normal-derivative jump. This academic exercise has permitted to illustrate the importance of applying an appropriate scaling to the gradients prior to calculating the descent direction [61] [47]. This has led us to define, a novel form of MGDA, consisting of a direct algorithm [62] based on a Gram-Schmidt orthogonalization conducted with a special normalization. The direct method was found more accurate and more efficient. Subsequently, we proposed two enhancements [63], the first to define the order in which the gradients are introduced in the Gram-Schmidt process uniquely and to interrupt the process as soon as the current estimate of the search direction is proved to satisfy the descent property, and the second to optimally scale the gradients when the Hessians are known, or approximated (e.g. by the BFGS estimate).

6.2.1.2. Meta-model-assisted 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 meta-models of the functionals of interest (lift, drag, etc) and to calculate approximate gradients by local finite differences. These meta-models 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 has been extended successfully to an internal flow optimization problem related to an automobile air-conditioning system and governed by the Navier-Stokes equations [55]. This more difficult problem has been proposed by Renault within the OMD2 ANR project.

6.2.1.3. Exact shape gradients

MGDA has successfully been tested over a two-objective optimization problem governed by two-dimensional elasticity. The deformation of a plate is calculated using an isogeometric approximation (see 6.6) and compliance derived from it. The exact parametric shape gradient is calculated, yielding the gradient of the objective function in two antagonistic situations differing by the loading. Pareto-fronts are thus identified.

6.2.1.4. Perspectives

MGDA offers the possibility to handle in a rational way several objective-functions for which gradients are known or approximated concurrently. This potential opens methodological paths to several themes of interest in high-fidelity simulation-based optimization: optimization of complex systems whose performance is evaluated w.r.t. several criteria originating from different, coupled disciplines; optimization under uncertainties, by introducing sensitivities as additional objectives; optimization of time-dependent systems, such as optimization of flow-control devices that generate a periodic flow (see next subsection), by converting the problem into a multi-point problem by time-discretization of the time and parameter-dependent functional; etc.

6.2.2. Flow control

Participants: Jean-Antoine Désidéri, 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 detachment. 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 resolution 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[37].

In this perspective, unsteady Reynolds Averaged Navier-Stokes equations are considered, with the Spalart-Allmaras turbulence closure. A numerical model for synthetic jets has been implemented to simulate the actuation[48], based on imposed velocity boundary conditions. Particular developments have then be carried out to include a noise term into Gaussian Process metamodels, which is used to filter errors arising from unsteady simulations/citelabroquere:hal-00742940. First results have demonstrated the feasibility of the proposed method. A systematic assessment of modeling and numerical errors is in progress, for a backward facing step test-case, with the objective of controlling the re-attachment point location.

This activity is conducted in collaboration with the CFD team of Ecole Centrale de Nantes.

6.2.3. Robust design

Participants: Jean-Antoine Désidéri, Régis Duvigneau, Daïgo Maruyama.

This work aims to develop robust design tools for aircraft design w.r.t. aerodynamic performance subject to uncertainties arising from geometrical features and fluctuations of inflow conditions. The robust design process is considered as a multi-objective optimization problem consisting of minimizing statistical quantities such as mean and variance of a cost function, typically the drag coefficient under lift constraint. MGDA is used for this purpose.

At present, analytical test cases have been tested, confirming the validity of our approach to identify the Pareto set.

One aspect of the problem is that the evaluation of these statistics and performing their optimization is very cost demanding. One solution could be, for aerodynamic design, to identify the most important variables to be treated as uncertain, possibly by the ANOVA approach, and construct adequate meta-models.

6.2.4. 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 have been set up to evaluate accurately the various physical contributions to drag. As a result, a realistic aircraft wing optimization has been conducted using a configuration provided by Airbus France as initial design. I. Ghazlane defended successfully her doctoral thesis thesis in December 2012 [34].

Besides, I. Ghazlane has realized a two-objective optimization (drag and mass reduction) via a Nash game using our optimization platform FAMOSA. These results will be included in a common publication on Nash games in preparation.

6.2.5. *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 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 [68] for the calculation of the Whitham function corresponding to the lowest-boom (axisymmetric) shape. Method and results for more general geometries have been presented internationally in [50].

Besides, aero-acoustic optimizations have been realized successfully by coupling the aerodynamic optimizer (based on Euler calculations by the elsA software) 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. These results will be included in a common publication on Nash games in preparation.

6.2.6. *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 is conducting a CIFRE thesis supported by EUROCOPTER (Marignane) at ONERA DAAP. 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 second criterion is the power necessary to maintain the forward motion. The first phase of thesis work has been devoted to the set up of a hierarchy of models from low to high fidelity, in order to calibrate appropriate functional criteria. In the current work, actual two-objective optimizations are conducted via our Nash game approach to competitive optimization with territory splitting based on reduced Hessian diagonalization. A first successful experiment has been realized in which the twist angle along the wing is optimized to reduce the power in forward motion while maintaining sub-optimality of the drag in hover. These results have been accepted for presentation at a forthcoming AIAA Conference, and will also contribute to a common publication on Nash games in preparation.

6.2.7. *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 modelling of hull shapes by parametric surfaces. The objective was to be able to modify existing hull shapes by controlling a small number of parameters, that are meaningful for naval architects. We have considered as test-case the bow shape for trawler ships[58]. As a second step, an optimum shape procedure has been set up, based on a metamodel-based optimizer, the developed CAD model and the simulation tool for free-surface flows provided by K-Epsilon company. The objective was to reduce the wave drag of a trawler ship by adding a bow, whose parameters are optimized.

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 coupling is successfully applied to a small scale industrial case, namely the shape optimization of a can bottom vis à vis dome reversal pressure and dome growth criteria. We have then defined a Nash game between criteria where the latter are approximated by the RBF metamodels. First, we validated the computation of a Nash equilibrium for mathematical functions, then we computed Nash equilibria for the small scale industrial case of the shape optimization of the can bottom. In both cases, only arbitrary territory splitting was used. Application to large scale 3D industrial problems, and the study of intelligent territory splitting algorithms 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.

We have successfully applied the NBI approach coupled to the hybrid SA+SPSA minimizer (Simulated Annealing with local search using the Simultaneous Perturbation Stochastic Approximation) to capture the Pareto front of a simple cross stamping of a high performance steel sheet. The use of cubic spline approximation of the costs (spring-back and failure criteria) turned out to be more reliable than e.g. a krigeage method.

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].
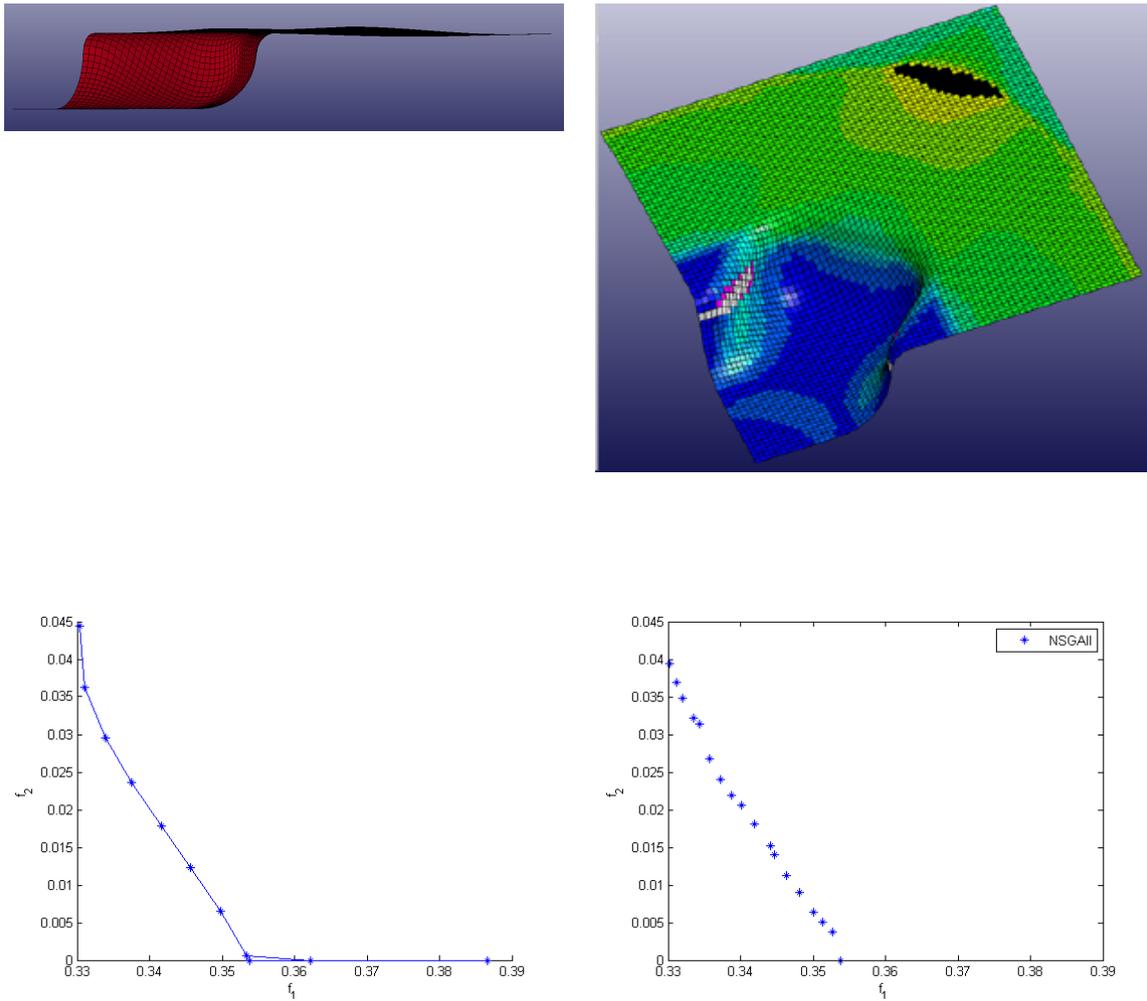
Figure 3. Multiobjective design of the stamping process of a high performance steel sheet. The costs are elastic spring-back (upper-left) and failure (upper-right). The Pareto front obtained by NNCM (lower-left) is compared to a NSGA-II one (lower-right).

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 [2] is extended in order to include the non-homogeneous wound healing or non-homogeneous dorsal closure The new model is obtained by adding extra terms that describe the particular process we want to model (lamellipodial crawling, granulation tissue contraction, extension of actin protrusions, epithelial resistance, etc.). We concentrate on the treatment of non-homogeneous forces, i.e. non-constant boundary terms which can be associated with a non-uniform cable, internal pull or zipping force due to the non-uniformity of the biological or physical properties of the boundary cells or of the connective tissue [35].

We also consider a particular yet major aspect of wound healing, namely the one related to the movement of wounded epithelial cell monolayers. The epithelial monolayer cell population, also referred to as cell-sheet, can be seen as a 2 dimensional structure, although it is well known that apical and basal sites play distinctive important roles during the migration, as well as the substrate itself. Immediately after a wound is created, the cells start to move in order to fill in the empty space. This movement, the wound closure, is a highly-coordinated collective behavior yielding a structured cohesive front, the wound leading edge. Even though wound closure involves biochemical and biomechanical processes, still far from being well understood, which are distributed over the whole monolayer, much specific attention was paid to the leading edge evolution, seen as the front of a traveling wave of the cell density function. We show that, for non inhibited wound assays, closure occurs at constant speed of the leading edge, a fact that is commonly shared by biologists and biomathematicians. But we also show that the leading edge may exhibit accelerated profiles, and that when inhibited, then the F-KPP has poor performances in modeling the leading edge dynamics.

6.5. Particular applications of simulation methods

6.5.1. *Hermitian interpolation under uncertainties*

Participants: Jean-Antoine Désideri, Manuel Bompard [Doctoral Student, ONERA/DSNA until December 2011; currently post-doctoral fellow in Toulouse], Jacques Peter [Research Engineer, ONERA/DSNA].

In PDE-constrained global optimization, iterative algorithms are commonly efficiently accelerated by techniques relying on approximate evaluations of the functional to be minimized by an economical, but lower-fidelity model (meta-model), in a so-called Design of Experiment (DoE). Various types of meta-models exist (interpolation polynomials, neural networks, Kriging models, etc). Such meta-models are constructed by pre-calculation of a database of functional values by the costly high-fidelity model. In adjoint-based numerical methods, derivatives of the functional are also available at the same cost, although usually with poorer accuracy. Thus, a question arises : should the derivative information, available but known to be less accurate, be used to construct the meta-model or ignored ? As a first step to investigate this issue, we have considered the case of the Hermitian interpolation of a function of a single variable, when the function values are known exactly, and the derivatives only approximately, assuming a uniform upper bound ε on this approximation is known. The classical notion of best approximation has been revisited in this context, and a criterion introduced to define the best set of interpolation points. This set was identified by either analytical or numerical means. If $n + 1$ is the number of interpolation points, it is advantageous to account for the derivative information when $\varepsilon \leq \varepsilon_0$, where ε_0 decreases with n, and this is in favor of piecewise, low-degree Hermitian interpolants. In all our numerical tests, we have found that the distribution of Chebyshev points is always close to optimal, and provides bounded approximants with close-to-least sensitivity to the uncertainties [56].

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.

(a)   (b)   (c)

(d)   (e)   (f)
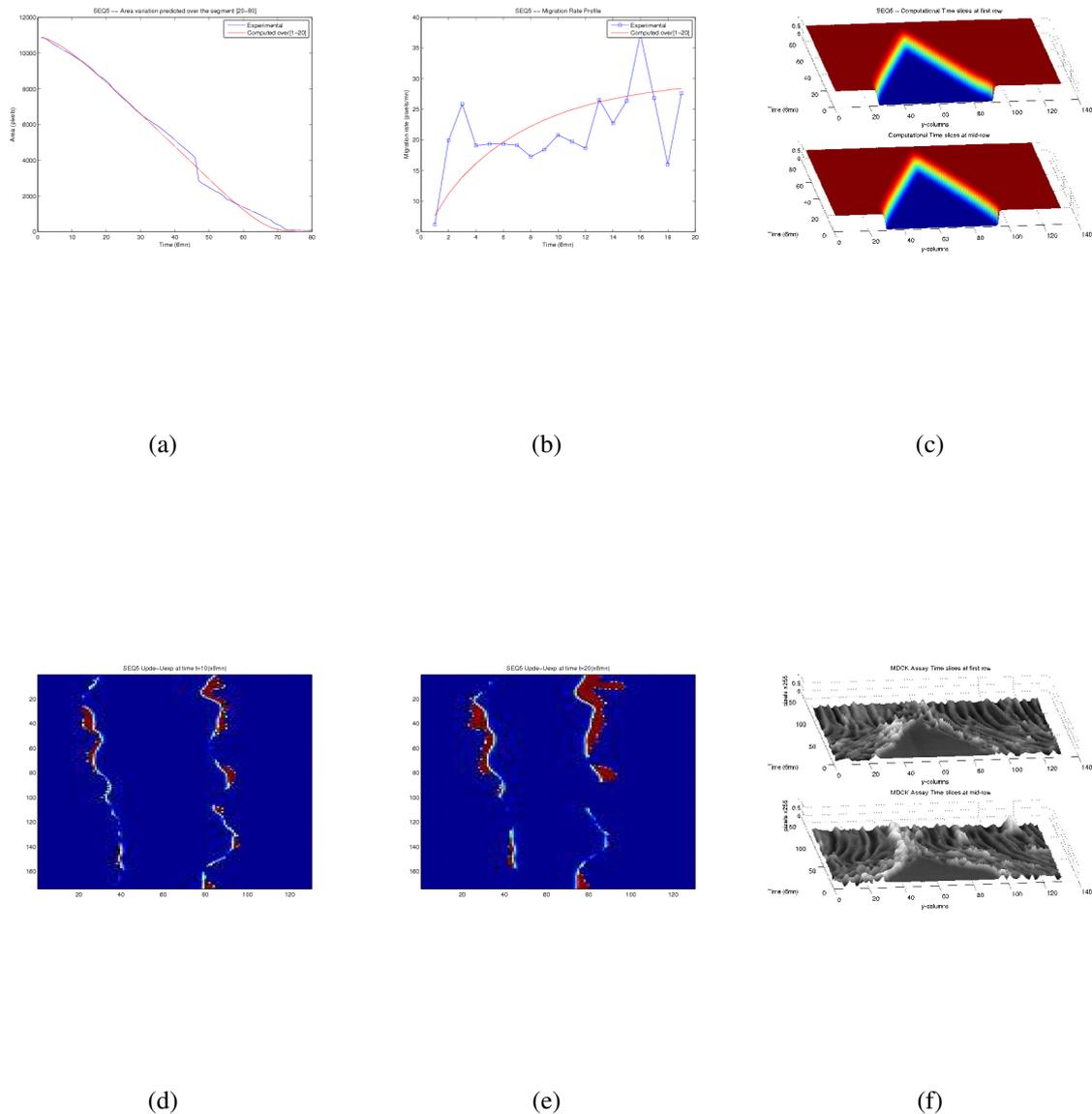
*Figure 4. **Sequence-5. Computational vs experimental wound evolution.** (a) Time variation of experimental (blue) versus computed (red) wound area (in pixels). (b) Time variation of the experimental (blue-dot) versus computed (red) migration rate (in pixels/mn). (c) Computed 3D XT view at first and mid-rows. (d) (e) (f) Traces of the difference between the experimental segmented and binarized cell-sheet images and the computed ones at different times, respectively 1hour (d), and 2hours (e) after the wounding. (f) Experimental 3D XT view at first and mid-rows.*

Several criteria have been assessed for mesh qualification in the context of inviscid-flow simulation and are currently being extended to the RANS context. These results have been presented internationally in the communication [54] and the publication [44].

6.5.3. *Hybrid meshes*

Participants: Sébastien Bourasseau, Jean-Antoine Désideri, Jacques Peter [Research Engineer, ON-ERA/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.

The on-going developments are related to the extension to the hybrid-mesh context of the full shape gradient in a 3D Eulerian flow computation.

6.5.4. *Data Completion Problems Solved as Nash Games*

Participants: Abderrahmane Habbal, Moez Kallel [University of Tunis].

The Cauchy problem for an elliptic operator is formulated as a two-player Nash game.

- Player (1) is given the known Dirichlet data, and *uses as strategy variable the Neumann condition* prescribed over the inaccessible part of the boundary.
- Player (2) is given the known Neumann data, and *plays with the Dirichlet condition* prescribed over the inaccessible boundary.
- The two players solve in parallel the associated Boundary Value Problems. Their respective objectives involve the *gap between the non used Neumann/Dirichlet known data and the traces of the BVP's solutions* over the accessible boundary, and are *coupled through a difference term.*

We prove the existence of a unique Nash equilibrium, which turns out to be the reconstructed data when the Cauchy problem has a solution. We also prove that the completion algorithm is stable with respect to noise. Many 3D experiments were performed which illustrate the efficiency and stability of our algorithm [42].

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 studies on elliptic problems have been conducted in collaboration with GALAAD Project-Team, such as the development of methods for adaptive parameterization including an a posteriori error estimate[46], [45]. A collaborative work has also been carried out with the Technical University of Kaiserslautern, concerning the computation of shape gradients for linear elasticity problems[59].

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

6. New Results

6.1. Towards Data-Centric Networking

Participants: Chadi Barakat, Damien Saucez, Jonathan Detchart, Mohamed Ali Kaafar, Ferdaouss Mattoussi, Marc Mendonca, Xuan-Nam Nguyen, Vincent Roca, Thierry Turletti.

- **DTN**

  Delay Tolerant Networks (DTNs) stand for 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. Along the main ideas of HBSD, we worked on a content optimal-delivery algorithm, CODA, that distributes content to multiple receivers over a DTN. CODA assigns a utility to each content item published in the network; this value gauges the contribution of a single content replica to the network's overall delivery-rate. CODA performs buffer management by first calculating the delivery-rate utility of each cached content-replica and then discarding the least-useful item. When an application requests content, the node supporting the application will look for the content in its cache. It will immediately deliver it to the application if the content is stored in memory. In case the request cannot be satisfied immediately, the node will store the pending request in a table. When the node meets another device, it will send the list of all pending requests to its peer; the peer device will try to satisfy this list by sending the requester all the matching content stored in its own buffer. A meeting between a pair of devices might not last long enough for all requested content to be sent. We address this problem by sequencing transmissions of data in order of decreasing delivery-rate utility. A content item with few replicas in the network has a high delivery rate utility; these items must be transmitted first to avoid degrading the content delivery-rate metric. The node delivers the requested content to the application

as soon as it receives it in its buffer. We implement CODA over the CCNx protocol, which provides the basic tools for requesting, storing, and forwarding content. Detailed information on CODA and the implementation work carried out herein can be found in [76].

- **Naming and Routing in Content Centric Networks**

  Content distribution prevails in todays 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 proposed 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. See [100] for more details.

- **On the fairness of CCN**

  Content-centric networking (CCN) is a new paradigm to better handle contents in the future Internet. Under the assumption that CCN networks will deploy a similar congestion control mechanism than in today's TCP/IP (i.e., AIMD), we built an analytical model of the bandwidth sharing in CCN based on the "square-root formula of TCP". With this model we can compare CCN download performance to what users get today. We consider different factors such as the way CCN routers are deployed, the popularity of contents, or the capacity of links and observe that when AIMD is used in a CCN network less popular content throughput is massively penalised whilst the individual gain for popular content is negligible. Finally, the main advantage of using CCN is the decrease of load at the server side. Our observations advocate the necessity to clearly define the notion of fairness in CCN and to design a proper congestion control to avoid less popular contents to become hardly accessible in tomorrow's Internet.

  Our results [75] clearly point to a fairness issue if AIMD is used with CCN. Indeed, combining blindly AIMD and CCN can severely worsen the download throughput of less popular contents with respect to the today's Internet due to subtle interactions with in-network caching strategies. The way cache memories are distributed within chain topologies has been investigated too, showing that for small and heterogeneous cache spaces, placing the biggest caches close to clients improves performance due to a smaller RTT on average. On the other hand, CCN can significantly reduce the load at the server side independently of the cache allocation strategy. Our findings advocate the urge of clearly defining the notion of fairness in CCN and designing congestion control algorithms able to limit the unfairness observed between contents of different popularities. The work is currently used within the IRTF ICNRG research group in order to motivate and define an appropriate congestion control mechanism for information centric networks like CCN. Moreover, we are currently validating the analytical results with an implementation of CCN where we can evaluate how much our model

deviates from the reality when contents are of various size or small. The implementation will also be a support to test different congestion control mechanism.

- **CCN to enable profitable collaborative OTT services**

  The ubiquity of broadband Internet and the proliferation of connected devices like laptops, tablets, or TV result in a high demand of multimedia content such as high definition video on demand (VOD) for which the Internet has been poorly designed with the Internet Protocol (IP). Information-Centric Networking and more precisely Content Centric Networking (CCN) overtake the limitation of IP by considering content as the essential element of the network instead of the topology. CCN and its content caching capabilities is particularly adapted to Over-The-Top (OTT) services like Netflix, Hulu, Xbox Live, or YouTube that distribute high-definition multimedia content to millions of consumers, independently of their location. However, bringing content as the most important component of the network implies fundamental changes in the Internet and the transition to a fully CCN Internet might take a long time. Despite this transition period where CCN and IP will co-exist, we have shown that OTT service providers and consumers have strong incentives for migrating to CCN. We also propose a transition mechanism based on the Locator/Identifier Separation Protocol (LISP) [28] that allows the provider to track the demands from its consumers even though they do not download the contents from another consumers instead of the producer itself.

  CCN, compared to IP, provides better security and performance. This last point is very interesting for OTT service providers that deliver multimedia content where performance is a key factor for the adoption of the service by consumers. With CCN, the content can be retrieved from the caches in the different CCN islands, instead of always being delivered by the content publisher. As a result, content retrieval is faster for the consumer and the operational cost of the publisher is reduced. Moreover, as the content is cached by the consumers and because the consumer can provide the content to other consumers, the overall performance increases with the number of consumers instead of decreasing as it is the case in IP today where the content is delivered by the hosting server. This property is particularly interesting because it dampens the effect of flash crowds which are normally very costly for OTT service providers as they have to provision their servers and networks to support them. Using CCN with caching at the consumers has then a direct impact on the profit earned by the OTT service provider as its costs are reduced. However, to benefit from the caching capabilities of consumers, the producer must propose real incentives to its consumers to *collaborate* and cache the content. To understand how incentives can be provided, it is necessary to remember that content in OTT is provided either freely to the consumer or in exchange of a fee. When the content is provided freely, the incomes for the publisher are ensured by advertisements dispersed in the content (e.g., banner, commercial interruptions...). A consumer has incentives to collaborate with the system if it receives some sort of discount, expressed in advertisement reduction or fee reduction. On the one hand, the discount has a cost for the publisher as its revenues will be reduced. On the other hand, the collaboration from its consumers reduces its operational costs. Hence, the publisher must determine the optimal discount, such that it maximises its profit. The situation for the consumer is the exact opposite: its costs are increasing because it is providing content to other consumers but its revenues also increase as it receives a discount on its expenses. We have determined the conditions to respect when deploying OTT with loosely collaborative consumers [99]. We currently refine the results using game theory.

- **Software-Defined Networking in Heterogeneous Networked Environments**

Software-Defined Networking (SDN) has been proposed as a way to facilitate network evolution by allowing networks and their infrastructure to be programmable. In the context of the COMMU-NITY associated team with University of California Santa Cruz (see URL http://inrg.cse.ucsc.edu/community/) , we are studying the potential of SDN to facilitate the deployment and management of new architectures and services in heterogeneous environments. In particular, we focus on the funda-mental issues related to enabling SDN in infrastructure-less/decentralized networked environments and we use OpenFlow as our target SDN platform. Our plan is to develop a hybrid SDN framework that strikes a balance between a completely decentralized approach like Active Networking and a centralized one such as OpenFlow~[58].

We are also currently evaluating the efficiency of SDN for optimizing caching in content-centric networks. CCN advocates in-network caching, i.e., to cache contents on the path from content providers to requesters. Akthough this on-path caching achieves good overall performance, we have shown that this strategy is far from being the optimal inside a domain. On this purpose, we proposed the notion of off-path caching by allowing deflection of the most popular traffic off the optimal path towards off-path caches available across the domain[100]. Off-path caching improves the global hit ratio and permits to reduce the peering links' bandwidth usage. We are now investigating whether SDN functionalities can be used to implement this optimal caching technique, in particular to identify of the most popular contents, and to configure deflection mechanisms within routers~[94].

- **Application-Level Forward Error Correction Codes (AL-FEC) and their Applications to Broadcast/Multicast Systems**

With the advent of broadcast/multicast systems (e.g., 3GPP MBMS services), 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 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, with a commercial service that started in April 2012. This is the first adoption of these codes in an international standard. These codes, along with our FLUTE/ALC software, are now part of the server and terminal protocol stack: http://www.rapidtvnews.com/index.php/2012041721327/ntt-data-mse-and-expways-joint-solution-powers-japanese-mobile-tv-service.html.

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 recent Internet-Drafts [91][92] define 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 for 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 im-proved decoding speed and reduced memory requirements, is commercialized through an industrial partner, Expway.

Since May 2012, along with the Expway French company, we are proposing the Reed-Solomon + LDPC-Staircase codes for the 3GPP-eMBMS call for technology, as a candidate for next generation AL-FEC codes for multimedia services. We have shown that these codes offer very good erasure

recovery capabilities, in line with 3GPP requirements, and extremely high decoding speeds, usually significantly faster than that of the other proposals. The final decision is expected for end of January 2013. In any case we have once again showed that these codes provide very good performance, often ahead of the competitors, and an excellent balance between several technical and non technical criteria.

Finally our activities in the context of the PhD of F. Mattoussi include 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 (IETF RFC 5170), while keeping backward compatibility (i.e. LDPC-Staircase "codewords" can be decoded with a GPLDPC-Staircase codec). More information is available in [56][57][55].

- **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 remaing) 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 independant 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 [91] 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 [92], [91].

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 [64]. We have shown that the GOE approach, by the flexibility it offers, its simplicity, its backward compatibility and its good recovery capabilities (under finite of 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 RFC6363.

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: [85] for Reed-Solomon and [88] for LDPC-Staircase. Both documents are close to being published as RFCs.

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. 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 [95] 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 [82]. 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.

- **Security of the Broadcast/Multicast Systems**

Sooner or later, broadcasting systems will require security services. This is all the more true as heterogeneous broadcasting technologies are used, some of them being by nature open, such as WiFi networks. Therefore, one of the key security services is the authentication of the packet origin and the packet integrity check. To that purpose, 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 and the standard is now published as IETF RFC 6584 [98].

- **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 2011-2012 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 [51].

6.2. Network Security and Privacy

Participants: Claude Castelluccia, Gergely Acs, Mathieu Cunche, Daniele Perito, Lukasz Olejnik, Mohamed Ali Kaafar, Abdelberi Chaabane, Cédric Lauradoux, Minh-Dung Tran.

- *Private Big Data Publication* Public datasets are used in a variety of applications spanning from genome and web usage analysis to location-based and recommendation systems. Publishing such datasets is important since they can help us analyzing and understanding interesting patterns. For example, mobility trajectories have become widely collected in recent years and have opened the possibility to improve our understanding of large-scale social networks by investigating how people exchange information, interact, and develop social interactions. With billion of handsets in use worldwide, the quantity of mobility data is gigantic. When aggregated, they can help understand complex processes, such as the spread of viruses, and build better transportation systems, prevent traffic congestion. While the benefits provided by these datasets are indisputable, they unfortunately pose a considerable threat to individual privacy. In fact, mobility trajectories might be used by a malicious attacker to discover potential sensitive information about a user, such as his habits, religion or relationships. Because privacy is so important to people, companies and researchers are reluctant to publish datasets by fear of being held responsible for potential privacy breaches. As a result, only very few of them are actually released and available. This limits our ability to analyze such data to derive information that could benefit the general public. Here follows some recent results of our activities in this domain.

  Privacy-Preserving Sequential Data Publication [41]: Sequential data is being increasingly used in a variety of applications, spanning from genome and web usage analysis to location-based recommendation systems. Publishing sequential data is of vital importance to the advancement of these applications since they can enable researchers to analyze and understand interesting sequential patterns. However, as shown by the re-identification attacks on the AOL and Netflix datasets, releasing sequential data may pose considerable threats to individual privacy. Recent research has indicated the failure of existing sanitization techniques to provide claimed privacy guarantees. It is therefore urgent to respond to this failure by developing new schemes with provable privacy guarantees. Differential privacy is one of the only models that can be used to provide such guarantees. Due to the inherent sequentiality and high-dimensionality, it is challenging to apply differential privacy to sequential data. In this work, we address this challenge by employing a variable-length n-gram model, which extracts the essential information of a sequential database in terms of a set of variable-length n-grams. Our approach makes use of a carefully designed exploration tree structure and a set of novel techniques based on the Markov assumption in order to lower the magnitude of added noise. The published n-grams are useful for many purposes. Furthermore, we develop a solution for generating a synthetic database, which enables a wider spectrum of data analysis tasks. Extensive experiments on real-life datasets demonstrate that our approach substantially outperforms the state-of-the-art techniques.

  Private Histogram Publishing [33]:

  Differential privacy can be used to release different types of data, and, in particular, histograms, which provide useful summaries of a dataset. Several differentially private histogram releasing schemes have been proposed recently. However, most of them directly add noise to the histogram counts, resulting in undesirable accuracy. In this work, we propose two sanitization techniques that exploit the inherent redundancy of real-life datasets in order to boost the accuracy of histograms. They lossily compress the data and sanitize the compressed data. Our first scheme is an optimization of the Fourier Perturbation Algorithm (FPA) presented in [13]. It improves the accuracy of the initial FPA by a factor of 10. The other scheme relies on clustering and exploits the redundancy between bins. Our extensive experimental evaluation over various real-life and synthetic datasets demonstrates that our techniques preserve very accurate distributions and considerably improve the accuracy of range queries over attributed histograms.

- *Privacy Issues on the Internet* Internet users are being increasingly tracked and profiled. Companies utilize profiling to provide customized, i.e. personalized services to their customers, and hence increase revenues.

  Privacy issues of Targeted Advertising [37]: Behavioral advertising takes advantage from profiles of users' interests, characteristics (such as gender, age and ethnicity) and purchasing activities. For example, advertising or publishing companies use behavioral targeting to display advertisements that closely reflect users' interests (e.g. 'sports enthusiasts'). Typically, these interests are inferred from users' web browsing activities, which in turn allows building of users' profiles. It can be argued that customization resulting from profiling is also beneficial to users who receive useful information and relevant online ads in line with their interests. However, behavioral targeting is often perceived as a threat to privacy mainly because it heavily relies on users' personal information, collected by only a few companies. In this work, we show that behavioral advertising poses an additional privacy threat because targeted ads expose users' private data to any entity that has access to a small portion of these ads. More specifically, we show that an adversary who has access to a user's targeted ads can retrieve a large part of his interest profile. This constitutes a privacy breach because interest profiles often contain private and sensitive information.

  On the Uniqueness of Web Browsing History Patterns [60]: We present the results of the first large-scale study of the uniqueness of Web browsing histories, gathered from a total of $368,284$ Internet users who visited a history detection demonstration website. Our results show that for a majority of users (69%), the browsing history is unique and that users for whom we could detect at least 4 visited websites were uniquely identified by their histories in 97% of cases. We observe a high rate of stability in browser history fingerprints: for repeat visitors, 80% of fingerprints are identical over time, and differing ones were strongly correlated with original history contents, indicating static browsing preferences. We report a striking result that it is enough to test for a small number of pages in order to both enumerate users' interests and perform an efficient and unique behavioral fingerprint; we show that testing 50 web pages is enough to fingerprint 42% of users in our database, increasing to 70% with 500 web pages. Finally, we show that indirect history data, such as information about *categories* of visited websites can also be effective in fingerprinting users, and that similar fingerprinting can be performed by common script providers such as Google or Facebook.

- **Adaptive Password-Strength Meters from Markov Models [38]**

  Passwords are a traditional and widespread method of authentication, both on the Internet and off-line. Passwords are portable, easy to understand for laypersons, and easy to implement for the operator. Thus, password-based authentication is likely to stay for the foreseeable future.

  To ensure an acceptable level of security of user-chosen passwords, sites often use mechanisms to test the strength of a password (often called *pro-active password checkers*) and then reject weak passwords. Hopefully this ensures that passwords are reasonably strong on average and makes guessing passwords infeasible or at least too expensive for the adversary. Commonly used password checkers rely on rules such as requiring a number and a special character to be used. However, as we will show and also has been observed in previous work, the accuracy of such password checkers is low, which means that often insecure passwords are accepted and secure passwords are rejected. This adversely affects both security and usability.

  In this work, we propose to use password strength meters based on Markov-models, which estimate the true strength of a password more accurately than rule-based strength meters. Roughly speaking, the Markov-model estimates the strength of a password by estimating the probability of the n-grams that compose said password. Best results can be obtained when the Markov-models are trained on the actual password database. We show, in this work, how to do so without sacrificing the security of the password database, even when the n-gram database is leaked.

  We show how to build secure adaptive password strength meters, where security should hold even when the n-gram database leaks. This is similar to traditional password databases, where one tries

to minimize the effects of a database breach by hashing and salting the stored passwords. This is not a trivial task. One potential problem is that, particularly strong passwords, can be leaked entirely by an n-gram database (without noise added).

- **Fast Zero-Knowledge Authentication [47]** We explore new area/throughput trade-offs for the Girault, Poupard and Stern authentication protocol (GPS). This authentication protocol was selected in the NESSIE competition and is even part of the standard ISO/IEC 9798. The originality of our work comes from the fact that we exploit a fixed key to increase the throughput. It leads us to implement GPS using the Chapman constant multiplier. This parallel implementation is 40 times faster but 10 times bigger than the reference serial one. We propose to serialize this multiplier to reduce its area at the cost of lower throughput. Our hybrid Chapman's multiplier is 8 times faster but only twice bigger than the reference. Results presented here allow designers to adapt the performance of GPS authentication to their hardware resources. The complete GPS prover side is also integrated in the network stack of the PowWow sensor which contains an Actel IGLOO AGL250 FPGA as a proof of concept.

- **Energy Efficient Authentication Strategies for Network Coding [26]**

  Recent advances in information theory and networking, e.g. aggregation, network coding or rateless codes, have significantly modified data dissemination in wireless networks. These new paradigms create new threats for security such as pollution attacks and denial of services (DoS). These attacks exploit the difficulty to authenticate data in such contexts. The particular case of xor network coding is considered herein. We investigate different strategies based on message authentication codes algorithms (MACs) to thwart these attacks. Yet, classical MAC designs are not compatible with the linear combination of network coding. Fortunately, MACs based on universal hash functions (UHFs) match nicely the needs of network coding: some of these functions are linear $h(x_1 \oplus x_2) = h(x_1) \oplus h(x_2)$. To demonstrate their efficiency, we consider the case of wireless sensor networks (WSNs). Although these functions can drastically reduce the energy consumption of authentication (up to 68% gain over the classical designs is observed), they increase the threat of DoS. Indeed, an adversary can disrupt all communications by polluting few messages. To overcome this problem, a group testing algorithm is introduced for authentication resulting in a complexity linear in the number of attacks. The energy consumption is analyzed for cross-point and butterfly network topologies with respect to the possible attack scenarios. The results highlight the trade-offs between energy efficiency, authentication and the effective throughput for the different MAC modes.

- **Towards Stronger Jamming Model: Application to TH-UWB Radio [35]**

  With the great expansion of wireless communications, jamming becomes a real threat. We propose a new model to evaluate the robustness of a communication system to jamming. The model results in more scenarios to be considered ranging from the favorable case to the worst case. The model is applied to a TH-UWB radio. The performance of such a radio in presence of the different jamming scenarios is analyzed. We introduce a mitigation solution based on stream cipher that restricts the jamming problem of the TH-UWB communication to the more favorable case while preserving confidentiality.

- **Privacy risks quantification in Online social networks**

  In this project, we analyze the different capabilities of online social networks and aim to quantify the privacy risks users are undertaking in this context. Online Social Networks (OSNs) are a rich source of information about individuals. It may be difficult to justify the claim that the existence of public profiles breaches the privacy of their owners, as they are the ones who entered the data and made them publicly available in the first place. However, aggregation of multiple OSN public profiles is debatably a source of privacy loss, as profile owners may have expected each profile's information to stay within the boundaries of the OSN service in which it was created. First we present an empirical study of personal information revealed in public profiles of people who use multiple Online Social Networks (OSNs). This study aims to examine how users reveal their personal information across multiple OSNs. We consider the number of publicly available attributes in public

profiles, based on various demographics and show a correlation between the amount of information revealed in OSN profiles and specific occupations and the use of pseudonyms. Then, we measure the complementarity of information across OSNs and contrast it with our observations about users who share a larger amount of information. We also measure the consistency of information revelation patterns across OSNs, finding that users have preferred patterns when revealing information across OSNs. To evaluate the quality of aggregated profiles we introduce a consistency measure for attribute values, and show that aggregation also improves information granularity. Finally, we demonstrate how the availability of multiple OSN profiles can be exploited to improve the success of obtaining users' detailed contact information, by cross-linking with publicly available data sources such as online phone directories. This work has been published in ACM SIGCOMM WOSN [42].

In a second study, we examine the user tracking capabilities of the three major global Online Social Networks (OSNs). We study the mechanisms which enable these services to persistently and accurately follow users web activity, and evaluate to which extent this phenomena is spread across the web. Through a study of the top 10K websites, our findings indicate that OSN tracking is diffused among almost all website categories, independently from the content and from the audience. We also evaluate the tracking capabilities in practice and demonstrate by analyzing a real traffic traces that OSNs can reconstruct a significant portion of users web profile and browsing history. We finally provide insights into the relation between the browsing history characteristics and the OSN tracking potential, highlighting the high risk properties. This work has also been published in ACM SIGCOMM WOSN [40].

In a third study, we also analyzed the inference capabilities of third parties from seemingly harmless and unconsciously publicly shared data. Interests (or "likes") of users is one of the highly-available on-line information on the web. In this study, 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 friend relationships 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. This is the first time that user interests are used for profiling, and more generally, semantics-driven inference of private data is addressed. Our work received many media attention and was published in the prestigious NDSS symposium [39].

- **On the Privacy threats of hidden information in Wireless communication**

Wi-Fi protocol has the potential to leak personal information. Wi-Fi capable devices commonly use active discovery mode to find the available Wi-Fi access points (APs). This mechanism includes broadcast of the AP names to which the mobile device has previously been connected to, in plain text, which may be easily observed and captured by any Wi-Fi device monitoring the control traffic. The combination of the AP names belonging to any mobile device can be considered as a Wi-Fi fingerprint, which can be used to identify the mobile device user. Our research investigates how it is possible to exploit these fingerprints to identify links between users i.e. owners of the mobile devices broadcasting such links. In this project, we have used an approach based on the similarity between the Wi-Fi fingerprints, which is equated to the likelihood of the corresponding users being linked. When computing the similarity between two Wi-Fi fingerprints, two dimensions need to be considered : (i) The number of network names in common. Indeed, sharing a network is an indication of the existence of a link, e.g. friends and family that share multiple Wi-Fi networks. (ii) The rarity of the network names in common. Some network names are very common and sharing them does not imply a link between the users. This is the case for public network names such as McDonalds Free Wi-Fi, or default network names such as NETGEAR and Linksys. On the other hand, uncommon network names such as Griffin Family Network or Orange-3EF50 are likely to

indicate a strong link between the users of these networks. Utilising a carefully designed similarity metric, we have been able to infer the existence of social links with a high confidence: 80% of the links were detected with an error rate of 7%. We show that through real-life experiments that owners of smartphones are particularly exposed to this threat, as indeed these devices are carried on persons throughout the day, connecting to multiple Wi-Fi networks and also broadcasting their connection history. There are a number of industry and research initiatives aiming to address Wi-Fi related privacy issues. The deployment of new technology i.e. privacy preserving discovery services, would necessitate software modifications in currently deployed APs and devices. The obvious solution to disable active discovery mode, comes at the expense of performance and usability, i.e. with an extended time duration for the Wi-Fi capable device to find and connect to an available AP. As a possible first step, users should be encouraged to remove the obsolete connection history entries, which may lower the similarity metric and thus reduce the ease of linkage. Our papers illustrating this study have been presented in the WoWMoM'12 conference [45] and in the IEEE MILCOM conference [43].

- **Information leakage in Ads networks**

  In targeted (or behavioral) advertising, users' behaviors are tracked over time in order to customize served ads to their interests. This creates serious privacy concerns since for the purpose of profiling, private information is collected and centralized by a limited number of companies. Despite claims that this information is secure, there is a potential for this information to be leaked through the customized services these companies are offering. In this study, we show that targeted ads expose users' private data not only to ad providers but also to any entity that has access to users' ads. We propose a methodology to filter targeted ads and infer users' interests from them. We show that an adversary that has access to only a small number of websites containing Google ads can infer users' interests with an accuracy of more than 79% (Precision) and reconstruct as much as 58% of a Google Ads profile in general (Recall). This study is the first work that identifies and quantifies information leakage through ads served in targeted advertising. We published a paper illustrating these results in the prestigious Privacy Enhancing Technologies Symposium PETS 2012 [37].

- **Privacy in P2P file sharing systems**

  In this study, we aim at characterizing anonymous file sharing systems from a privacy perspective. We concentrate on a recently deployed privacy-preserving file sharing system: OneSwarm. Our characterisation is based on measurement of several aspects of the OneSwarm system such as the nature of the shared and searched content and the geolocation and number of users. Our findings indicate that, as opposed to common belief, there is no significant difference in downloaded content between this system and the classical BitTorrent ecosystem. We also found that a majority of users appear to be located in countries where anti-piracy laws have been recently adopted and enforced (France, Sweden and U.S). Finally, we evaluate the level of privacy provided by OneSwarm, and show that, although the system has strong overall privacy, a collusion attack could potentially identify content providers. This work has been published in [46].

- **Privacy leakage on mobile devices: the Mobilitics Inria-CNIL project**

  This joint Inria-CNIL (the French data protection agency) project aims at assessing the privacy risks associated to the use of smartphones and tablets, in particular because of personal information leakage to remote third parties. Both applications and the base OS services are considered as potential source of information leakage. More precisely, the goals are to define a platform and a methodology to identify, measure, and see the evolution over the time of privacy risks.

  If similar risks exist with a PC, the situation is more worrying with mobile terminals. The reasons are:

  - the intrusive feature of these terminals that their owner continuously keep with them;
  - the amount of personnal information available on these terminals (mobile terminals aggregate personnal information but also create them, for instance with geolocalisation information);

– the facility with which the owner can personnalize its terminal with new applications;

– the financial incentives that lead companies to collect and use personnal information;

– the fact that the terminal user has no tool (e.g. a "privacy" firewall) to control precisely what information is exchanged with whom. The permissions provided by Android is too coarse grained to be useful, and the new privacy dashboard of IOS 6 does not enable the user to have an idea of how personnal information is used by an authorized application (a one time access to a personnal information and local processing within the application can be acceptable, whereas the periodic transmission of this information to remote servers is not);

The final goals of the Mobilitics project are both to study the situation and trend, but also to make mobile terminal users aware of the situation, and to provide tools that may help them to better control the personnal information flow of their terminal.

6.3. Formal and legal issues of privacy

Participants: Thibaud Antignac, Denis Butin, Daniel Le Métayer.

- **Verification of privacy properties** The increasing official use of security protocols for electronic voting deepens the need for their trustworthiness, hence for their formal verification. The impossibility of linking a voter to her vote, often called voter privacy or ballot secrecy, is the core property of many such protocols. Most existing work relies on equivalence statements in cryptographic extensions of process calculi. We have proposed the first theorem-proving based verification of voter privacy which overcomes some of the limitations inherent to process calculi-based analysis [36]. Unlinkability between two pieces of information is specified as an extension to the Inductive Method for security protocol verification in Isabelle/HOL. New message operators for association extraction and synthesis are defined. Proving voter privacy demanded substantial effort and provided novel insights into both electronic voting protocols themselves and the analysed security goals. The central proof elements have been shown to be reusable for different protocols with minimal interaction.

- **Privacy by design** The privacy by design approach is often praised by lawyers as well as computer scientists as an essential step towards a better privacy protection. The general philosophy of privacy by design is that privacy should not be treated as an afterthought but rather as a first-class requirement during the design of a system. The approach has been applied in different areas such as smart metering, electronic traffic pricing, ubiquitous computing or location based services. More generally, it is possible to identify a number of core principles that are widely accepted and can form a basis for privacy by design. For example, the Organization for Economic Co-operation and Development (OECD) has put forward principles such as the consent, limitation of use, data quality, security and accountability. One must admit however that the take-up of privacy by design in the industry is still rather limited. This situation is partly due to legal and economic reasons: as long as the law does not impose binding commitments, ICT providers and data collectors do not have sufficient incentives to invest into privacy by design. The situation on the legal side might change in Europe though because the regulation proposed by the European Commission in January 2012 (to replace the European Directive 95/46/EC) includes binding commitments on privacy by design.

But the reasons for the lack of adoption of privacy by design are not only legal and economic: even though computer scientists have devised a wide range of privacy enhancing tools, no general methodology is available to integrate them in a consistent way to meet a set of privacy requirements. The next challenge in this area is thus to go beyond individual cases and to establish sound foundations and methodologies for privacy by design. As a first step in this direction, we have focused on the data minimization principle which stipulates that the collection should be limited to the pieces of data strictly necessary for the purpose, and we have proposed a framework to reason about the choices of architecture and their impact in terms of privacy [53]. The first strategic choices are the allocation of the computation tasks to the nodes of the architecture and the types of communications between the nodes. For example, data can be encrypted or hashed, either to protect

their confidentiality or to provide guarantees with respect to their correctness or origin. The main benefit of a centralized architecture for the "central" actor is that he can trust the result because he keeps full control over its computation. However, the loss of control by a single actor in decentralized architectures can be offset by extra requirements ensuring that errors (or frauds) can be detected *a posteriori*. In order to help the designer grasp the combination of possible options, our framework provides means to express the parameters to be taken into account (the service to be performed, the actors involved, their respective requirements, etc.) and an inference system to derive properties such as the possibility for an actor to detect potential errors (or frauds) in the computation of a variable. This inference system can be used in the design phase to check if an architecture meets the requirements of the parties or to point out conflicting requirements.

- **Privacy and discrimination**

  Actually, the interactions between personal data protection, privacy and protection against discriminations are increasingly numerous and complex. For example, there is no doubt that misuses of personal data can adversely affect privacy and self-development (for example, resulting in the unwanted disclosure of personal data to third parties, in identity theft, or harassment through email or phone calls), or lead to a loss of choices or opportunities (for example, enabling a recruiter to obtain information over the internet about political opinions or religious beliefs of a candidate and to use this information against him). It could even be suggested that privacy breaches and discriminations based on data processing are probably the two most frequent and the most serious types of consequences of personal data breaches. We have studied these interactions from a multidisciplinary (legal and technical) perspective and argued that an extended application of the application of non-discrimination regulations could help strengthening data protection [52]. We have analysed and compared personal data protection, privacy and protection against discriminations considering both the types of data concerned and the *modus operandi* (*a priori* versus *a posteriori* controls, actors in charge of the control, etc.). From this comparison, we have drawn some conclusions with respect to their relative effectiveness and argued that *a posteriori* controls on the use of personal data should be strengthened and the victims of data misuse should get compensations which are significant enough to represent a deterrence for data controllers. We have also advocated the establishment of stronger connections between anti-discrimination and data protection laws, in particular to ensure that any data processing leading to unfair differences of treatments between individuals is prohibited and can be effectively punished [29].

6.4. Network measurement, modeling and understanding

Participants: Chadi Barakat, Arnaud Legout, Ashwin Rao, Walid Dabbous, Tessema Mindaye, Mohamed Ali Kaafar, Dong Wang, Vincent Roca, Ludovic Jacquin, Byungchul Park.

The main objective of our work in this domain is a better monitoring of the Internet and a better understanding of its traffic. 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 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.

- **Checking Traffic Differentiation at the Internet Access**

  In the last few years, ISPs have been reported to discriminate against specific user traffic, especially if generated by bandwidth-hungry applications. The so-called network neutrality, advocating that an ISP should treat all incoming packets equally, has been a hot topic ever since. We propose Chkdiff, a novel method to detect network neutrality violations that takes a radically different approach from existing work: it aims at both application and differentiation technique agnosticism. We achieve this in three steps. Firstly, we perform measurements with the user's real traffic instead of using specific

application traces. Secondly, we do not assume that discrimination takes place on any particular packet field, which requires us to preserve the integrity of all the traffic we intend to test. Thirdly, we detect differentiation by comparing the performance of a traffic flow against that of all other traffic flows from the same user, considered as a whole.

Chkdiff is based on the following key ideas:

Idea 1: **Use real user traffic**. We want to test the existence of traffic discrimination for the exact set of applications run by the end user. Hence, we only consider user-generated traffic.

Idea 2: **Leave user traffic unchanged, or almost**. All methods performing active measurements send probes made of real application packets and of packets that are similar, but slightly modified, so that they do not get discriminated along their path. This is quite an assumption, as we do not know exactly what ISPs do behind the scenes. In the extreme case, ISPs could even white-list traffic generated by differentiation detecting tools. It is therefore crucial to preserve as much of the original packets as possible, as well as their original per-flow order. We will see that the modifications introduced by our tool affect only the ordering of packets, their TTL value or their IP identification field.

Idea 3: **Baseline is the entire traffic performance**. Since we do not want to make any hypothesis in advance on what kind of mechanisms - if any - are deployed, we claim that the performance of each single non-differentiated flow should present the same behaviour as that of the rest of our traffic as a whole. Differentiated flows, on the other hand, should stand out when compared to all other flows grouped together, where a large fraction of non-differentiated flows should mitigate the impact of differentiated ones.

Chkdiff is currently the subject of a collaboration with I3S around the PhD thesis of Riccardo Ravaioli (funded by the Labex UCN@Sophia). A first description of the tool is presented in [63].

- **Lightweight Enhanced Monitoring for High-Speed Networks**

  Within the collaboration with Politecnico di Bari, we worked on LEMON, a lightweight enhanced monitoring algorithm based on packet sampling. This solution targets a pre-assigned accuracy on bitrate estimates, for each monitored flow at a router interface. To this end, LEMON takes into account some basic properties of the flows, which can be easily inferred from a sampled stream, and exploits them to dynamically adapt the monitoring time-window on a per-flow basis. Its effectiveness is tested using real packet traces. Experimental results show that LEMON is able to finely tune, in real-time, the monitoring window associated to each flow and its communication overhead can be kept low enough by choosing an appropriate aggregation policy in message exporting. Moreover, compared to a classic fixed-scale monitoring approach, it is able to better satisfy the accuracy requirements of bitrate estimates. Finally, LEMON incurs a low processing overhead, which can be easily sustained by currently deployed routers, such as a CISCO 12000 device. This work is currently under submission.

- **The Complete Picture of the Twitter Social Graph**

  In this work [49], we collected the entire Twitter social graph that consists of 537 million Twitter accounts connected by 23.95 billion links, and performed a preliminary analysis of the collected data. In order to collect the social graph, we implemented a distributed crawler on the PlanetLab infrastructure that collected all information in 4 months. Our preliminary analysis already revealed some interesting properties. Whereas there are 537 million Twitter accounts, only 268 million already sent at least one tweet and no more than 54 million have been recently active. In addition, 40% of the accounts are not followed by anybody and 25% do not follow anybody. Finally, we found that the Twitter policies, but also social conventions (like the followback convention) have a huge impact on the structure of the Twitter social graph.

- **Meddle: Middleboxes for Increased Transparency and Control of Mobile Traffic**

  Mobile networks are the most popular, fastest growing and least understood systems in today's Internet ecosystem. Despite a large collection of privacy, policy and performance issues in mobile networks users and researchers are faced with few options to characterize and address them. In this work [62] we designed Meddle, a framework aimed at enhancing transparency in mobile networks and providing a platform that enables users (and researchers) control mobile traffic. In the mobile environment, users are forced to interact with a single operating system tied to their device, generally run closedsource apps that routinely violate user privacy, and subscribe to network providers that can (and do) transparently modify, block or otherwise interfere with network traffic. Researchers face a similar set of challenges for characterizing and experimenting with mobile systems. To characterize mobile traffic and design new protocols and services that are better tailored to the mobile environment, we would like a framework that allows us to intercept and potentially modify traffic generated by mobile devices as they move with users, regardless of the device, OS, wireless technology, or carrier. However, implementing this functionality is difficult on mobile devices because it requires warrantyvoiding techniques such as jail breaking to access and manipulate traffic at the network layer. Even when using such an approach, carriers may manipulate traffic once it leaves the mobile device, thus rendering some research impractical. Furthermore, researchers generally have no ability to deploy solutions and services such as prefetching and security filters, that should be implemented in the network. In this work, we designed Meddle, a framework that combines virtual private networks (VPNs) with middleboxes to provide an experimental platform that aligns the interests of users and researchers.

- **Mobile users' behavior modeling in Video on Demand systems and its implication on user privacy and caching strategies**

  In this project, we examine mobile users' behavior and their corresponding video viewing patterns from logs extracted from the servers of a large scale VoD system. We focus on the analysis of the main discrepancies that might exist when users access the VoD system catalog from WiFi or 3G connections. We also study factors that might impact mobile users' interests and video popularity. The users' behavior exhibits strong daily and weekly patterns, with mobile users' interests being surprisingly spread across almost all categories and video lengths, independently of the connection type. However, by examining the activity of users individually, we observed a concentration of interests and peculiar access patterns, which allows to classify the users and thus better predict their behavior. We also find the skewed video popularity distribution and demonstrate that the popularity of a video can be predicted using its very early popularity level. We then analyzed the sources of video viewing and found that even if search engines are the dominant sources for a majority of videos, they represent less than 10% (resp. 20%) of the sources for the highly popular videos in 3G (resp. WiFi) network. We also report that both the type of connection and the type of mobile device used have an impact on the viewing time and the source of viewing. Using our findings, we provide insights and recommendations that can be used to design intelligent mobile VoD systems and help in improving personalized services on these platforms. This work has been published in IMC 2012 [54].

- **Explicative models for Information Spreading on the web from a user profiling perspective**

Microblog services offer a unique approach to online information sharing allowing microblog users to forward messages to others. We study the process of information diffusion in a microblog service developing Galton-Watson with Killing (GWK) model, which has many implications ranging from privacy protection to experiments validation and benchmarking. We describe an information propagation as a discrete GWK process based on Galton-Watson model which models the evolution of family names. Our model explains the interaction between the topology of the social graph and the intrinsic interest of the message. We validate our models on dataset collected from Sina Weibo and Twitter microblogs. Sina Weibo is a Chinese microblog web service which reached over 100 million users as for January 2011. Our Sina Weibo dataset contains over 261 thousand tweets which have retweets and 2 million retweets from 500 thousand users. Twitter dataset contains over 1.1 million tweets which have retweets and 3.3 million retweets from 4.3 million users. The results of the validation show that our proposed GWK model fits the information diffusion of microblog service very well in terms of the number of message receivers. We show that our model can be used in generating tweets load and also analyze the relationships between parameters of our model and popularity of the diffused information. Our work is the first to give a systemic and comprehensive analysis for the information diffusion on microblog services, to be used in tweets-like load generators while still guaranteeing popularity distribution characteristics. Our paper illustrating this study will be presented in IEEE Infocom 2013 [69].

- **Tracking ICMP black holes at an Internet Scale**

  ICMP is a key protocol to exchange control and error messages over the Internet. An appropriate ICMP's processing throughout a path is therefore a key requirement both for troubleshooting operations (e.g. debugging routing problems) and for several functionalities (e.g. Path Maximum Transmission Unit Discovery, PMTUD). Unfortunately it is common to see ICMP malfunctions, thereby causing various levels of problems. In our study, we first introduce a taxonomy of the way routers process ICMP, which is of great help to understand for instance certain traceroute outputs. Secondly we introduce IBTrack, a tool that any user can use to automatically characterize ICMP issues within the Internet, without requiring any additional in-network assistance (e.g. there is no vantage point). Finally we validate our IBTrack tool with large scale experiments and we take advantage of this opportunity to provide some statistics on how ICMP is managed by Internet routers. This work has been presented in IEEE Globecom [51].

6.5. Experimental Environment for Future Internet Architecture

Participants: Walid Dabbous, Thierry Parmentelat, Fréderic Urbani, Daniel Camara, Alina Quereilhac, Shafqat Ur-Rehman, Mohamed Larabi, Thierry Turletti, Julien Tribino.

- **SFA Federation of experimental testbeds**

  We are now involved in the NOVI (E.U. STREP) project, the F-Lab (French A.N.R.) project, the FED4FIRE (E.U. IP) project and have the lead of the "Control Plane Extensions" WorkPackage of OpenLab (E.U. IP) project. Within these frameworks, as part of the co-development agreement between the Planète team and Princeton University, we have made a great deal of contributions into one of the most visible and renown implementations of 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 2012, we have focused on the maturation of the SFA specifications and the SfaWrap codebase, with several objectives in mind. Firstly, we have contributed within the GENI (N.S.F.) project to the specifications of the Version 3 of the AM-API (Aggregate Manager API), which defines the primitives that a testbed management infrastructure has to provide in order to be SFA-compliant.

Secondly, knowing that our former SFA implementation was targeting PlanetLab testbeds only, we needed on the one hand, to make generic this SFA implementation, by completely redesign and refactor its codebase, and on the other hand, we needed to support all the resources allocation strategies supported by the testbeds, namely the allocation of both 'shared' and 'exclusive' resources. As a result of this redesign and development effort, out new SFA implementation is now disseminated and started to be known, under the name of SfaWrap, and we believe that it can be used as a production-grade alternative to quickly add SFA compatibility on top of many heterogenous testbed management frameworks.

Finally, in order to allow the community of networking researchers to execute cross-testbed experiments, involving heterogeneous resources, Planète team has been instrumental in federating a set of well-known testbeds through the SfaWrap, namely PlanetLab Europe, Senslab - developed in other Inria Project-teams -, FEDERICA, the outcome of another E.U.-funded project and more recently NITOS, an OMF-enabled wireless testbed. See [96] and [97] for more details.

- **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 developpers 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 completly 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 scenarii, 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 and is currently under submission.

- **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. For more details see the Bake web page http://planete.inria.fr/software/bake/index.html

- **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.

- **Automated Deployment and Customization of Routing Overlays Across Heterogeneous Experimentation Platforms**

  During the last decades, many institutions and companies around the world have invested great effort into building new network experimentation platforms. These platforms range from simulators, to emulators and live testbeds, and provide very heterogeneous ways to access resources and to run experiments.

  Currently, a growing concern among platform owners is how to encourage researchers from different platform communities to take advantage of the resources they offer. However, one important aspect that needs to be overcome in order to appeal researchers to use as many experimentation platforms as necessary to best validate their results, is to decrease the inherent complexity to run experiments in different platforms. Even more so, to decrease the complexity of mixing resources from different platforms on a same experiment, to achieve the combination of resources best suited to the experiment needs.

  To address this concern, we developed the Network Experiment Programming Interface (NEPI) whose goal is to make easier the use of different experimentation platforms, and switch among them easily. 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. On 2010 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.

  During 2011, we 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. Also on 2011, we started work on a prototype PlanetLab backend.

  Last year, we extended NEPI to provide automated deployment and customization of routing overlays using resources from heterogeneous experimentation platforms. The main contribution of this work is to enable researchers to easily integrate different resources, such as simulated, emulated or physical nodes, on a same experiment, using a network overlay, thus addressing one of the main concerns previously mentioned.

We started by adding support to easily build routing overlays on PlanetLab, and providing the ability to customize network traffic by adding user defined filters to packets traversing the overlay tunnels [48]. We then improved this work by adding the ability to include simulated nodes from the ns-3 backend and emulated nodes from the linux containers backend into a single overlay network. We demonstrated the use of NEPI to build adn control routing overlays which incorporate resources from different on the ns-3 2012 community workshop [74].

- **Content Centric Networks Live Experimentation**

  Realistic experimentation on top of Internet-like environments is key to evaluate the feasibility of world wide deployment of CCNx, and to assess the impact of existing Internet traffic conditions on CCN traffic. However, deploying live experiments on the Internet is a difficult and error prone task, specially when performed manually.

  To address this issue, during the last year, we extended NEPI, a framework for managing network experiments, to support easy design, and automated deployment and control, of CCNx experiments on the PlanetLab testbed. Among other features, NEPI now enables the deployment of user modified CCNx sources on arbitrary PlanetLab nodes, and the creation of tunnels to enable the use of multicast FIB entries between CCNx daemons over the Internet. By supporting easy CCNx experimentation on PlanetLab, NEPI can help to explore the co-existence of CCN and TCP/IP architecture.

  This work was presented as a poster and a demo at CCNxCon 2012, the CCNx http://www.ccnx.org/ community meeting [73]. The work had a very good reception and gained NEPI some new users.

  An online tutorial and demo were also made available at NEPI's web page http://nepi.inria.fr/wiki/ nepi/CCNxOnPlanetLabEurope, for dissemination purposes.

- **Smooth-transition: a new methodology for dealing with various network experiment environments**

  The smooth-transition is a new methodology, which supports various network experiment environments covering from pure simulation through realistic emulation consistently. The reproducibility in experimental network research is getting important feature for iterative experiments in short-term and long-term period. The main idea of this concept is providing the reproducibility in a broader sense. So far, we had to implement different experiments by different environment, such as simulation, application-level emulation, and link-level emulation. Whereas the smooth-transition is able to keep the context of the experiments started from a pure simulation up to a realistic emulation gradually. That means the user does not need to waste time any more for learning and following a lot of documents and manuals from each different environment. Moreover, anyone can easily start to use the testbed and to develop inside (i.e. protocol stack). Because NS3 which is the most popular and powerful network simulator has been used in this concept as an experiment engine.

  The smooth-transition employees Network Experiment Programming Interface (NEPI) to conduct all functions, such as composing scenario, node deployment, experiment control, and resource management. The core of building this concept is NS3 which has Emulation (EMU) and Direct Code Execution (DCE) modules. EMU supports to use real network devices instead of NS3 MAC and PHY layer implementations. DCE is able to launch real application on top of NS3 protocol stacks. Furthermore, real Linux kernel (currently, net-next 2.6 is available) can replace NS3 Internet protocols by its advanced mode. This concept needs back-end system covering all experiment nodes. Control and Management Framework (OMF) plays an important role as a software framework to control and manage an wireless network testbed, and all messages are exchanged by Extensible Messaging and Presence Protocol (XMPP). Nitos scheduler has been adopted as a reservation system http://nitlab.inf.uth.gr/NITlab/index.php/scheduler. The user can reserve a time slot, nodes,

and wireless channels through its web page. In addition, SFA supports that the testbed is federated with other ones of outside.

The testbed provides PCAP files as a common outcome, and this file contains captured in and out packets. However, the file size is easily over gigabytes, then it makes a very long delay to process dozens of that files. To reduce the processing time efficiently, we are using an indexing scheme for fast collecting desired packets by filtering. In particular, this scheme is very useful to find packets occurred rarely, when an detailed analysis is required for an network event, such as retransmission, intrusion detection, and node association/disassociation. The indexing information is stored in a database file, and it does not need to be modified after making the file. The size of the file is very small compared with the PCAP file, so it provides fast packet filtering permanently, even after leaving the testbed. This work, post-processing of PCAP files, is in a collaboration with Diego Dujovne and Luciano Ahumada from the Universidad Diego Portales of Chili. Especially, YoungHwan Kim, a postdoc of the Planète group, has been currently dispatched for this collaboration for fourteen weeks (September 15 2012 ∼ January 26 2013) in Santiago, Chile.

- **The FIT experimental platform**

We have started, since 2011, the procedure of building a new experimental platform at Sophia-Antipolis, in the context of the FIT Equipment of Excellence project. This platform has two main goals : the first one is to enable highly controllable experiments due to its anechoic environment. These experiments can be either hybrid-experiments (as NEPI will be deployed) or federated-experiments through several testbeds. The second goal is to make resource consuming experiments (like CCNx) possible due to some powerful servers that will be installed and connected to the PlanetLab testbed. During 2012, the specifications has been defined and the procedure will continue during the next year.

- **Network Simulations on a Grid**

We studied an hybrid approach for the evaluation of networking protocols based on the ns-3 network simulator and a Grid testbed. We analyzed the performance of the approach using a simple use case. Our evaluation shows that the scalability of our approach is mainly limited by the processor speed and memory capacities of the simulation node. We showed that by exploiting the emulation capacity of ns-3, it is possible to map complex network scenarios on grid nodes. We also proposed a basic mapping algorithm to distribute a network scenario on several node [32].

REVES Project-Team

5. New Results

5.1. Plausible Image Rendering

5.1.1. Rich Intrinsic Image Decomposition of Outdoor Scenes from Multiple Views

Participants: Pierre-Yves Laffont, Adrien Bousseau, George Drettakis.

Intrinsic image decomposition aims at separating photographs into independent reflectance and illumination layers. We show that this ill-posed problem can be solved by using multiple views of the scene from which we derive additional constraints on the decomposition.

Our first method uses pictures from multiple views at a *single time of day* to automatically reconstruct a 3D point cloud of an outdoor scene. Although this point cloud is sparse and incomplete, it is sufficient to compute plausible sky and indirect illumination at each oriented 3D point, given an environment map that represents incoming distant radiance. We introduce an optimization method to estimate sun visibility over the point cloud, which compensates for the lack of accurate geometry and allows the extraction of precise cast shadows. We finally use image-guided propagation algorithms to propagate the illumination computed over the sparse point cloud to every pixel, and to separate the illumination into distinct sun, sky, and indirect components. This *rich intrinsic image decomposition* enables advanced image manipulations, illustrated in Figure 3 .

This work has led to the RID software (Section 4.1) and to a technology transfer agreement with Autodesk (Section 6.1.1.1). A paper will be published in the IEEE Transactions on Visualization and Computer Graphics journal [18] (in press). It has also been presented at SIGGRAPH 2012 in the Poster and Talk sessions [22].
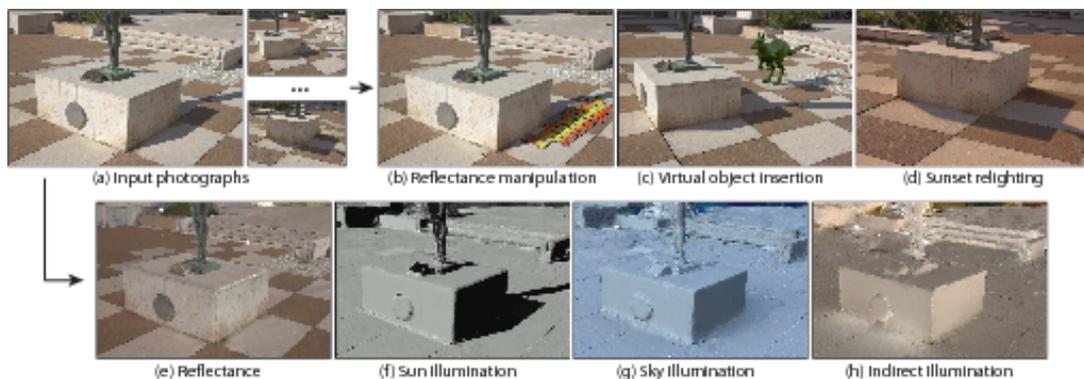
Figure 3. Starting from multiple views of the scene (a), our method decomposes photographs into four intrinsic layers — the reflectance (e), the illumination due to sun (f), the illumination due to sky (g) and the indirect illumination (h). Each layer can then be manipulated independently for advanced image editing applications (b-d).

5.1.2. Coherent Intrinsic Images from Photo Collections

Participants: Pierre-Yves Laffont, Adrien Bousseau, George Drettakis.

We propose a second method to compute intrinsic images in the presence of varying lighting conditions. Our method exploits the rich information provided by *multiple viewpoints and illuminations* in an image collection to process complex scenes without user assistance, nor precise and complete geometry. Such collections can be gathered from photo-sharing websites, or captured indoors with a light source which is moved around the scene.

We use multi-view stereo to automatically reconstruct 3D points and normals, from which we derive relationships between reflectance values at different locations, across multiple views, and consequently across different lighting conditions. In addition, we propose an optimization approach which enforces coherent reflectance in all views of a scene.

The resulting *coherent intrinsic images* enable image-based illumination transfer between photographs of the collection, as illustrated in Figure 4 .

This work is a collaboration with Frédo Durand (MIT) and Sylvain Paris (Adobe), and started with a visit of Pierre-Yves Laffont at MIT during Summer 2011. It has been published in the ACM Transactions on Graphics journal [19], and has been presented at SIGGRAPH Asia 2012.

(a) Input photograph (source lighting)
(b) Estimated reflectance
(c) Estimated illumination
lighting transfer
(d) Input photographs for lighting transfer (target viewpoints)
(e) Synthesized images (source lighting with target viewpoints)
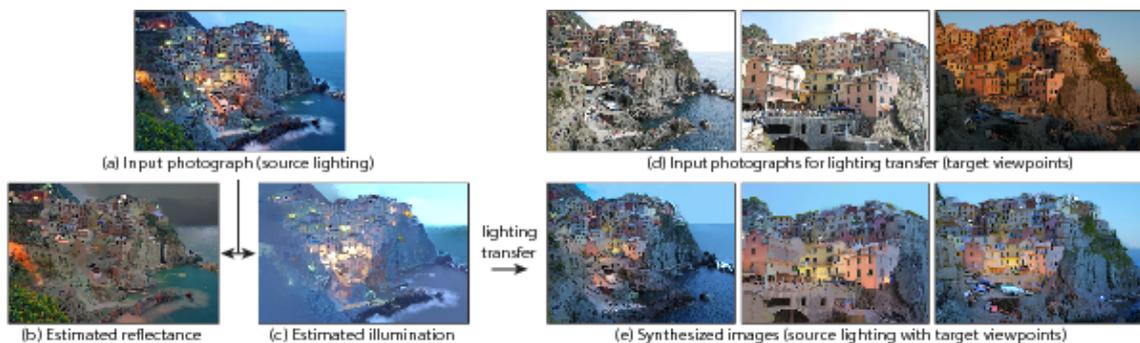
Figure 4. Our method automatically decomposes each image of a photo collection into reflectance and illumination (a-c). Transferring the illumination layer (c) to other viewpoints (d) yields synthetic images with novel viewpoint/lighting combinations (e).

5.1.3. Intrinsic Images by Clustering

Participant: Jorge Lopez Moreno.

Decomposing an input image into its intrinsic illumination and reflectance components is a long-standing ill-posed problem. We present a novel algorithm that requires no user strokes and works on a single image. Based on simple assumptions about its reflectance and luminance, we first find clusters of similar reflectance in the image, and build a linear system describing the connections and relations between them. Our assumptions are less restrictive than widely-adopted Retinex-based approaches, and can be further relaxed in conflicting situations. The resulting system is robust even in the presence of areas where our assumptions do not hold. We show a wide variety of results, including natural images, objects from the MIT dataset and texture images, along with several applications, proving the versatility of our method (see Figure 5).

This work is a collaboration with Elena Garces, Adolfo Munoz and Diego Gutierrez from University of Zaragoza (Spain). The work was published in an special issue of the journal Computer Graphics Forum and presented at the Eurographics Symposium on Rendering 2012 [16].

5.1.4. Relighting for Image Based Rendering

Participants: Sylvain Duchêne, Jorge Lopez Moreno, Stefan Popov, George Drettakis.

Figure 5. Decomposition by our method of the input image (left) into illumination (center) and reflectance (right) components.

Image-based rendering generates realistic virtual images from a small set of photographs. However, while current methods can simulate novel viewpoints from the input pictures, they cannot produce novel illumination conditions that differ from the lighting at the time of capture. The goal of this project is to provide such relighting capabilities. Our method first rely on multi-view stereo algorithms to estimate a coarse geometry of the scene. This geometry is often innacurate and incomplete. We complement it with image-based propagation algorithms that fill-in the missing data using the high-resolution input pictures. This combination of geometric and image-based cues allows us to generate plausible shadow motion and simulate novel sun directions.

5.1.5. *Depth Synthesis and Local Warps for Plausible Image-based Navigation*
Participants: Gaurav Chaurasia, Sylvain Duchêne, George Drettakis.

Modern multi-view stereo algorithms can estimate 3D geometry from a small set of unstrcutured photographs. However, the 3D reconstruction often fails on vegetation, vehicles and other complex geometry present in everyday urban scenes. We introduce a new Image-Based Rendering algorithm that is robust to unreliable geometry. Our algorithm segments the image into superpixels, *synthesizes* depth in superpixels with missing depth, warps them using a shape-preserving warp and blends them to create real-time plausible novel views for challenging target scenes, resulting in convincing immersive navigation experience.

This work is in collaboration with Dr. Olga Sorkine at ETH Zürich. and has been submitted to ACM Transactions on Graphics.

5.1.6. *Perception of Slant for Image-Based Rendering*
Participants: Christian Richardt, Peter Vangorp, George Drettakis.

Image-based rendering can create images with a high level of realism using simple geometry. However, as soon as the viewer moves away from the correct viewpoint, the image appears deformed. This work investigates the parameters which influence the perception of these image deformations. We propose a novel model of slant perception, which we validate using psychophysical experiments.

This work is a collaboration with Peter Vangorp at MPI Informatik, and Emily Cooper and Martin Banks from the University of California, Berkeley; in the context of the Associate Team CRISP (see also Section 7.3.1.1).

5.1.7. *Lightfield Editing*
Participant: Adrien Bousseau.

Lightfields capture multiple nearby views of a scene and are consolidating themselves as the successors of conventional photographs. As the field grows and evolves, the need for tools to process and manipulate lightfields arises. However, traditional image manipulation software such as Adobe Photoshop are designed to handle single views and their interfaces cannot cope with multiple views coherently. In this work we evaluate different user interface designs for lightfield editing. Our interfaces differ mainly in the way depth is presented to the user and build uppon different depth perception cues.

This work is a collaboration with Adrian Jarabo, Belen Masia and Diego Gutierrez from Universidad de Zaragoza and Fabio Pellacini from Sapienza Università di Roma.

5.1.8. Example-Based Fractured Appearance

Participants: Carles Bosch, George Drettakis.

A common weathering effect is the appearance of cracks due to material fractures. Previous exemplar-based aging and weathering methods have either reused images or sought to replicate observed patterns exactly. We propose an approach to exemplar-based modeling that creates weathered patterns by matching the statistics of fracture patterns in a photograph. We conducted a user study to determine which statistics are correlated to visual similarity and how they are perceived by the user. We describe a physically-based fracture model capable of producing similar crack patterns at interactive rates and an optimization method to determine its parameters based on key statistics of the exemplar. Our approach is able to produce a variety of fracture effects from simple crack photographs at interactive rates, as shown in Figure 6 .

Figure 6. Application of our example-based fracturing method on different scenes. Photographs of input fracture patterns are shown in the insets.

This work is a collaboration with Loeiz Glondu, Maud Marchal and George Dumont from IRISA-INSA/Inria Rennes - Bretagne Atlantique, Lien Muguercia from the University of Girona, and Holly Rushmeier from Yale University. The work was published in the Computer Graphics Forum journal and presented at the 23rd Eurographics Symposium on Rendering [17].

5.1.9. Real-Time Rendering of Rough Refraction

Participant: Adrien Bousseau.

We propose an algorithm to render objects made of transparent materials with rough surfaces in real-time, under all-frequency distant illumination. Rough surfaces cause wide scattering as light enters and exits objects, which significantly complicates the rendering of such materials. We present 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. We demonstrate the quality of our approximation by comparison against stochastic ray-tracing (see Figure 7).

Furthermore we propose two extensions to our method for supporting spatially varying roughness on object surfaces and local lighting for thin objects.

(a) Ground truth   (b) Our method

Figure 7. Compared to an expensive ray-traced reference (a), our method produces plausible results in real-time (b).

This work is a collaboration with Charles De Rousiers , Kartic Subr, Nicolas Holzschuch from Inria Grenoble, and Ravi Ramamoorthi from UC Berkeley in the context of the Associate Team CRISP (see also Section 7.3.1.1). A paper describing the method was published in the IEEE Transactions on Visualization and Computer Graphics journal [14].

5.1.10. Gabor Noise by Example

Participants: Ares Lagae, George Drettakis.

Procedural noise is a fundamental tool in Computer Graphics. However, designing noise patterns is hard. In this project, we propose *Gabor noise by example*, a method to estimate the parameters of bandwidth-quantized Gabor noise, a procedural noise function that can generate noise with an arbitrary power spectrum, from exemplar Gaussian textures, a class of textures that is completely characterized by their power spectrum (see Figure 8).

More specifically, we introduce (i) bandwidth-quantized Gabor noise, a generalization of Gabor noise to arbitrary power spectra that enables robust parameter estimation and efficient procedural evaluation; (ii) a robust parameter estimation technique for quantized-bandwidth Gabor noise, that automatically decomposes the noisy power spectrum estimate of an exemplar into a sparse sum of Gaussians using non-negative basis pursuit denoising; and (iii) an efficient procedural evaluation scheme for bandwidth-quantized Gabor noise, that uses multi-grid evaluation and importance sampling of the kernel parameters. Gabor noise by example preserves the traditional advantages of procedural noise, including a compact representation and a fast on-the-fly evaluation, and is mathematically well-founded.

This work is a collaboration with Bruno Galerne from MAP5, Université Paris Descartes and CNRS, Sorbonne Paris Cité; Ares Lagae from KU Leuven; and Sylvain Lefebvre from the ALICE project team, Inria Nancy - Grand Est. This work was presented at SIGGRAPH 2012 and published in ACM Transactions on Graphics [15].

5.1.11. Structured Gabor noise

Participants: Gaurav Chaurasia, Ares Lagae, George Drettakis.

Current procedural noise synthesis techniques [15] are limited to Gaussian random field textures. This project aims to generalize procedural noise to a broader class of structured textures.
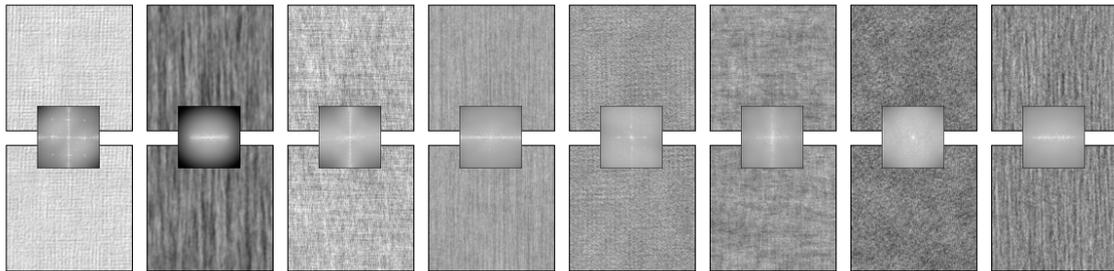
Figure 8. Gabor noise by example is a method to estimate the parameters of bandwidth-quantized Gabor noise, a procedural noise function that can generate noise with an arbitrary power spectrum, from exemplar Gaussian textures, a class of textures that is completely characterized by their power spectrum. (row 1) Gaussian texture. (row 2) Procedural noise. (insets) Estimated power spectrum.

This work is in collaboration with Dr. Ares Lagae (Katholieke Universiteit Leuven, Belgium), Dr. Bruno Galerne (Université Paris Descartes) and Prof. Ravi Ramamoorthi (UC Berkeley), in the contect of the Associate Team CRISP (Section 7.3.1.1).

5.1.12. Gloss Perception in Painterly and Cartoon Rendering

Participant: Adrien Bousseau.

Depictions with traditional media such as painting and drawing represent scene content in a stylized manner. It is unclear however how well stylized images depict scene properties like shape, material and lighting. In this project, we use non photorealistic rendering algorithms to evaluate how stylization alters the perception of gloss (see Figure 9). Our study reveals a compression of the range of representable gloss in stylized images so that shiny materials appear more diffuse in painterly rendering, while diffuse materials appear shinier in cartoon images.

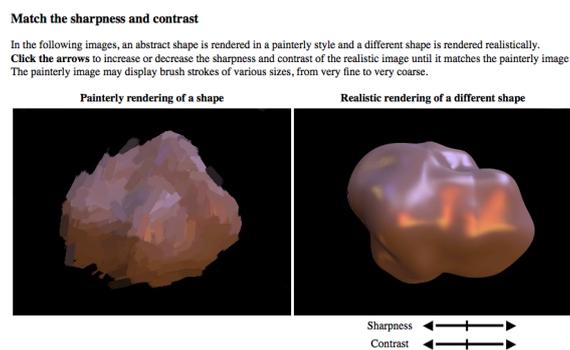*Figure 9. The experimental task used for studying gloss perception in stylized images.*

From our measurements we estimate the function that maps realistic gloss parameters to their perception in a stylized rendering. This mapping allows users of NPR algorithms to predict the perception of gloss in their images. The inverse of this function exaggerates gloss properties to make the contrast between materials in

a stylized image more faithful. We have conducted our experiment both in a lab and on a crowdsourcing website. While crowdsourcing allows us to quickly design our pilot study, a lab experiment provides more control on how subjects perform the task. We provide a detailed comparison of the results obtained with the two approaches and discuss their advantages and drawbacks for studies like ours.

This work is a collaboration with James O'Shea, Ravi Ramamoorthi and Maneesh Agrawala from UC Berkeley in the context of the Associate Team CRISP (see also Section 7.3.1.1) and Frédo Durand from MIT. It will be published in ACM Transactions on Graphics 2013 [12] (in press).

5.2. Interaction and Design for Audiovisual Virtual Environments

5.2.1. *Auditory-visual integration of emotional signals in a virtual environment for cynophobia*
Participants: Emmanuelle Chapoulie, Adrien David, Rachid Guerchouche, George Drettakis.

Cynophobia (dog phobia) has both visual and auditory relevant components. In order to investigate the efficacy of virtual reality exposure-based treatment for cynophobia, we studied the efficiency of auditory-visual environments in generating presence and emotion. We conducted an evaluation test with healthy participants sensitive to cynophobia in order to assess the capacity of auditory-visual virtual environments to generate fear reactions. Our application involves both high fidelity visual stimulation displayed in an immersive space and 3D sound. This specificity enables us to present and spatially manipulate fearful stimuli in the auditory modality, the visual modality and both.

We conducted a study where participants were presented with virtual dogs in realistic environments. Dogs were presented in a progressive manner, from unimodal and static to audiovisual and dynamic. Participants were also submitted a Behavioral Assessment Test at the beginning and end of the experiment where they were presented a virtual dog walking towards them step by step until it was extremely close. Finally, they completed several questionnaires and were asked to comment on their experience. The participants reported higher anxiety levels in response to auditory-visual stimuli compared to unimodal stimuli. Our results strongly suggest that manipulating auditory-visual integration might be a good way to modulate affective reactions and that auditory-visual VR are a promising tool for the treatment of cynophobia.

This work is a collaboration with Marine TAFFOU and Isabelle VIAUD-DELMON from IRCAM, in the context of ARC NIEVE (see also Section 7.1.4). The work was published in the Annual Review of Cybertherapy and Telemedicine in 2012.

5.2.2. *Procedural audio modeling for particle-based environmental effects*
Participants: Charles Verron, George Drettakis.

In this project we proposed a sound synthesizer dedicated to particle-based environmental effects, for use in interactive virtual environments. The synthesis engine is based on five physically-inspired basic elements which we call sound atoms, that can be parameterized and stochastically distributed in time and space. Based on this set of atomic elements, models are presented for reproducing several environmental sound sources. Compared to pre-recorded sound samples, procedural synthesis provides extra flexibility to manipulate and control the sound source properties with physically-inspired parameters. The controls are used to simultaneously modify particle-based graphical models, resulting in synchronous audio/graphics environmental effects. The approach is illustrated with three models, that are commonly used in video games: fire, wind, and rain. The physically-inspired controls simultaneously drive graphical parameters (e.g., distribution of particles, average particles velocity) and sound parameters (e.g., distribution of sound atoms, spectral modifications) as illustrated on Figure 10 for fire. The joint audio/graphics control results in a tightly-coupled interaction between the two modalities that enhances the naturalness of the scene.

The work was presented at the 133rd AES convention in October 2012 [23].

5.2.3. *Perception of crowd sounds*
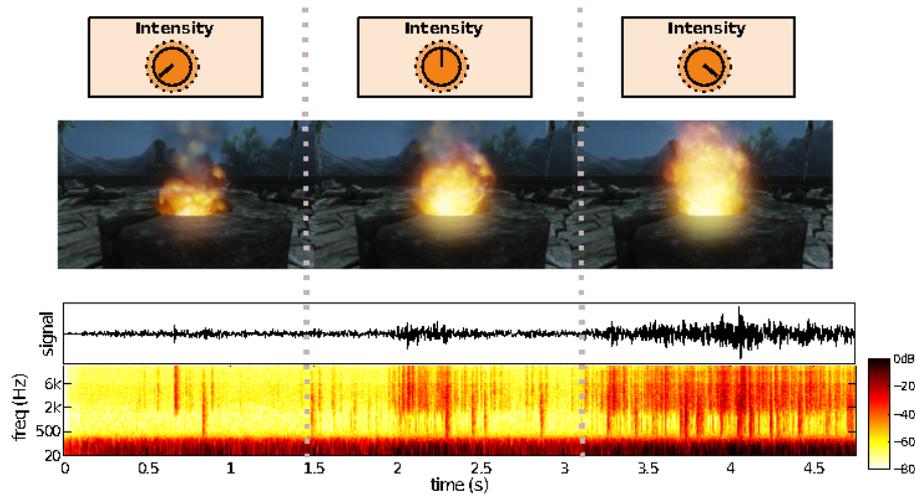**Participants:** Charles Verron, George Drettakis.

Figure 10. Audio/graphics high-level control of a fire. The control Intensity changes the rate and gain of noisy impacts, and the combustion noise of the fire sound model. Simultaneously, it controls the flame/smoke particle spawn rate for the graphics simulation.

Simulating realistic crowd scenes is an important challenge for virtual reality and games. Motion capture techniques allow to reproduce efficiently characters that look, move and sound realistic in virtual environments. However a huge amount of data is required to ensure that all agents behave differently in a big crowd. A common approach to solve this issue is to "clone" the same appearance, motion or sound several times, which can lead to perceived repetitions and break the realism of the scene. In this study we further investigate our perception of crowd scenes. Using a database of motions and sounds captured for 40 actors, along with a database of 40 different appearance templates, we propose an experimental framework to evaluate the perceptual degradations caused by clones. A particular attention is given to evaluate the influence of appearance, motion and sound, either separately or in multimodal conditions. This study aims at providing useful insights on our perception of crowd scenes, and guidelines to designers in order to reduce the amount of resources to produce convincing crowd scenes.

This ongoing project is a collaboration between Inria, CNRS-LMA (Marseille, France) and Trinity College (Dublin, Ireland).

5.2.4. *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). Locomotion techniques for such restricted workspaces should satisfy 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 have proposed three novel locomotion techniques that attempt to satisfy these goals in innovative ways. We constrain traditional Wand locomotion by turning off the Wand controls for directions that can be reached by real walking instead, and we display warning signs when the user approaches the limits of the real workspace (Figure 11 (a)). We also extend the Magic Barrier Tape paradigm with "blinders" to avoid rotation towards the

missing back screen (Figure 11 (b)). Finally, we introduce the "Virtual Companion", which uses a small bird to guide the user through virtual environments larger than the physical space (Figure 11 (c,d)).

We evaluate the three new techniques through a user study with travel-to-target and path following tasks. The study provides insight into the relative strengths of each new technique for the three aforementioned goals. Specifically, if speed and accuracy are paramount, traditional controller interfaces augmented with our novel warning techniques may be more appropriate; if physical walking is more important, two of our paradigms, the extended Magic Barrier Tape and the Constrained Wand, should be preferred; and finally, fun and ecological criteria would favor the Virtual Companion.
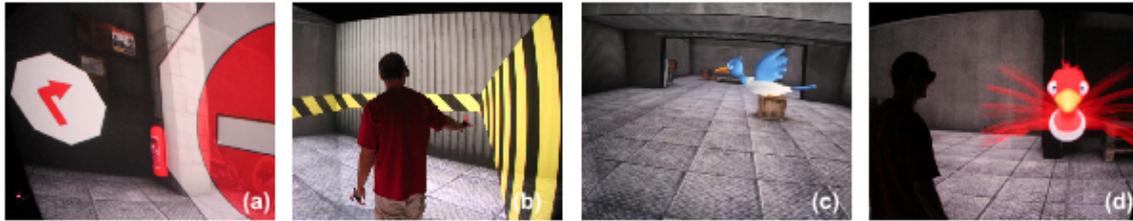

Figure 11. Screenshots illustrating the three novel locomotion techniques. From left to right: (a) Constrained Wand and signs: the "no-way" and "turn right" signs. (b) Extended Magic Barrier Tape: the tape and blinders. (c,d) Virtual Companion: the bird in "rest mode" (c) and "protection mode" (d).

This work is a collaboration with Gabriel Cirio, Maud Marchal and Anatole Lécuyer (VR4I project team, IRISA-INSA/Inria Rennes - Bretagne Atlantique) in the context of ARC NIEVE (see Section 7.1.4). The work was published in the special issue of the journal IEEE Transactions on Visualization and Computer Graphics (TVCG) [13], and presented at the IEEE Virtual Reality conference 2012.

5.2.5. *Natural Gesture-based Interaction for Complex Tasks in an Immersive Cube*
Participants: Emmanuelle Chapoulie, Jean-Christophe Lombardo, George Drettakis.

We present a solution for natural gesture interaction in an immersive cube in which users can manipulate objects with fingers of both hands in a close-to-natural manner for moderately complex, general purpose tasks. To do this, we develop a solution using finger tracking coupled with a real-time physics engine, combined with a comprehensive approach for hand gestures, which is robust to tracker noise and simulation instabilities. To determine if our natural gestures are a feasible interface in an immersive cube, we perform an exploratory study for tasks involving the user walking in the cube while performing complex manipulations such as balancing objects. We compare gestures to a traditional 6-DOF Wand, and we also compare both gestures and Wand with the same task, faithfully reproduced in the real world. Users are also asked to perform a free task, allowing us to observe their perceived level of presence in the scene. Our results show that our robust approach provides a feasible natural gesture interface for immersive cube-like environments and is perceived by users as being closer to the real experience compared to the Wand.

This work is a collaboration with Evanthia Dimara and Maria Roussou from the University of Athens and with Maud Marchal from IRISA-INSA/Inria Rennes - Bretagne Atlantique. The work has been submitted to 3DUI 2013.

5.2.6. *CrossShade: Shading Concept Sketches Using Cross-Section Curves*
Participant: Adrien Bousseau.

We facilitate the creation of 3D-looking shaded production drawings from concept sketches. The key to our approach is a class of commonly used construction curves known as cross-sections, that function as an aid to both sketch creation and viewer understanding of the depicted 3D shape. In particular, intersections of these curves, or cross-hairs, convey valuable 3D information, that viewers compose into a mental model of the overall sketch. We use the artist-drawn cross-sections to automatically infer the 3D normals across the sketch, enabling 3D-like rendering (see Figure 12).

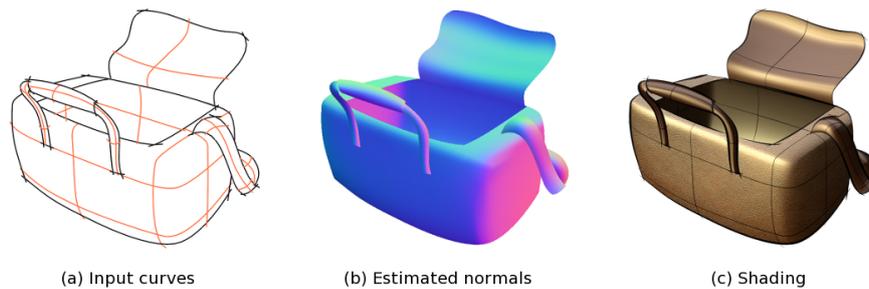(a) Input curves   (b) Estimated normals   (c) Shading

Figure 12. Concept sketches (a) frequently use cross-sections (drawn in orange) to convey 3D shape with just a handful of lines. We derive the mathematical properties of cross-section curves and leverage them to automatically estimate surface normals across the drawn objects (b). The resulting normal field allow users to shade the objects using a variety of shading styles and setups (c).

The technical contribution of our work is twofold. First, we distill artistic guidelines for drawing cross-sections and insights from perception literature to introduce an explicit mathematical formulation of the relationships between cross-section curves and the ge- ometry they aim to convey. We then use these relationships to develop an algorithm for estimating a normal field from cross-section curve networks and other curves present in concept sketches. We validate our formulation and algorithm through a user study and a ground truth normal comparison. These contributions enable us to shade a wide range of concept sketches with a variety of rendering styles.

This work is a collaboration with Cloud Shao and Karan Singh from the University of Toronto and Alla Sheffer from the University of British Columbia. It has been published at ACM Transactions on Graphics, proceedings of the SIGGRAPH 2012 conference.

5.2.7. CrossShape

Participant: Adrien Bousseau.

We facilitate the automatic creation of surfaced 3D models from design sketches that employ a commonly drawn network of cross-section curves. Our previous method generates 3D renderings of input sketches by creating a 3D surface normal field that interpolates the sketched cross-sections. This normal field however, incorporates the inevitable inaccuracy of sketched curves, making it inappropriate for 3D surface construction.

Successful construction of the 3D surface perceived from sketches requires cross-section properties and other perceived curve relationships such as symmetry and parallelism, to be met precisely. We present a novel formulation where these geometric constraints are satisfied while minimizing the difference between the sketch and the 3D cross-sections projected on it. We validate our approach by producing accurate surface reconstructions of existing 3D models represented using a network of cross-sections as well on a variety of sketch input. Finally we illustrate our surfacing solution within an interactive sketch based modeling framework.

This work is a collaboration with James McCrae and Karan Singh from the University of Toronto and Xu Baoxuan and Alla Sheffer from the University of British Columbia.

5.2.8. *Computer-assisted drawing*

Participants: Emmanuel Iarussi, Adrien Bousseau.

A major challenge in drawing from observation is to trust what we *see* rather than what we *know*. Drawing books and tutorials provide simple techniques to gain consciousness of the shapes that we observe and their relationships. Common techniques include drawing simple geometrical shapes first – also known as *blocking in* – and checking for alignments and equal proportions. While very effective, these techniques are usually illustrated on few examples and it takes significant effort to generalize them to an arbitrary model. In addition, books and tutorials only contain static instructions and cannot provide feedback to people willing to practice drawing.

In this project, we develop an interactive drawing tool that assists users in their practice of common drawing techniques. Our *drawing assistant* helps users to draw from any model photograph and provides corrective feedback interactively.

This work is a collaboration with Theophanis Tsandilas from the InSitu project team, Inria Saclay - Ile de France, in the context of the ANR DRAO project (see Section 7.1.2).

5.2.9. *Depicting materials in vector graphics*

Participants: Jorge Lopez-Moreno, Adrien Bousseau, Stefan Popov, George Drettakis.

Vector drawing tools like Illustrator and InkScape enjoy great popularity in illustration and design because of their flexibility, directness and distinctive look. Within such tools, skillful artists depict convincing material and lighting effects using 2D vector primitives like gradients and paths. However, it takes significant expertise to convey plausible material appearance in vector drawings. Instead, novice users often fill-in regions with a constant color, sacrificing plausibility for simplicity. In this project we present the first vector drawing tool that automates the depiction of material appearance. Users can use our tool to either fill-in regions automatically, or to generate an initial set of vector primitives that they can refine at will.

This work is a collaboration with Maneesh Agrawala from the University of Berkeley in the context of the Associate Team CRISP (see Section 7.3.1.1).

5.2.10. *Gradient Art: Creation and Vectorization (survey)*

Participant: Adrien Bousseau.

We survey the main two categories of methods for producing vector gradients. One is mainly interested in converting existing photographs into dense vector representations. By vector it is meant that one can zoom infinitely inside images, and that control values do not have to lie onto a grid but must represent subtle color gradients found in input images. The other category is tailored to the creation of images from scratch, using a sparse set of vector primitives. In this case, we still have the infinite zoom property, but also an advanced model of how space should be filled in-between primitives, since there is no input photograph to rely on. These two categories are actually extreme cases, and seem to exclude each other: a dense representation is difficult to manipulate, especially when one wants to modify topology; a sparse representation is hardly adapted to photo vectorization, especially in the presence of texture. Very few methods lie in the middle, and the ones that do require user assistance.

We published our survey in the book *Image and Video based Artistic Stylization* [25] editied by Springer. The survey was written in Collaboration with Pascal Barla from the MANAO project team, Inria Bordeaux - Sud Ouest, in the context of the ANR DRAO project (see Section 7.1.2).

<div align="center">

SCIPORT Team

</div>

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)].

Together with colleagues in Argonne National Lab. and RWTH Aachen, we are studying how AD tools can handle MPI-parallel codes, especially in adjoint mode.

This year, we have presented our strategy [16] to extend Data-Flow analysis to Message-Passing communication. This strategy is specially designed for a program representation like that of TAPENADE, i.e. based on a Call-Graph whose nodes are indeed Flow-Graphs. This representation makes it easier to implement analyses in a way that is both context-sensitive and flow-sensitive. Our strategy also relies on the fixed-point implementation of the analyses, which uses a "wait-list".

At the same time, we continue the design of a adjoint-mode AD adapted to MPI communication. In our framework of AD by source transformation, we have pushed far in the direction of static data-flow analyses and static source transformation of individual MPI calls. We obtained results on classical cases of message-passing [38]. However, experience shows [11] that general usage of message-passing defies static analysis. A purely static analysis and transformation must resort too often to conservative choices, yielding a poor efficiency.

As a consequence, we are now going in the direction of a more dynamic, run-time treatment of adjoint MPI calls. This means designing a wrapper library "AMPI" on top of MPI, that takes care during execution of the adjoint code of the bookkeeping to send the adjoint messages in the reverse direction. This wrapper library should also be independent from the particular AD tool, as it will be used not only with TAPENADE but with the tools developed at Argonne and RWTH Aachen.

6.2. Finer control on AD transformation

Participants: Valérie Pascual, Laurent Hascoët.

We explore methods to provide the AD end-user with a better control on the AD transformation. We want to organize a progressive AD process in which the end-user can choose among a set of available AD code optimizations. In a first stage, the end-user may deactivate most of these optimizations, thus obtaining a differentiated code that is easier to understand and hopefully more robust. If problems do occur, this differentiated code is easier to debug with the debugging tools that we provide. In the next stages, the end-user may progressively turn the optimizations on, and at the same time check that the derivatives remain correct.

Another goal closely related is the comparison and evaluation of the existing corpus of AD code optimizations. TAPENADE is one of the AD tools that incorporate most of AD optimizations proposed in litterature. If a few missing optimizations are included, TAPENADE with its relatively large set of validation applications can be the common ground for a credible evaluation of the benefit brought by each optimization.

In this direction, we have extended TAPENADE to turn some classical optimizations that were automatically applied into optional optimizations. The emblematic example is activity analysis. This required some code cleanup. Also, we are extending TAPENADE to give the option of "association by address" instead of "association by name". This means bundling each variable with its derivative into a structured object, instead of creating new variables with new names to hold the derivatives. Which option is better is a difficult question, related to memory locality issues. This extension will allow us to make accurate measurements on our set of validation codes. This is also a step towards a better collaboration of TAPENADE with overloading-based AD tools, that natively use association by address.

6.3. Formal specification of AD

Participant: Laurent Hascoët.

There is very little formal specification of AD as a program transformation, and consequently no formal proof of its correctness. Correctness of tangent AD is problematic: if defined as equivalence of the tangent program semantics with the mathematical derivative of the semantics of the original code, correctness is mostly granted for simple staight-line programs, and in general not granted for programs with control. Therefore formal proofs of correctness appear unreachable at present. Fortunately, there is little concern about the practical relevance of tangent AD. The confidence of end-users regarding tangent AD is justified by everyday experience.

Adjoint-mode AD poses a different challenge. The adjoint AD transformation is by no means simple nor intuitive. Its specification is informal, so that end-users of AD cannot gain a strong confidence in the process. Moreover, the constant quest for efficiency of the adjoint code has introduced a number of improvements and tradeoffs that are defined informally. These improvements make the adjoint code intricate and sometimes interact to cause subtle bugs. On the other hand, the good news is that the difference between the adjoint code and the tangent code only lies in the order of the derivatives computations and not in their nature. A formal proof of semantic equivalence is thus conceivable.

The first step towards such a proof is a formal specification of both tangent-mode and adjoint-mode AD, including the specification of the program static data-flow analyses that the transformations require. We have provided this specification in terms of Data-Flow equations for the analyses, and in terms of Structural Operational Semantics (more precisely Natural Semantics) for the AD transformations themselves [19]. This specification will be the basis for future formal proofs of equivalence between tangent AD and adjoint AD.

6.4. Resolution of linearised systems

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

Increased sophistication of solution algorithms pose a challenge to Automatic Differentiation. Time-stepping iterations create numerous updates of the iterated solution vector. Other additional nonlinear iterative processes occur such as:

- 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).

Adjoint-mode AD applied to these algorithms produces a "linearised iterative algorithm" which is transposed and therefore follows the original iterations in the reverse order, needing each of the iterated state solution vectors. One such extreme case is the simulation of unsteady phenomena with implicit numerical schemes: simulating high Reynolds turbulent flows by a Large Eddy Simulation (LES) and RANS-LES models requires hundreds of thousands time steps, each of them involving a modern iterative solution algorithm. This is the case targeted by the 4-year ANR project "ECINADS", jointly with university of Montpellier 2, the Institut de Mécanique des fluides de Toulouse and Lemma company, started in 2009.

In ECINADS, we design more efficient solution algorithms and we examine the questions risen by their adjoint differentiation. Our goal is practical scalability of the direct simulation and of its adjoint on a large number of processors. ECINADS also addresses the scalable solution of new approximations.

In 2012, the novel three-level method studied by H. Alcin on a model problem has been extended to compressible viscous flows by B. Koobus and C. Moussaed from university of Montpellier.

Hubert Alcin, Bruno Koobus, Olivier Allain and Alain Dervieux published their work on a two-level Schwarz algorithm in IJNMFD [12]. H. Alcin has presented his work in the Parallel CFD conference of Altlanta [14]. H. Alcin wrote his thesis [11], defended in december, on the three main subjects of ECINADS: the two- and three-level Schwarz algorithms, Automatic Differentiation and mesh adaptation.

6.5. Automatic Differentiation of a CFD code

Participants: Hubert Alcin, Valérie Pascual, Laurent Hascoët, Alain Dervieux.

The ECINADS workplan includes the building of an adjoint state for a CFD kernel. We have chosen AIRONUM 5.1 , a real life kernel that combines two particular features:

- it uses intensively the Fortran95 dynamic memory allocation
- it uses MPI parallelization.

This work is reported in H. Alcin's PhD [11].

6.6. Perturbation Methods

Participants: Alain Dervieux, Laurent Hascoët.

In the context of the European project NODESIM-CFD (ended 2010), the contribution of Sciport involved mainly the derivation 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, Tangent-on-Adjoint, and Adjoint-on-Tangent. These new methods are promoted through short courses, e.g. by Alain Dervieux at an ERCOFTAC session (Chatou, 15-16 mai 2012). The application and extension of these methods are part of a FP7 proposal (Proposal UMRIDA, nov. 2012).

6.7. Control of approximation errors

Participants: Frédéric Alauzet [GAMMA team, Inria-Rocquencourt], Estelle Mbinky [GAMMA team, Inria-Rocquencourt], Olivier Allain [Lemma], Alexandre Carabias, Hubert Alcin, Alain Dervieux.

This is a joint research between Inria teams Gamma (Rocquencourt), Sciport, Castor and the Lemma company. Gamma brings mesh and approximation expertise, Sciport brings adjoint methods, and CFD applications are developed by CASTOR and Lemma.

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

In 2012, this activity is amplifying. A work on goal-oriented mesh adaptation for unsteady Euler flows submitted to the journal JCP has been accepted and published [13]. Its extension to the compressible Navier-Stokes model has been developed in 2D [22] and in 3D [11]. A further extension to Large Eddy Simulation has been defined and developed in the WOLF demonstrator. A communication at ECCOMAS (Vienna) has been presented and papers are being written for publication in journal.

The method is being extended to a third-order approximation, the Vertex-CENO. This approximation was defined collaboratively between university of Montpellier, IMM-Moscow and Sciport. A more accurate version is studied by Alexandre Carabias. A new mesh adaptation theory involving error estimates and criteria has been developed by Gamma and Sciport. The extension of the multiscale adaptation method is considered by Estelle Mbinky at Rocquencourt and has been presented at ECCOMAS (Vienna). The extension of the goal-oriented method is considered by Alexandre Carabias and first results were presented at ECCOMAS (Vienna). A cooperation with CEMEF and university of Nice is considered and a ERC common proposal, CMILE, has been built. Anisotropic mesh adaptation allows for better convergence towards continuous solutions, and in particular more accurate a posteriori error estimates and correctors. The synergy between correctors and mesh adaptation is the subject of a joint contribution (Gamma and Sciport) for the FP7 UMRIDA proposal (nov. 2012).

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

6. New Results

6.1. Introduction

This year Stars has proposed new algorithms related to its three main research axes : perception for activity recognition, semantic activity recognition and software engineering for activity recognition.

6.1.1. Perception for Activity Recognition

Participants: Julien Badie, Slawomir Bak, Vasanth Bathrinarayanan, Piotr Bilinski, Bernard Boulay, François Brémond, Sorana Capalnean, Guillaume Charpiat, Duc Phu Chau, Etienne Corvée, Eben Freeman, Carolina Garate, Jihed Joober, Vaibhav Katiyar, Ratnesh Kumar, Srinidhi Mukanahallipatna, Sabine Moisan, Silviu Serban, Malik Souded, Anh Tuan Nghiem, Monique Thonnat, Sofia Zaidenberg.

This year Stars has extended an efficient algorithm for detecting people. We have also proposed a new algorithm for re-identification of people through a camera network. We have realized a new algorithm for the recognition of short actions and validated also its performance on several benchmarking databases (e.g. ADL). We have improved a generic event recognition algorithm by handling event uncertainty at several processing levels. More precisely, the new results for perception for activity recognition concern:

- Image Compression and Modelization (6.2)
- Background Subtraction (6.3)
- Fiber Based Video Segmentation (6.4)
- Enforcement of Monotonous Shape Growth/Shrinkage in Video Segmentation (6.5)
- Dynamic and Robust Object Tracking in a Single Camera View (6.6)
- Optimized Cascade of Classifiers for People Detection Using Covariance Features (6.7)
- Learning to Match Appearances by Correlations in a Covariance Metric Space (6.8)
- Recovering Tracking Errors with Human Re-identification (6.9)
- Human Action Recognition in Videos (6.10)
- Group Interaction and Group Tracking for Video-surveillance in Underground Railway Stations (6.11)
- Crowd Event Monitoring Using Texture and Motion Analysis (6.12)
- Detecting Falling People (6.13)
- People Detection Framework (6.14)

6.1.2. Semantic Activity Recognition

Participants: Sorana Capalnean, Guillaume Charpiat, Cintia Corti, Carlos -Fernando Crispim Junior, Hervé Falciani, Baptiste Fosty, Qioa Ma, Firat Ozemir, Jose-Luis Patino Vilchis, Guido-Tomas Pusiol, Rim Romdhame, Bertrand Simon, Abhineshwar Tomar.

Concerning semantic activity recognition, the contributions are :

- A Model-based Framework for Activity Recognition of Older People using Multiple sensors (6.15)
- Activity Recognition for Older People using Kinect (6.16)
- Descriptors of Depth-Camera Videos for Alzheimer Symptom Detection (6.17)
- Online Activity Learning from Subway Surveillance Videos (6.18)
- Automatic Activity Detection Modeling and Recognition: ADMR (6.19)

6.1.3. Software Engineering for Activity Recognition

Participants: François Brémond, Daniel Gaffé, Julien Gueytat, Baptiste Fosty, Sabine Moisan, Anh tuan Nghiem, Annie Ressouche, Jean-Paul Rigault, Leonardo Rocha, Luis-Emiliano Sanchez, Swaminathan Sankaranarayanan.

This year Stars has continued the development of the SUP platform. This latter is the backbone of the team experiments to implement the new algorithms. We continue to improve our meta-modelling approach to support the development of video surveillance applications based on SUP. This year we have focused on an architecture for run time adaptation and on metrics to drive dynamic architecture changes. We continue the development of a scenario analysis module (SAM) relying on formal methods to support activity recognition in SUP platform. We improve the theoretical foundations of CLEM toolkit and we rely on it to build SAM. Finally, we are improving the way we perform adaptation in the definition of a multiple services for device adaptive platform for scenario recognition.

The contributions for this research axis are:

- SUP Software Platform (6.20)
- Qualitative Evaluation of Detection and Tracking Performance (6.21)
- Model-Driven Engineering and Video-surveillance (6.22)
- Synchronous Modelling and Activity Recognition (6.23)

6.2. Image Compression and Modelization

Participants: Guillaume Charpiat, Eben Freeman.

Recent results in statistical learning have established the best strategy to combine several advices from different experts, for the problem of sequential prediction of times series. The notions of prediction and compression are tightly linked, in that a good predictor can be turned into a good compressor via entropy coding (such as Huffman coding or arithmetic coding), based on the predicted probabilities of the events to come : the more predictable an event E is, the easier to compress it will be, with coding cost $-\log(p(E))$ with such techniques.

The initial idea here, by Yann Ollivier (TAO team), within a collaboration with G. Charpiat and Jamal Atif (TAO team), was to adapt these results to the case of image compression, where time series are replaced with 2D series of pixel colors, and where experts are predictors of the color of a pixel given the colors of neighbors. The main difference is that there is no canonical physically-relevant 1D ordering of the pixels in an image, so that a sequential order (of the pixels to predict their colors) had to be defined first. Preliminary results with a hierarchical ordering scheme already competed with standard techniques in lossless compression (png, lossless jpeg2000).

During his internship in the Stars team, Eben Freeman developed this approach, by building relevant experts able to predict a variety of image features (regions of homogeneous color, edges, noise, . . .). We also considered random orderings of pixels, using kernels to express probabilities in a spatially-coherent manner. Using such modellings of images with experts, we were also able to generate new images, that are typical of these models, and show more structure than the ones associated to standard compression schemes (typical images highly compressed).

6.3. Background Subtraction

Participants: Vasanth Bathrinarayanan, Anh-Tuan Nghiem, Duc-Phu CHAU, François Brémond.

Keywords: Gaussian Mixture Model, Shadow removal, Parameter controller, Codebook model, Context based information

6.3.1. Statistical Background Subtraction for Video Surveillance Platform

Anh-Tuan Nghiem work on background subtraction is an extended version of Gaussian Mixture Models [73]. The algorithm compares each pixel of current frame to background representation which is developed based on the pixel information from previous frames. It includes shadow and highlight removal to give better results. Selective background updating method based on the feedback from the object detection helps to better model background and remove noise and ghosts.

Figure 10 shows a sample illustration of the output of the background subtraction, where blue are foreground pixels and red are shadow or illumination change pixels and a green bounding box is a foreground blob. Also we have compared our algorithm with few other such as OpenCV and also IDIAP's background subtraction(not tuned perfectly, used default parameters) and the results are shown in Figure 11 where the green background refers to best performance of the comparisons. This evaluation is done on PETS 2009 data-set with our obtained foreground blobs to the manually annotated bounding boxes of people.

6.3.2. Parameter controller using Contextual features

The above method has some parameters that has to be tuned every time for each video, which is a time consuming work. The work of Chau et al [59] learns the contextual information from the video and controls object tracking algorithm parameters during the run-time of the algorithm. This approach is at preliminary stage for background subtraction algorithm to automatically adapt parameters. These parameters are learned as described in the offline learning process block diagram 12 over several ground truth videos and clustered into a database. The contextual feature which are used presently include object density, occlusion, contrast, 2D area, contrast variance, 2D area variance. Figure 13 shows a sample of video chunks based on contextual feature similarity for a video from caviar data-set.

The controller's preliminary results are promising and we are experimenting and evaluating with different features to learn the parameters. The results will be published in upcoming top computer vision conferences.

Figure 10. Sample illustration of output of EGMM

6.4. Fiber Based Video Segmentation

Participants: Ratnesh Kumar, Guillaume Charpiat, Monique Thonnat.

Keywords: Video Volume, Fibers, Trajectory

PETS dataset Evaluation on Background subtraction blob with GT of people--			
Metrics	**IDIAP**	**SUP**	**OPENCV**
Metric M1.1			
Global results:			
Number of True Positives	3761	4046	4057
Number of False Positives	1807	317	1099
Number of False Negatives 0	1070	793	774
Precision (mean by frame)	0.71	0.93	0.81
Sensitivity 0 (mean by frame)	0.79	0.84	0.85
Precision (global)	0.68	0.93	0.79
Sensitivity 0 (global)	0.78	0.84	0.84

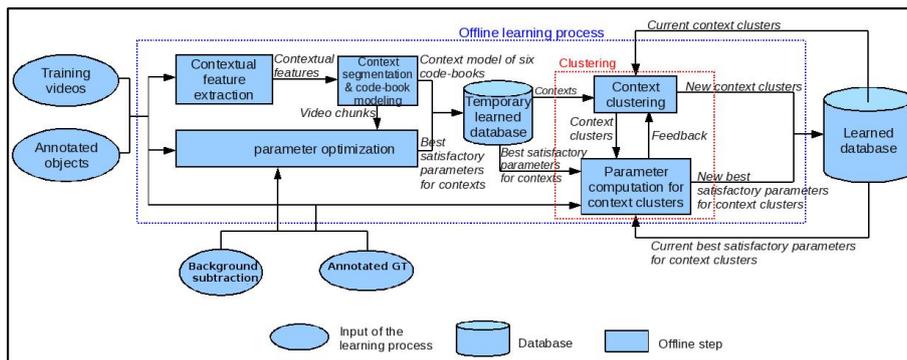*Figure 11. Evaluation with some background subtraction algorithms*

Figure 12. Block representation of the offline learning scheme to build a database
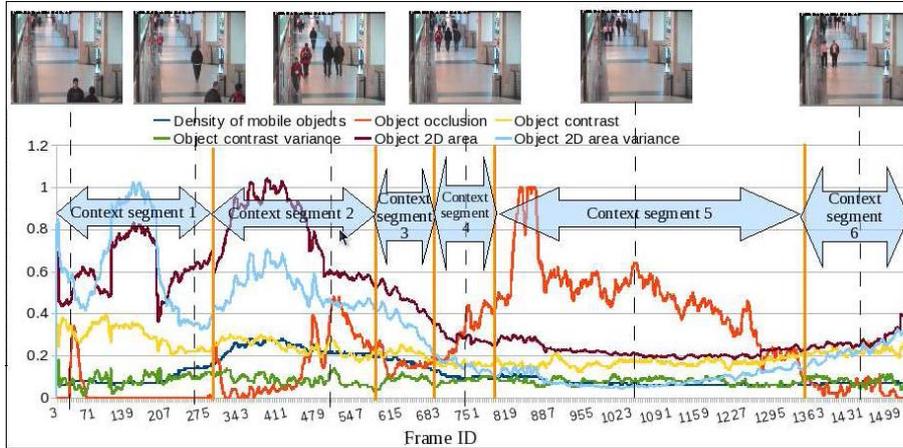
Figure 13. Context segmentation of the sequence ThreePastShop2cor (belonging to the Caviar dataset). The context segments are separated by the vertical orange lines. The control parameters are then learned for each context segment

The aim of this work is to segment objects in videos by considering videos as 3D volumetric data (2D×time). Figure 14 shows an input video and its corresponding partition in terms of fiber at a particular hierarchy level. Particularly, it 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. Analogically to the fibers in MRI images of human brains, we term *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.

Initial fibers are built using correspondences computing algorithms like optical flow and descriptor matching. As these algorithms are reliable near corners and edges, we build fibers at these locations for a video. Our subsequent goal is to partition this video in terms of fibers built, by extending them (both spatially and temporally) to the rest of the video.

To extend fibers, we compute geodesics from pixels (not belonging to the initially built fibers) to fibers. For a reliable extension, the cost of moving along a geodesic is proportional to the trajectory similarity of a pixel *wrt* a fiber, wherein a pixel trajectory is similar to the fiber trajectory. This cost function quantifies the color homogeneity of a pixel trajectory along with its color similarity *wrt* a fiber. A pixel is then associated to a fiber for which this cost is minimum.

With the above mentioned steps we obtain a partition of a video in terms of fibers wherein we have a trajectory associated with each pixel. This hierarchical partition provides a mid-level representation of a video, which can be seen as a facilitator or a pre-processing step towards higher level video understanding systems *eg* activity recognition.

6.5. Enforcement of Monotonous Shape Growth/Shrinkage in Video Segmentation

Participant: Guillaume Charpiat.

keywords: graph cuts, video segmentation, shape growth

*Figure 14. **Left:** Input Video and Spatio-Temporal Slices. **Right:** Segmented Results at a Particular Hierarchy Level*

The segmentation of noisy videos or time series is a difficult problem, not to say an impossible or ill-posed task when the noise level is very high. While individual frames can be analysed independently, time coherence in image sequences provides a lot of information not available for a single image. Most of the state-of-art works explored short-term temporal continuity for object segmentation in image sequences, *i.e.*, each next frame is segmented by using information from one or several images at previous time points. It is, however, more advantageous to simultaneously segment many frames in the data set, so that segmentation of the entire image set supports each of the individual segmentations.

In this work, we focus on segmenting shapes in image sequences which only grow or shrink in time, and on making use of this knowledge as a constraint to help the segmentation process. Examples of growing shapes are forest fires in satellite images and organ development in medical imaging. We propose a segmentation framework based on graph cuts for the joint segmentation of a multi-dimensional image set. By minimizing an energy computed on the resulting spatio-temporal graph of the image sequence, the proposed method yields a *globally optimal solution*, and runs in practice in linear complexity in the total number of pixels.

Two applications are performed. First, with Yuliya Tarabalka (Ayin team), we segment multiyear sea ice floes in a set of satellite images acquired through different satellite sensors, after rigid alignment (see Figure 15). The method returns accurate melting profiles of sea ice, which is important for building climate models. The second application, with Bjoern Menze (ETH Zurich, also MIT and collaborator of Asclepios team), deals with the segmentation of brain tumors from longitudinal sets of multimodal MRI volumes. In this task we impose an additional inter-modal inclusion constraint for joint segmentation of different image sequences, finally also returning highly sensitive time-volume plots of tumor growth.

6.6. Dynamic and Robust Object Tracking in a Single Camera View

Participants: Duc-Phu Chau, Julien Badie, François Brémond, Monique Thonnat.

Keywords: Object tracking, online parameter tuning, controller, self-adaptation and machine learning
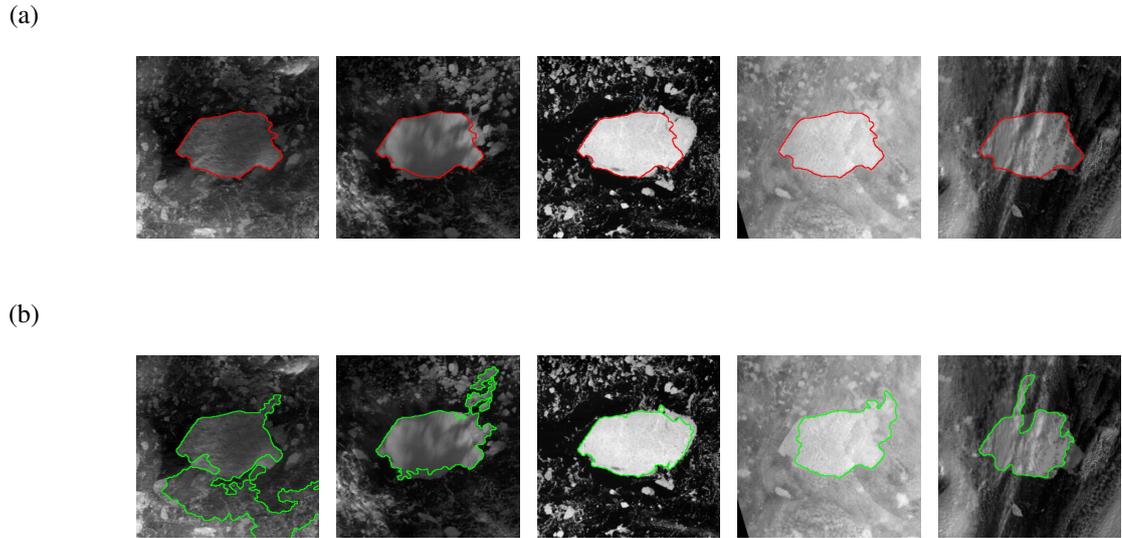
(a)

(b)

Figure 15. (a) Aligned satellite images captured each four days superposed with segmentation contours computed by our approach. (b) Segmentation contours for images (a) obtained by applying graph cut segmentation to each image at a single time moment. Note that the segmentations (a) are pixelwise precise, and that the white regions surrounding sometimes the boundaries are other ice blocks, agglomerating temporarily only, thus correctly labelled. Hence the importance of enforcing time coherence.

Object tracking quality usually depends on video scene conditions (e.g. illumination, density of objects, object occlusion level). In order to overcome this limitation, we present a new control approach to adapt the object tracking process to the scene condition variations. The proposed approach is composed of two tasks.

The objective of the first task is to select a convenient tracker for each mobile object among a Kanade-Lucas-Tomasi-based (KLT) tracker and a discriminative appearance-based tracker. The KLT feature tracker is used to decide whether an object is correctly detected. For badly detected objects, the KLT feature tracking is performed to correct object detection. A decision task is then performed using a Dynamic Bayesian Network (DBN) to select the best tracker among the discriminative appearance and KLT trackers.

The objective of the second task is to tune online the tracker parameters to cope with the tracking context variations. The tracking context, or context, of a video sequence is defined as a set of six features: density of mobile objects, their occlusion level, their contrast with regard to the surrounding background, their contrast variance, their 2D area and their 2D area variance. Each contextual feature is represented by a code-book model. In an offline phase, training video sequences are classified by clustering their contextual features. Each context cluster is then associated with satisfactory tracking parameters. In the online control phase, once a context change is detected, the tracking parameters are tuned using the learned values. This work has been published in [29], [35].

We have tested the proposed approach on several public datasets such as Caviar and PETS. Figure 16 illustrates the results of the object detection correction using the KLT feature tracker.

Figure 17 illustrates the tracking output for a Caviar video (on the left image) and for a PETS video (on the right image). The experimental results show that our method gets the best performance compared to some recent state of the art trackers.

Table 1 presents the tracking results for 20 videos from the Caviar dataset. The proposed approach obtains the best MT value (i.e. mostly tracked trajectories) compared to some recent state of the art trackers.

Figure 16. Illustration of the object detection correction for a Caviar video. The green bounding box is the output of the object detection process. The red bounding boxes are the results of the detection correction task.

a.   b.

Figure 17. Tracking results for Caviar and PETS videos

Table 1. Tracking results on the Caviar dataset. MT: Mostly tracked trajectories, higher is better. PT: Partially tracked trajectories. ML: Most lost trajectories, lower is better. The best values are printed bold.

Method	MT (%)	PT (%)	ML (%)
Zhang et al., CVPR 2008 [89]	85.7	10.7	3.6
Li et al., CVPR 2009 [71]	84.6	14.0	1.4
Kuo et al., CVPR 2010 [69]	84.6	14.7	**0.7**
Proposed approach	**86.4**	10.6	3.0

Table 2 presents the tracking results of the proposed approach and three recent approaches [56], [82], [67] for a PETS video. With the proposed approach, we obtain the best values in both metrics MOTA (i.e. Multi-object tracking accuracy) and MOTP (i.e. Multi-object tracking precision). The authors in [56], [82], [67] do not present the tracking results with the MT, PT and ML metrics.

Table 2. Tracking results on the PETS sequence S2.L1, camera view 1, sequence time 12.34. MOTA: Multi-object tracking accuracy, higher is better. MOTP: Multi-object tracking precision, higher is better. The best values are printed bold.

Method	MOTA	MOTP	MT (%)	PT (%)	ML (%)
Berclaz et al., PAMI 2011 [56]	0.80	0.58	-	-	-
Shitrit et al., ICCV 2011 [82]	0.81	0.58	-	-	-
Henriques et al., ICCV 2011 [67]	0.85	0.69	-	-	-
Proposed approach	**0.86**	**0.72**	71.43	19.05	9.52

6.7. Optimized Cascade of Classifiers for People Detection Using Covariance Features

Participants: Malik Souded, François Brémond.

keywords: People detection, Covariance descriptor, LogitBoost.

We propose a new method to optimize a state of the art approach for people detection, which is based on classification on Riemannian manifolds using covariance matrices in a boosting scheme. Our approach makes training and detection faster while maintaining equivalent or better performances. This optimisation is achieved by clustering negative samples before training, providing a smaller number of cascade levels and less weak classifiers in most levels in comparison with the original approach.

Our approach is based on Tuzel et al. [86] work which was improved by Yao et al. [87]. We keep the same scheme to achieve our people detector: train a cascade of classifiers based on covariance descriptors, using a Logitboost training algorithm which was modified by Tuzel et al. to deal with the Riemannian manifolds metrics and using the operators which were presented in [75]. In fact, Covariance matrices do not belong to vector space but to the Riemannian manifold of (d x d) symmetric positive definite matrices. The trained cascade of classifiers is applied for detection after training.

We propose an additional step to speed up training and detection process. We propose to apply a clustering step on negative training dataset before training the classifiers. This clustering step is performed both in Riemannian manifold and in the vector space of mapped covariance matrices, using the operators and metrics previously cited.

The idea consists in regrouping all similar negative samples, with regard to their covariance information, into decreasing size clusters. Each classifier of the cascade is trained on one cluster, specializing this classifier for a given kind of covariance information, and then, speeding up the training step and providing shorter classifier, which accelerate its response when applied on image. In the same time, the specialization of each cascade classifier shortens the cascade too, speeding up the detection (see Figure 18 and Figure 19).

A paper describing this approach has been accepted in VISAPP 2013 conference [50].

6.8. Learning to Match Appearances by Correlations in a Covariance Metric Space

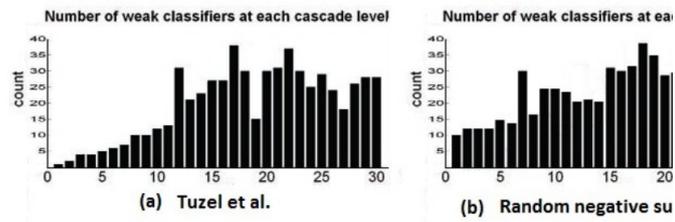**Participants:** Sławomir Bąk, Guillaume Charpiat, Etienne Corvée, Francois Brémond, Monique Thonnat.

Figure 18. Comparison of classifiers cascade structure

(a) (b) (c)

Figure 19. Cascade classifiers: some detection results

keywords: covariance matrix, re-identification, appearance matching

This work addresses the problem of appearance matching across disjoint camera views. Significant appearance changes, caused by variations in view angle, illumination and object pose, make the problem challenging.

We propose to formulate the appearance matching problem as the task of learning a model that selects the most descriptive features for a specific class of objects. Our main idea is that different regions of the object appearance ought to be matched using various strategies to obtain a distinctive representation. Extracting region-dependent features allows us to characterize the appearance of a given object class (*e.g.* class of humans) in a more efficient and informative way. Different kinds of features characterizing various regions of an object is fundamental to our appearance matching method.

We propose to model the object appearance using covariance descriptor yielding rotation and illumination invariance. Covariance descriptor has already been successfully used in the literature for appearance matching. In contrast to state of the art approaches, we do not define *a priori* feature vector for extracting covariance, but we learn which features are the most descriptive and distinctive depending on their localization in the object appearance (see figure 20). Learning is performed in a covariance metric space using an entropy-driven criterion. Characterizing a specific class of objects, we select only essential features for this class, removing irrelevant redundancy from covariance feature vectors and ensuring low computational cost.

The proposed technique has been successfully applied to the person re-identification problem, in which a human appearance has to be matched across non-overlapping cameras [34]. We demonstrated that: (1) by using different kinds of covariance features *w.r.t.* the region of an object, we obtain clear improvement in appearance matching performance; (2) our method outperforms state of the art methods in the context of pedestrian recognition on publicly available datasets (i-LIDS-119, i-LIDS-MA and i-LIDS-AA); (3) using 4×4 covariance matrices we significantly speed-up the processing time offering an efficient and distinctive representation of the object appearance.
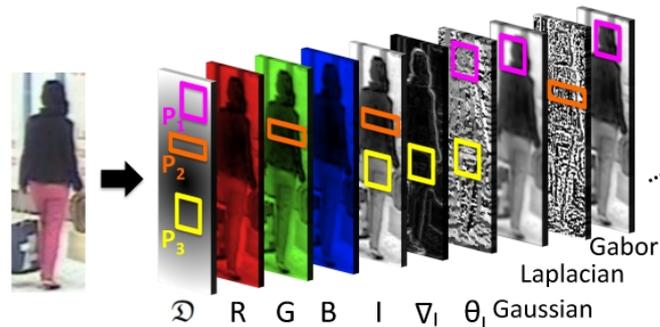
Figure 20. Example of three different covariance features. Every covariance is extracted from a region (P), distance layer (D) and three channel functions (e.g. bottom covariance feature is extracted from region P_3 using layers: D, I-intensity, ∇_I-gradient magnitude and θ_I-gradient orientation).

6.9. Recovering Tracking Errors with Human Re-identification

Participants: Julien Badie, Slawomir Bak, Duc-Phu Chau, François Brémond, Monique Thonnat.

keywords: tracking error correction, re-identification

This work addresses the problem of people tracking at long range even if the target people are lost several times by the tracking algorithm. We have identified two main reasons for tracking interruption. The first one concerns interruptions that can be quickly recovered, which includes short mis-detections, occlusions with other persons or static obstacles. The second one occurs when a person is occluded or mis-detected for a long time or when the person leaves the scene and comes back latter. Our main objective is to design a framework that can track people even if their trajectory is very segmented and/or associated with different IDs. We called this problem the global tracking challenge (see Figure 21).

Figure 21. The global tracking challenge : correcting errors due to occlusions (ID 142 on the first frame becomes 147 on the last frame) and tracking people that are leaving the scene and reentering (ID 133 on the first frame becomes 151 on the last frame).

In order to describe a person's tracklet (segment of trajectory), we use a visual signature called Mean Riemannian Covariance Grid and a discriminative method to emphasize the main differences between each tracklet. This step improves the reliability and the accuracy of the results. By computing the distance between the visual signatures, we are able to link tracklets belonging to the same person into a tracklet cluster. Only tuples of tracklets that are not overlapping each other are used as initial candidates. Then, we use Mean Shift to create the clusters.

We evaluated this method on several datasets (i-LIDS, Caviar, PETS 2009). We have shown that our approach can perform as well as the other state of the art methods on Caviar and can perform better on i-LIDS. On PETS 2009 dataset, our approach performs better than standard tracker but cannot be compared with the best state of the art methods due to unadapted metrics.

This approach is described in detail in two articles : one published in ICIP 2012 [35], which is focused on computing the covariance signature and the way to discriminate it and the other one published in PETS 2012 workshop (part of AVSS 2012 conference) [33], which is focused on the method to link the tracklets.

This work will be added to a more general tracking controller that should be able to detect several kinds of detection and tracking errors and try to correct them.

6.10. Human Action Recognition in Videos

Participants: Piotr Bilinski, François Brémond.

keywords: Action Recognition, Contextual Features, Pairwise Features, Relative Tracklets, Spatio-Temporal Interest Points, Tracklets, Head Estimation.

The goal of this work is to automatically recognize human actions and activities in diverse and realistic video settings.

Over the last few years, the bag-of-words approach has become a popular method to represent video actions. However, it only represents a global distribution of features and thus might not be discriminative enough. In particular, the bag-of-words model does not use information about: local density of features, pairwise relations among the features, relative position of features and space-time order of features. Therefore, we propose three new, higher-level feature representations that are based on commonly extracted features (e.g. spatio-temporal interest points used to evaluate the first two feature representations or tracklets used to evaluate the last approach). Our representations are designed to capture information not taken into account by the model, and thus to overcome its limitations.

In the first method, we propose new and complex contextual features that encode spatio-temporal distribution of commonly extracted features. Our feature representation captures not only global statistics of features but also local density of features, pairwise relations among the features and space-time order of local features. Using two benchmark datasets for human action recognition, we demonstrate that our representation enhances the discriminative power of commonly extracted features and improves action recognition performance, achieving 96.16% recognition rate on popular KTH action dataset and 93.33% on challenging ADL dataset. This work has been published in [36].

In the second approach, we design new representation of features encoding statistics of pairwise co-occurring local spatio-temporal features. This representation focuses on pairwise relations among the features. In particular, we introduce the geometric information to the model and associate geometric relations among the features with appearance relations among the features. Despite that local density of features and space-time order of local features are not captured, we are able to achieve similar results on the KTH dataset (96.30% recognition rate) and 82.05% recognition rate on UCF-ARG dataset. An additional advantage of this method is to reduce the processing time of training the model from one week on a PC cluster to one day. This work has been published in [37].

In the third approach, we propose a new feature representation based on point tracklets and a new head estimation algorithm. Our representation captures a global distribution of tracklets and relative positions of tracklet points according to the estimated head position. Our approach has been evaluated on three datasets, including KTH, ADL, and our locally collected Hospital dataset. This new dataset has been created in cooperation with the CHU Nice Hospital. It contains people performing daily living activities such as: standing up, sitting down, walking, reading a magazine, *etc*. Sample frames with extracted tracklets from video sequences of the ADL and Hospital datasets are illustrated on Figure 22 . Consistently, experiments show that our representation enhances the discriminative power of tracklet features and improves action recognition performance. This work has been accepted for publication in [38].

Figure 22. Sample frames with extracted tracklets from video sequences of the ADL (left column) and Hospital (right column) datasets.

6.11. 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.

Keywords: events detection, behaviour recognition, automatic video understanding, tracking

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). Analysing trajectories over a temporal window allows handling this instability more robustly. We use the event-description language described in [88] 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 behaviours such as violence or vandalism (alarming events) or a queue at the vending machine (non-alarming events).

The group tracking approach uses Mean-Shift clustering of trajectories to create groups. Two or more individuals are associated in a group if their trajectories have been clustered together by the Mean-Shift algorithm. The trajectories are given by the long-term tracker described in [60]. Each trajectory is composed of a person's positions (x, y) on the ground plane (in 3D) over the time window, and of their speed at each frame in the time window. Positions and speed are normalized using the minimum and maximum possible values (0 and $10m/s$ for the speed and the field of view of the camera for the position). The Mean-Shift algorithm requires a *tolerance* parameter which is set to 0.1, meaning that trajectories need to be distant by less than 10% of the maximum to be grouped.

Figure 23. Example of a group composed of non-similar individual trajectories.

As shown in Figure 23 , people in a group might not always have similar trajectories. For this reason, a group is also created when people are very close. A group is described by its coherence, a value calculated from the average distances of group members, their speed similarity and direction similarity. The update phase of the group uses the coherence value. A member will be kept in a group as long as the group coherence is above a threshold. This way, a member can temporarily move apart (for instance to buy a ticket at the vending machine) without being separated from the group.

This work has been applied to the benchmark CAVIAR dataset for testing, using the provided ground truth for evaluation. This dataset is composed of two parts: acted scenes in the Inria hall (9 sequences of 665 frames in average) and not acted recordings from a shopping mall corridor (7 sequences processed of 1722 frames in average). The following scenarios have been defined using the event-description language of [88]: *fighting*, *split up*, *joining*, *shop enter*, *shop exit*, *browsing*. These scenarios have been recognized in the videos with a high success rate (94%). The results of this evaluation and the above described method have been published in [45].

The group tracking algorithm is integrated at both Torino and Paris testing sites and runs in real time on live video streams. The global VANAHEIM system has been presented as a demonstration at the ECCV 2012 conference. A demonstration video has been compiled from the results of the group tracking on 60 sequences from the Paris subway showing interesting groups with various activities such as *waiting*, *walking*, *lost*, *kids* and *lively*.

6.12. Crowd Event Monitoring Using Texture and Motion Analysis

Participants: Vaibhav Katiyar, Jihed Joober, François Brémond.

keywords: Crowd Event, Texture Analysis, GLCM, Optical Flow

The aim of this work is to monitor crowd event using crowd density, change of speed and orientation of group of people. For reducing complexity we are using human density rather than individual human detection and tracking. In this study Human density is quantified mainly into three groups- (1) Empty (2) Sparse (3) Dense. These are approximated by calculating Haralick features from Grey Level Co-occurrence Matrix (GLCM).

We use Optical flow for getting motion information like current speed and orientation of selected FAST feature points. Subsequently we used this information for classifying crowd behaviour into normal or abnormal categories wherein we seek for sudden change in speed or orientation heterogeneity for abnormal behaviour.

In future work this abnormal behaviour may further be classified into different events like Running, Collecting, Dispersion, Stopping/Blocking.

6.13. Detecting Falling People

Participants: Etienne Corvee, Francois Bremond.

keywords: fall, tracking, event

We have developed a people falling algorithm based on our object detection and tracking algorithm [58] and using our ontology based event detector [57]. These algorithms extract moving object trajectories from videos and triggers alarms whenever the people activity fits event models. Most surveillance systems use a multi Gaussian technique [83] to model background scene pixels. This technique is very efficient in detecting in real-time moving objects in scenes captured by a static camera, with low level of shadows, few persons interacting in the scene and with as few as possible illumination changes. This technique does not analyse the content of the moving pixels but simply assign them as foreground or background pixels.

Many state of the art algorithms exist that can recognize objects such as a person human shape, a head, a face or a couch. However, these algorithms are quite time consuming or the database used for training the system is not well adapted to our application domain. For example, people detection algorithms use databases containing thousands of image instances of standing or walking persons taken by camera from a certain distance from the persons and from a facing position. In our indoor monitoring application, cameras are located on the roof with high tilt angle so that most of the scene (e.g.rooms) is viewed. With such camera spatial configuration, the image of a person on the screen rarely corresponds to the person images in the training database. In addition, people are often occluded by the image border (the image of the full body is not available), image distortion needs to be corrected and people often have poses that are not present in the database (e.g. a person bending or sitting).

Using our multi Gaussian technique [74], after having calibrated a camera scene, a detected object is associated with a 3D width and height in two positions : the standing and lying positions. This 3D information is checked against 3D human model and any object is then labelled as either a standing person, a lying person or unknown. Many 3D filtering thresholds are used ; for example, object speed should not be greater than a human possible running speed. Second, we use an ontology based event detector to build a hierarchy of event model complexity. We detect people to have fallen on the floor if the object has been detected as a person on the floor and outside the bed and couch for at least several seconds consecutively. An example of a fallen person is shown in Figure 24 .

Figure 24. Detection of a fallen person.

6.14. People Detection Framework

Participants: Srinidhi Mukanahallipatna, Silviu-Tudor Serban, François Brémond.

keywords: LBP, Adaboost, Cascades

We present a new framework called COFROD (Comprehensive Optimization Framework for Real-time Object Detection) for object detection that focuses on improving state of the art accuracy, while maintaining real-time detection speed. The general idea behind our work is to create an efficient environment for developing and analyzing novel or optimized approaches in terms of classification, features, usage of prior knowledge and custom strategies for training and detection. In our approach we opt for a standard linear classifier such as Adaboost. Inspired by the integral channel feature approach, we compute variants of LBP and Haar-like features on multiple channels of the input image. Thus, we obtain an elevated number of computationally inexpensive features that capture substantial information. We use an extensive training technique in order to obtain optimal classifier.

We propose a comprehensive framework for object detection with an intuitive modular design and high emphasis on performance and flexibility. Its components are organized by parent-modules, child-modules and auxiliary-modules. The parent-modules contain several child-modules and focus on a general task such as Training or Detection. Child-modules solve more specific tasks, such as feature extraction, training or testing and in most cases require auxiliary-modules. The later have precise intents, for instance computing a color channel transformation or a feature response.

We present two detection configurations. One relies on a single intensively trained detector and the other as a set of specialist detectors.

Our baseline detector uses cascades in order to speed up the classifier. By removing most false positive at first stages, computation time is significantly reduced. Classifier for each cascade is generated using the training approach.

Our contribution is in the form of a hierarchical design of specialized detectors. At first level we use a version of the baseline detector in order to remove irrelevant candidates. At the second level, specialist detectors are defined. These detectors can be either independent or can use third level detectors and cumulate their output. A specialist detector can take the role of solving an exact classification issue, such as sitting pose. In that case it is trained only with data relevant to that task. In some applications, a specialist detector can be trained to perform exceptionally on a specific situation. In this case training samples are adapted to the particularity of the testing, and possibly parts of the testing sets are used for training.

This is a versatile system for object detection that excels in both accuracy and speed. We present a valuable strategy for training and a hierarchy of specialized people detectors for dealing with difficult scenarios. We also propose an interesting feature channel and a method for loosing less detection speed-up. In our approach we build upon the ideas of feature scaling instead of resizing images and of transferring most computations from detection to training, thus achieving real-time performance on VGA resolution.

Figure 25 and Figure 26 illustrate our detections results. Figure 27 shows the performance of our system compared to other. IDIAP detector was used without tuning the parameters.
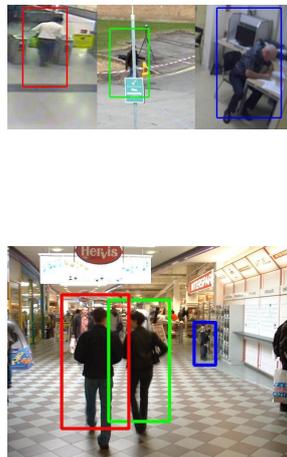
Figure 25. Detection Results

6.15. A Model-based Framework for Activity Recognition of Older People using Multiple sensors

Participants: Carlos -Fernando Crispim Junior, Qiao Ma, Baptiste Fosty, Cintia Corti, Véronique Joumier, Philippe Robert, Alexandra Konig, François Brémond, Monique Thonnat.

keywords: Activity Recognition, Multi-sensor Analysis, Surveillance System, Older people, Frailty assessment

We have been investigating a model-based activity recognition framework for the automatic detection of physical activity tests and instrumental activities of daily living (IADL, *e.g.*, preparing coffee, making a phone call) of older people. The activities are modelled using a constraint-based approach (using spatial, temporal, and *a priori* information of the scene), and a generic ontology based on natural terms which allows medical experts to easily modify the defined activity models. Activity models are organized in a hierarchical structure according to their complexity (Primitive state, Composite State, Primitive Event, and Composite Event). The framework has been tested as a system on the clinical protocol developed by the Memory Center of Nice hospital. This clinical protocol aims at studying how ICTs (Information and Communication Technologies)

Figure 26. PETS Detection Results

Figure 27. Detection results on the PETS dataset

can provide objective evidence of early symptoms of Alzheimer's disease (AD) and related conditions (like Memory Cognitive Impairment - M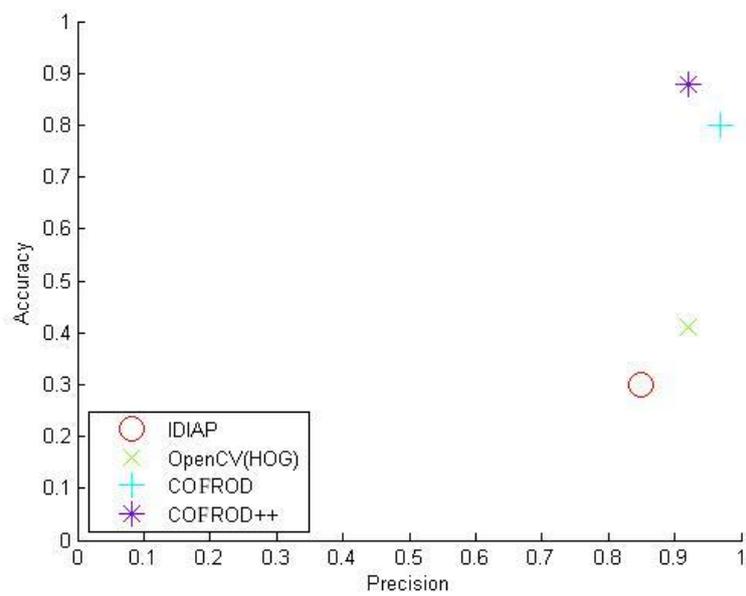CI). The Clinical protocol participants are recorded using a RGB video-camera (8 fps), a RGB-D Camera (Kinect - Microsoft), and an inertial sensor (MotionPod) which allows a multi-sensor evaluation of the activities of the participants in an observation room equipped with home appliances. A study of the use of a multi-sensor monitoring for Patient diagnosis using events annotated by experts has been performed in partnership with CHU-Nice and SMILE team of TAIWAN, and it has shown the feasibility of the use of these sensors for patient performance evaluation and differentiation of clinical protocol groups (Alzheimer's disease and healthy participants) [31] and [40]. The multi-sensor evaluation has used the proposed surveillance system prototype and has been able to detect the full set of physical activities of the scenario 1 of the clinical protocol (e.g., Guided a ctivities : Balance test, Repeated Transfer Test), with a true positive rate of 96.9% to 100% for a set of 38 patients (MCI=19, Alzheimer=9) using data of an ambient camera. An extension of the developed framework has been investigated to handle multiple sensors data in the event modeling. In this new scenario, information from the ambient camera and the inertial sensor worn on the participants chest is used (see Figure 28). The prototype using the extended framework has been tested on the automatic detection of IADLs, and preliminary results points to an average sensitivity of 91% and an average precision of 83.5%. This evaluation has been performed for 9 participants videos (15 min each, healthy: 4, MCI: 5). See [39] for more details. Future work will focus on a learning mechanism to automatic fuse events detected by a set of heterogeneous sensors, and at supporting clinicians at the task of studying differences between the activity profile of healthy participants and early to moderate stage Alzheimer's patients.
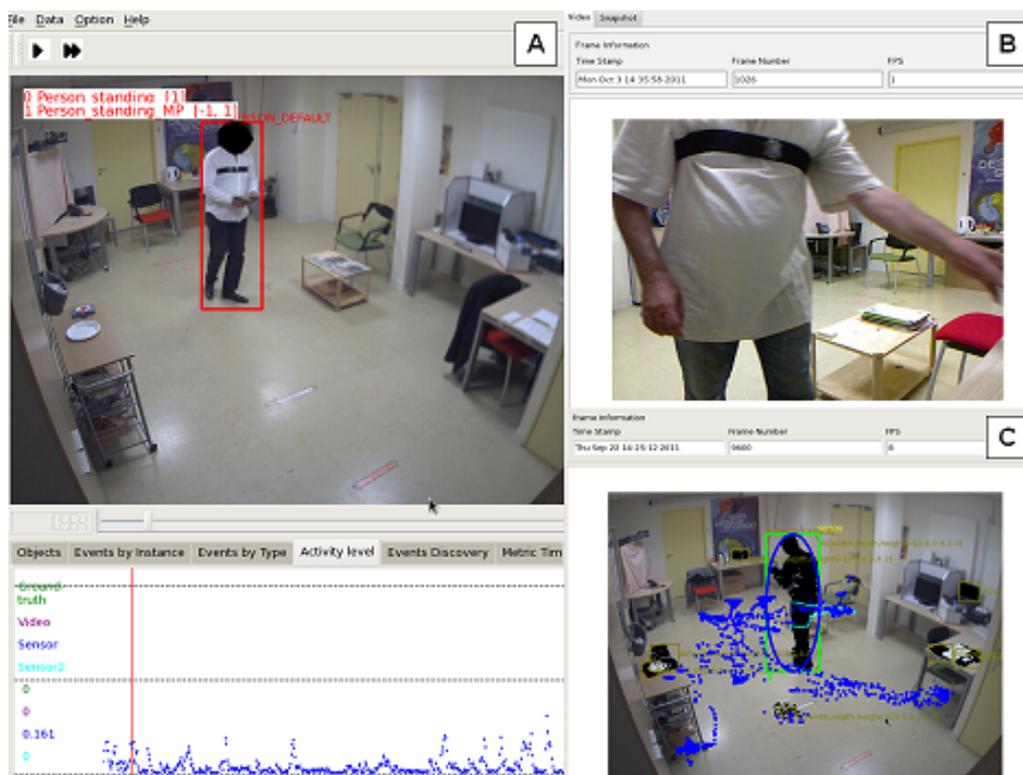

Figure 28. A: Ambient Camera View of Patient Activity. Actimetry captured by the inertial sensor is displayed at the bottom, B: RGB-D Camera View of Patient. The inertial sensor is worn by the patient by an accessory chest strap, C: Trajectory information of Patient Activity during the experimentation.

6.16. Activity Recognition for Older People using Kinect

Participants: Baptiste Fosty, Carlos -Fernando Crispim Junior, Véronique Joumier, Philippe Robert, Alexandra Konig, François Brémond, Monique Thonnat.

keywords: Activity Recognition, RGB-D camera analysis, Surveillance System, Older people, Frailty assessment

Within the context of the Dem@Care project, we have studied the potential of the RGB-D camera (Red Green Blue + Depth) from Microsoft (Kinect) for an activity recognition system developed to extract automatically and objectively evidences of early symptoms of Alzheimer's disease (AD) and related conditions (like Memory Cognitive Impairment - MCI) for older people. This system is designed on a model-based activity recognition framework. Using a constraint-based approach with contextual and spatio-temporal informations of the scene, we have developed activity models related to the physical activity part of the protocol (Scenario 1, guided activities : balance test, walking test, repeated transfers posture between sitting and standing). These models are organized in a hierarchical structure according to their complexity (Primitive state, Composite State, Primitive Event, and Composite Event). This work is an adaptation of the work performed for multi-sensor analysis [39].

Several steps are needed to adapt the processing. We had for example to generate new ground truth, or we had to design new 3D zones of interest according to Kinect point of view and referential (differing from the 2D camera). Moreover, in order to improve the reliability of the results, we had to solve several issues in the processing chain. For instance, Kinect and the detection algorithm provided by OpenNi and Nestk (free libraries) have several limitations which leads to wrong detection of human. We proposed in these cases several solutions like filtering wrong object detections by size (see Figure29 C) or recomputing the height of older people based on their head when wearing black pants (absorption of infrared) (see Figure 29 D).

For the experimentation, we have processed the data recorded for 30 patients. The results are shown in Figure 30 . With a true positive rate of almost 97% and a precision of 94.2%, our system is able to extract most of the activities performed by patients. Then, relevant and objective information can be delivered to clinicians, to assess the patient frailty. For further information on the performance of the detection process, we also generate the results frame by frame, which are shown in Figure 31 . We see there that the performance of the event detection in terms of true positive rate is almost as good as by events (94.5%). Nevertheless, if we focus on the precision, it is lower than previously. This means that we still need to improve detection accuracy of the beginning and the end of an event.

Future work will focus on using the human skeleton to extract finest information on the patient activity and to process more scenarios (semi-guided and free).

6.17. Descriptors of Depth-Camera Videos for Alzheimer Symptom Detection

Participants: Guillaume Charpiat, Sorana Capalnean, Bertrand Simon, Baptiste Fosty, Véronique Joumier.

keywords: Kinect, action description, video analysis

In a collaboration with the CHU hospital of Nice, a dataset of videos was recorded, where elderly are asked by doctors to perform a number of predefined exercises (like walking, standing-sitting, equilibrium test), and recorded with an RGBD camera (Kinect). Our task is to analyze the videos and detect automatically early Alzheimer symptoms, through statistical learning. Here we focus on the 3D depth sensor (no use of the RGB image), and aim at providing action descriptors that are accurate enough to be informative.

During her internship in the Stars team, Sorana Capalnean proposed descriptors relying directly on the 3D points of the scene. First, based on trajectory analysis, she proposed a way to recognize the different physical exercises. Then she proposed, for each exercise, specific descriptors aiming at providing the information asked by doctors, such as step length, frequency and asymmetry for the walking exercise, or sitting speed and acceleration for the second exercise, etc. Problems to deal with included the high level of noise in the 3D cloud of points given by the Kinect, as well as an accurate localization of the floor.
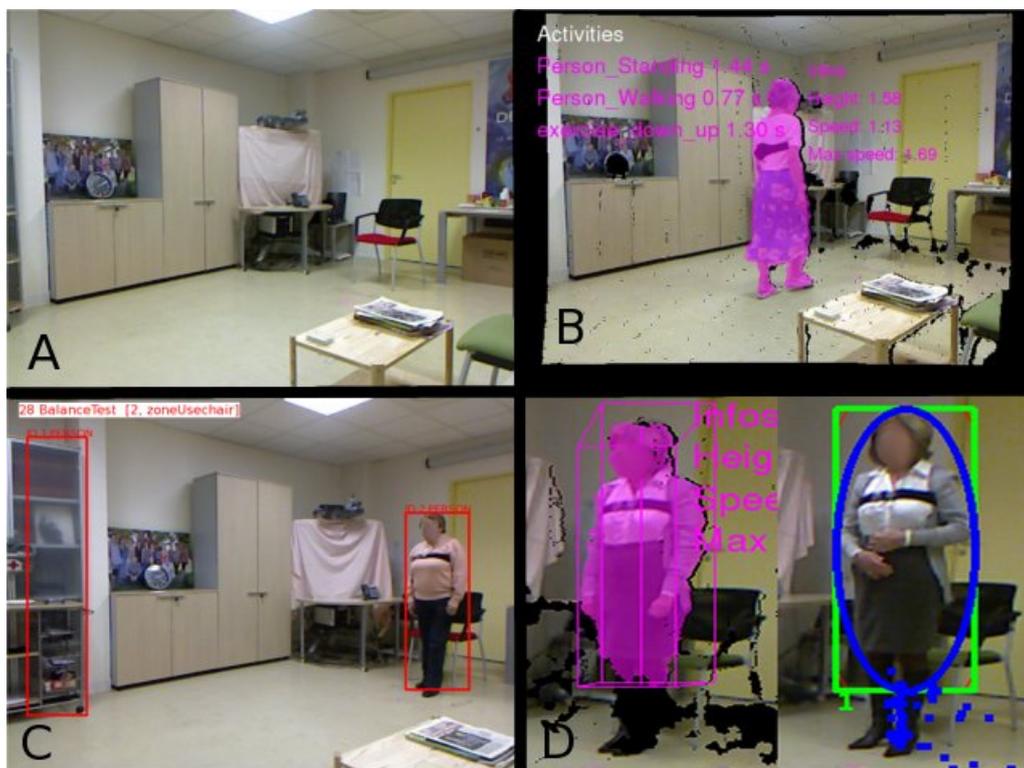
Figure 29. A: RGB-D camera view of the scene, B: 3D representation of the scene with some event detection, C: people detection problem (furniture detected as extra person), D: people detection problem (black clothes not detected).

	Nb GT	Detected		Sensitivity (%)	Precision (%)	Fscore (%)
BalanceTest	30	VP	30	100	100	100
		FP	0			
		FN	0			
WalkingTest_firstAttempt	30	TP	30	100	90,9	95,2
		FP	3			
		FN	0			
WalkingTest_secondAttempt	30	TP	27	90	100	94,7
		FP	0			
		FN	3			
RepeatedTransfersTest	30	TP	30	100	90,9	95,2
		FP	3			
		FN	0			
UpAndGoTest	30	TP	28	93,3	90,3	91,8
		FP	3			
		FN	2			
TOTAL	**150**	**TP**	**145**	**96,6**	**94,2**	**95,4**
		FP	**9**			
		FN	**5**			

Figure 30. Results by events (GT = ground truth, TP = true positive, FP = false positive, FN = false negative)

	Nb GT	Detected		Sensitivity (%)	Precision (%)	Fscore (%)
BalanceTest	37235	TP	37216	99,9	70,8	82,9
		FP	15322			
		FN	19			
WalkingTest_firstAttempt	1537	TP	849	55,2	94,1	69,6
		FP	53			
		FN	688			
WalkingTest_secondAttempt	1425	TP	857	60,1	62,6	61,3
		FP	513			
		FN	568			
RepeatedTransfersTest	5190	TP	4496	86,6	94,6	90,4
		FP	257			
		FN	694			
UpAndGoTest	3345	TP	2670	79,8	86,2	82,9
		FP	428			
		FN	675			
TOTAL	**48732**	**TP**	**46088**	**94,5**	**73,6**	**82,8**
		FP	**16573**			
		FN	**2644**			

Figure 31. Results by events (GT = ground truth, TP = true positive, FP = false positive, FN = false negative)

During his internship, Bertrand Simon proposed other kinds of descriptors, based on the articulations of the human skeleton given by OpenNI. These articulations are however very noisy too, so that a pre-filtering step of the data in time had to be performed. Various coordinate systems were studied, to reach the highest robustness. The work focused not only on descriptors but also on metrics suitable to compare gestures (in the phase space as well as in the space of trajectories). See figure 32 for an example.

These descriptors are designed to be robust to camera noise and to extract the relevant information from the videos; however their statistical analysis still remains to be done, to recognize Alzheimer symptoms during the different exercises.
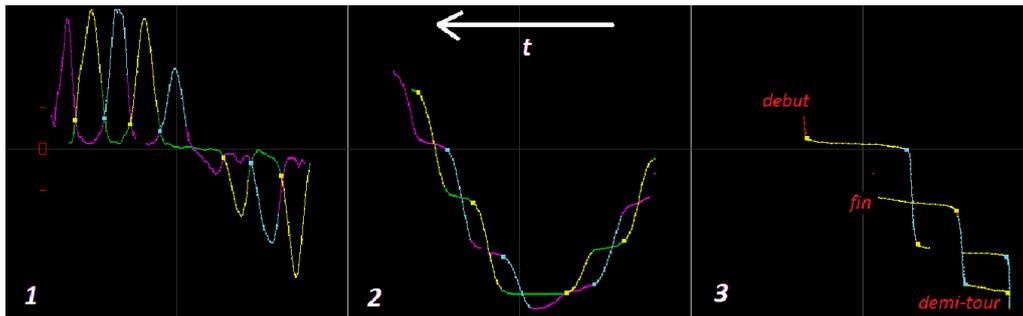

Figure 32. Curves obtained during a person's walk in backward then forward directions. Purple and blue curves stand for the right foot, while green and yellow ones stand for the left one. Graph 1 shows the speeds as a function of time; Graph 2 shows the locations as a function of time; Graph 3 shows the location of the right foot as a function of the location of the left foot.

6.18. Online Activity Learning from Subway Surveillance Videos

Participants: Jose-Luis Patino Vilchis, Abhineshwar Tomar, François Brémond, Monique Thonnat.

Keywords: Activity learning, clustering, trajectory analysis, subway surveillance

This work provides a new method for activity learning from subway surveillance videos. This is achieved by learning the main activity zones in the observed scene by taking as input the trajectories of detected mobile objects. This provides us the information on the occupancy of the different areas of the scene. In a second step, these learned zones are employed to extract people activities by relating mobile trajectories to the learned zones, in this way, the activity of a person can be summarised as the series of zones that the person has visited. If the person resides in the single zone this activity is also classified as a standing. For the analysis of the trajectory, a multiresolution analysis is set such that a trajectory is segmented into a series of tracklets based on changing speed points thus extracting the information when people stop to interact with elements of the scene or other people. Starting and ending tracklet points are fed to an advantageous incremental clustering algorithm to create an initial partition of the scene. Similarity relations between resulting clusters are modelled employing fuzzy relations. A clustering algorithm based on the transitive closure calculation of the fuzzy relations easily builds the final structure of the scene. To allow for incremental learning and update of activity zones (and thus people activities), fuzzy relations are defined with online learning terms. The approach is tested on the extraction of activities from the video recorded at one entrance hall in the Torino (Italy) underground system. Figure 33 presents the learned zones corresponding to the analyzed video. To test the validity of the activity extraction a one hour video was annotated with activities (corresponding to each trajectory) according to user defined ground-truth zones. After the comparison, following results were obtained: TP:26, FP:3, FN:1, Precision:0.89, Sensitivity:0.96. This work is published in [43].
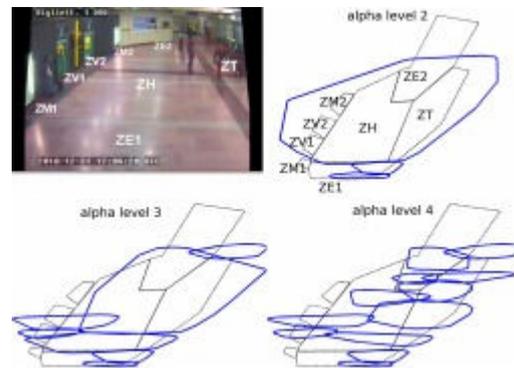
Figure 33. Left top panel: Original underground scene observed by the camera with user-defined areas delimiting the scene. Remaining panels: Learned zones in a 3D top view. They correspond to activity areas as discovered with our algorithm. Different granularity levels allow understanding the activity with different resolutions

6.19. Automatic Activity Detection Modeling and Recognition: ADMR

Participants: Guido-Tomas Pusiol, François Brémond.

This year a new Ph.D. thesis has been defended [30]. The main objective of the thesis is to propose a complete framework for the automatic activity discovery, modeling and recognition using video information. The framework uses perceptual information (e.g. trajectories) as input and goes up to activities (semantics). The framework is divided into five main parts:

1. We break the video into chunks to characterize activities. We propose different techniques to extract perceptual features from the chunks. This way, we build packages of perceptual features capable of describing activity occurring in small periods of time.

2. We propose to learn the video contextual information. We build scene models by learning salient perceptual features. The models end up containing interesting scene regions capable of describing basic semantics (i.e. region where interactions occur).

3. We propose to reduce the gap between low-level vision information and semantic interpretation, by building an intermediate layer composed of Primitive Events. The proposed representation for primitive events aims at describing the meaningful motions over the scene. This is achieved by abstracting perceptual features using contextual information in an unsupervised manner.

4. We propose a pattern-based method to discover activities at multiple resolutions (i.e. activities and sub-activities). Also, we propose a generative method to model multi-resolution activities. The models are built as a flexible probabilistic framework easy to update.

5. We propose an activity recognition method that finds in a deterministic manner the occurrences of modelled activities in unseen datasets. Semantics are provided by the method under user interaction. All this research work has been evaluated using real datasets of people living in an apartment (home-care application) and elder patients in a hospital.

The work has also been evaluated for other types of applications such as sleeping monitoring. For example, Figure 34 display the results of the activity discovery method during 6 hours (left to right) applied to the center of mass (3D) of a tracked sleeping person. The colored segments represent hierarchical (bottom-up is finer-coarse) discovered activity which matches with sleeping postural movements. The segments have similar color when postural movements are similar. For example, the segment (j) is the only time the person sleeps upside down. Also, health professionals analysed the results claiming that the segments corresponds to normal

sleeping cycle, where low motion is noticed at the beginning of the sleep and more motion is shown when the person have a lighter sleep when starts waking up.
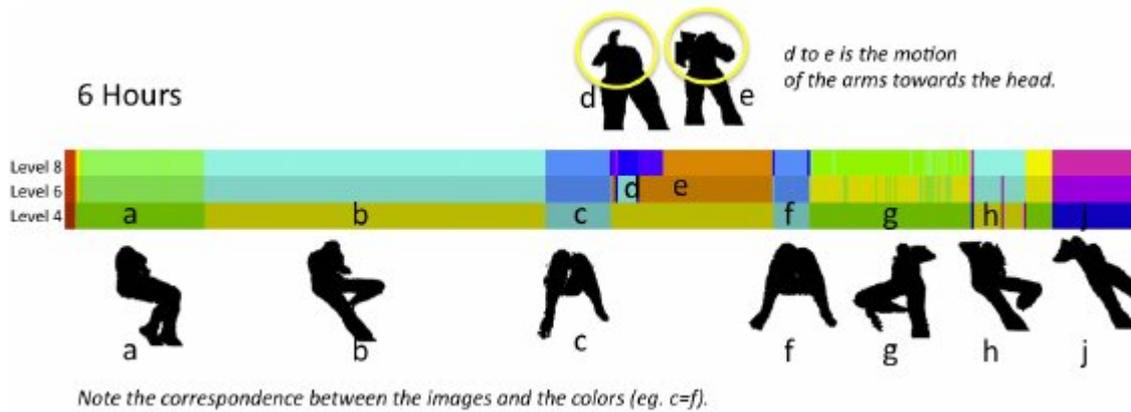

Figure 34. Results of the activity discovery method during 6 hours

6.20. SUP Software Platform

Participants: Julien Gueytat, Baptiste Fosty, Anh tuan Nghiem, Leonardo Rocha, François Brémond.

Our team focuses on developing Scene Understanding Platform (SUP) (see section 5.1). This platform has been designed for analyzing a video content. SUP is able to recognize simple events such as 'falling', 'walking' of a person. We can easily build new analyzing system thanks to a set of algorithms also called plugins. The order of those plugins and their parameters can be changed at run time and the result visualized. This platform has many more advantages such as easy serialization to save and replay a scene, portability to Mac, Windows or Linux, ... All those advantages are available since we are working together with the software developers team DREAM. Many Inria teams are pushing together to improve a common Inria development toolkit DTK. Our SUP framework is one of the DTK-like framework developed at Inria.

Currently, we have fully integrated OpenCV library with SUP and the next step is to integrate OpenNI to get depth map processing algorithms from PrimeSense running in SUP. Updates and presentations of our framework can be found on our team website http://team.inria.fr/stars. Detailed tips for users are given on our Wiki website http://wiki.inria.fr/stars and sources are hosted thanks to the new Source Control Management tool.

6.21. Qualitative Evaluation of Detection and Tracking Performance

Participants: Swaminathan Sankaranarayanan, François Brémond.

We study an evaluation approach for detection and tracking systems. Given an algorithm that detects people and simultaneously tracks them, we evaluate its output by considering the complexity of the input scene. Some videos used for the evaluation are recorded using the Kinect sensor which provides for an automated ground truth acquisition system. To analyse the algorithm performance, a number of reasons due to which an algorithm might fail is investigated and quantified over the entire video sequence. A set of features called Scene Complexity measures are obtained for each input frame. The variability in the algorithm performance is modelled by these complexity measures using a polynomial regression model. From the regression statistics, we show that we can compare the performance of two different algorithms and also quantify the relative influence of the scene complexity measures on a given algorithm. This work has been published in [44].

6.22. Model-Driven Engineering and Video-surveillance

Participants: Sabine Moisan, Jean-Paul Rigault, Luis-Emiliano Sanchez.

keywords: Feature Model Optimization, Software Metrics, Requirement specification, Component-based system, Dynamic Adaptive Systems, Model-Driven Engineering, Heuristic Search, Constraint Satisfaction Problems

The domain of video surveillance (VS) offers an ideal training ground for Software Engineering studies, because of the huge variability in both the surveillance tasks and the video analysis algorithms [41]. The various VS tasks (counting, intrusion detection, tracking, scenario recognition) have different requirements. Observation conditions, objects of interest, device configuration... may vary from one application to another. On the implementation side, selecting the components themselves, assembling them, and tuning their parameters to comply with context may lead to a multitude of variants. Moreover, the context is not fixed, it evolves dynamically and requires run time adaptation of the component assembly.

Our work relies on Feature Models, a well-known formalism to represent variability in software systems. This year we have focused on an architecture for run time adaptation and on metrics to drive dynamic architecture changes.

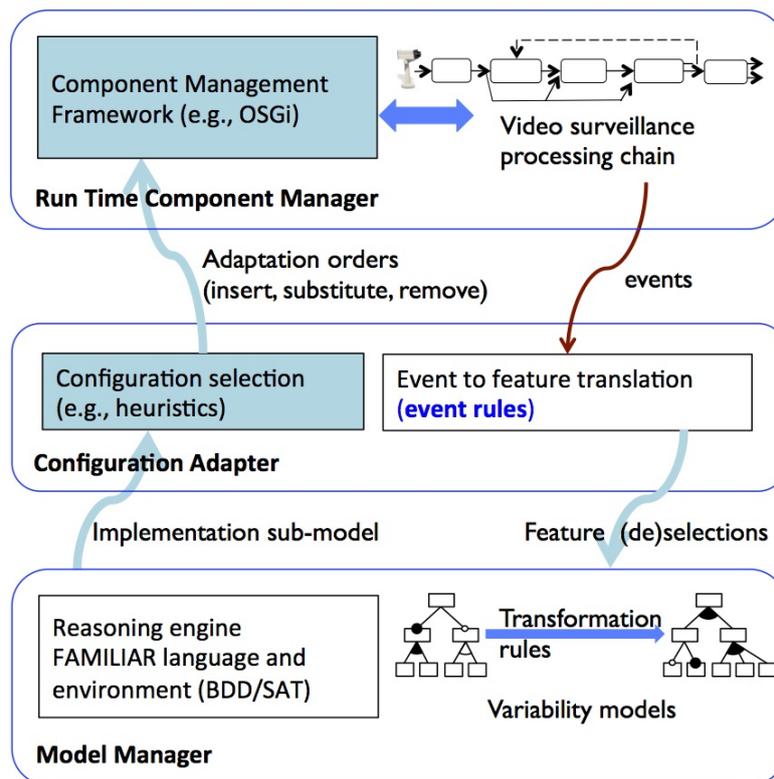### *6.22.1. Run Time Adaptation Architecture*

Figure 35. Architecture of an adaptive system using models at run time. (In light blue the elements studied this year.)

The architecture of the run time system (also used for initialization at deployment time) is based on three collaborating modules as shown in Figure 35 . A *Run Time Component Manager* (RTCM) cooperates with the low levels (to manage the software components and capture events) and applies configuration changes. A *Configuration Adapter* (CA) receives events from the RTCM, and propagates them as features into the models to obtain a new configuration. The *Model Manager* (MM) embeds a specialized scripting language for Feature Models (FAMILIAR [52], [53][1]) to manage the representation of the two specialized feature models and applies constraints and model transformations on them. The *Model Manager* produces new component configurations (a model specialization) that it sends to the CA. At its turn, the CA selects one single configuration (possibly using heuristics) and converts it into component operations to be applied by the RTCM.

This year we first finalized the interface between the Model Manager and the Configuration Adapter. On one hand, we transform the feature models obtained from FAMILIAR into C++ representations enriched with software component information. On the other hand, we dynamically transform context change events into requests to FAMILIAR.

Second, we searched for a suitable technology for handling components in the Run Time Component Manager. OSGi is an adequate de facto standard but it is mainly available in the Java world. However we could find a C++ implementation, complete enough for our needs (SOF, Service Oriented Framework [65]). However, SOF has to be completed to adjust to the needs of our end users who are the video system developers. Thus, we are currently building a multi-threaded service layer on top of SOF, easy to use and hiding most of the nitty-gritty technical details of thread programming and SOF component manipulation. This layer provides end users with a set of simple patterns and allow them to concentrate only on the code of video services (such as acquisition, segmentation, tracking...).

As a matter of feasability study we are building an experimental video self adaptive system based on the afore mentionned architecture. Software components are implemented with the OpenCV library. In the final system, feature models and software components continuously interact in real time, modifying the whole system in response to changes in its environment.

6.22.2. *Metrics on Feature Models to Optimize Configuration Adaptation at Run Time*

As shown on figure 35 , the Configuration Adapter has to set up a suitable component configuration of the run time system. For this, each time the context changes, it receives a set of valid configurations (a feature *sub*-model) from the Model Manager. In most cases, this set contains more than one configuration. Of course, only one configuration can be applied at a given time and the problem is to select the "best" one. Here, "best" is a trade-off between several non-functional aspects: performance, quality of service, time cost for replacing the current configuration, etc.

It is thus necessary to rank the configurations. Our approach is to define metrics suitable for comparing configurations. Then the problem comes down to the widely studied problem of *Feature model optimization* [55]. This problem is known to be an intractable combinatorial optimization problem in general.

We started with a study of the state of the art: metrics for general graphs as well specific to feature models, optimization and requirement specification on feature models... We obtained a structured catalog of quality and feature model metrics. Then we selected solutions based on heuristic search algorithms using quality and feature model metrics. We thus propose several strategies and heuristics offering different properties regarding optimality of results and execution efficiency.

These strategies and heuristics have been implemented, tested, and analyzed using random generated feature models. We got empirical measures about their properties, such as completeness, optimality, time and memory efficiency, scalability... This allows us to compare the performance of the different algorithms and heuristics, and to combine them in order to achieve a good trade-off between optimality and efficiency. Finally, the proposed algorithms have been introduced as part of the Configuration Adapter module.

---

[1]FAMILIAR has been developed at the I3S laboratory by the Modalis team.

This work is quite original from several aspects. First, we did not find any study using heuristic search algorithms for solving the feature optimization problem. Most studies apply Artificial Intelligence techniques such as CSP solvers, planning agents, genetic algorithms... Second, we do not restrict to the optimization of linear objective functions, but we also address non-linear ones allowing us to take into account a broader set of criteria. Among the possible criteria we consider quality of service of components, their performance, their set up delay, the cost of their replacement, etc. Finally, we apply our metrics at run time whereas most studies consider metrics only for static analysis of feature models.

Currently, we are still working on new variants of the search algorithms and new heuristics relying on techniques proposed in the domains of heuristic search and constraint satisfaction problems.

6.23. Synchronous Modelling and Activity Recognition

Participants: Annie Ressouche, Sabine Moisan, Jean-Paul Rigault, Daniel Gaffé.

6.23.1. Scenario Analysis Module (SAM)

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 other 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 [76] and we adapt usual techniques of synchronous modelling approach to express scenario behaviours. This approach facilitates scenario validation and allows us to generate a recognizer for each scenario model.

In addition, 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. To face this challenge, we introduced in the generation of 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. Indeed, we relied on Lustre [66] synchronous language to express the automata semantics of scenario models as Boolean equation systems. This approach was successful and shows that we can consider a synchronous framework to generate validated scenario recognition engines. This year, in order to improve efficiency (and to tackle the real time recognition problem), we begin to rely on CLEM (see section 6.23.2) toolkit to generate such recognition engines. The reason is threefold: (1) CLEM is becoming a mature synchronous programming environment; (2) we can use the CLEM compiler to build our own compiler; (3) CLEM supplies the possibility of using NuSMV [61] model checker, which is more powerful than the Lustre model-checker. Moreover, thanks to CLEM compiler into Boolean equation systems, we can compute the expected events of the next instant on the fly, by propagation of information related to the current instant.

6.23.2. The clem Workflow

This research axis concerns the theoretical study of a synchronous language LE with modular compilation and the development of a toolkit (see Figure 9) 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. This year, we mainly work on theoretical aspects of CLEM.

First, synchronous language semantics usually characterizes each output and local signal status (as present or absent) according to input signal status. To reach our goal, we defined a semantics that translates LE programs into equation systems. This semantics bears and grows richer the knowledge about signals and is never in contradiction with previous deduction (this property is called constructiveness). In such an approach, causality turns out to be a scheduling evaluation problem. We need to determine all the partial orders of equation systems and to compute them, we consider a 4-valued algebra to characterize the knowledge of signal status (unknown, present, absent, overknown). Previously, we relied on 4-valued Boolean algebra [19], [20] which defines the negation of unknown as overknown. The advantage of this way is to benefit from Boolean algebras laws to compute equation system solutions. The drawback concerns signal status evaluation which does not correspond to usual interpretation (not unknown = unknown and not overknown = overknown). To avoid this drawback, we study other kinds of algebras well suited to define synchronous languages semantics. In [49], we choose an algebra which is a bilattice and we show that it is well suited to solve our problem. It is a new application of general bilattice theory [64]. But, the algebra we defined is no more a Boolean algebra, but we prove (always in [49]), that the main laws of Boolean algebras hold as distributivity laws, associativity laws, idempotence laws, etc. After compilation, signals have to be projected into Boolean values. Bilattice theory offers an isomorphism between 4-valued status and pair of Boolean.

Second, the algorithm which computes partial orders relies 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. Inputs (resp. outputs) have date 0 and the algorithm recursively increases the dates of nodes in the upstream (resp downstream) dependencies graph. Hence, the algorithm determines an *earliest* date and a *latest* date for equation system variables. Moreover, we can compute the dates of variables of a global equation system starting from dates already computed for variables which were inputs and outputs in a sub equation system corresponding to a sub program of the global program[2]. This way of compiling is the corner stone of our approach [20]. We defined two approaches to compute all the valid partial orders of equation systems, either applying critical path scheduling technique (CPM) [3] or applying fix point theory: the vector of earliest (resp. latest) dates can be computed as the least fix point of a monotonic increasing function. This year we have proved that we can compute dates either starting from a global equation system or considering equation system where some variables are abstracted (i.e they have no definition) and whose dates have been already computed. To achieve the demonstration, we rely on an algebraic characterization of dates and thanks to uniqueness property of least fix points, we can deduce that the result is the same for a global equation systems as for its abstraction. We are in the process of publishing this result. From an implementation point of view, we use CPM approach to implement our scheduling algorithm since it is more efficient than fix point consideration. Of course both ways yield the same result. Indeed, fix point approach is useful for a theoretical concern.

6.23.3. *Multiple Services for Device Adaptive Platform for Scenario Recognition*

The aim of this research axis is to federate the inherent constraints of an activity recognition platform like SUP (see section 5.1) 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 [85], [84], [68]) 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 behaviour models associated with a well sound modelling of composition operation may guarantee the respect of the usage contract.

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 behavioural 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.

---

[2]these variables are local in the global equation system
[3] http://pmbook.ce.cmu.edu/10_Fundamental_Scheduling_Procedures.html

To define such synchronous component, user can specify a synchronous component per sub assembly corresponding to a concern and compose the synchronous components connected with the same critical component in order to get an only synchronous component. Thus, we supply a *composition under constraints* of synchronous components and we proved that this operation preserves already separately verified properties of synchronous components [79], [78].

The main challenge of this approach 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 other, then we get a combinatorial number of constraints with respect to the number of synchronous monitors and inputs of the critical component. To tackle this problem, we replace the effective description of constraints by a generic specification of them in the critical component. But, we must offer a way to express these generic constraints. Then, each synchronous component has a synchronous controller, which is the projection of the generic constraints on its output set. The global synchronous component is the synchronous parallel composition of all basic components and their synchronous controllers. Moreover, according to synchronous parallel composition features, the property preservation result we have still hold.

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, Antoine Lejay, Sylvain Maire, Sebastian Niklitschek Soto, Nicolas Perrin, Denis Talay, Etienne Tanré, Denis Villemonais, Laurent Violeau.

6.1.1. Published works and preprints

- In collaboration with P.-E. Jabin (University of Maryland), J.-F. Jabir and J. Fontbona (CMM and Universidad de Chile, Santiago de Chile), M. Bossy 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 to an incompressible Lagrangian stochastic model in periodic domain. The paper is in positive revision for publication in *Communications in Partial Differential Equations* [33]. http://hal.inria.fr/hal-00691712

- 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 [13], [14] 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, http://hal.inria.fr/inria-00616765. In collaboration with Mathieu Richard (Ecole Polytechnique, Palaiseau), they also extended some of these results to the case of splitting trees with mutations occuring at birth of individuals [15], http://hal.inria.fr/hal-00736036.

- N. Champagnat obtained with P. Diaconis (Stanford Univ.) and L. Miclo (Univ. Toulouse 3) the full spectral decomposition of the transition matrix of two-dimensional Markov chains $(X_n, Y_n)_{n \geq 0}$ in $\mathbb{Z}_+^2$, without immigration or mutation, which are *neutral* in the sense that $(X_n + Y_n)_{n \geq 0}$ is a Markov process. Because of the specific form of the eigenvectors, they were also able to characterize all the Dirichlet eigenvectors in subdomains of $\mathbb{Z}_+^2$ of the form $\{(i, j) \in \mathbb{Z}_+^2 : i + j \geq d\}$ for all $d \geq 0$. As an application, they could determine the quasi-stationary and quasi-limiting distributions of such processes [12], http://hal.inria.fr/hal-00672938.

- N. Champagnat studied with F. Campillo (EPI MODEMIC, Inria Sophia Antipolis — Méditerrannée) individual based models of clonal plants where plants interact through the network formed by the rizhomes or stolons linking plants. In the limit of large population, they obtained a PDE governing the dynamics of population densities in space [11], http://hal.inria.fr/hal-00723209.

- M. Deaconu and S. Herrmann introduced a new method for the simulation of the hitting times of nonlinear boundaries for Bessel processes. This method combines the method of images and the random walk on spheres method. They construct the so called walk on moving spheres algorithm. This approach can be applied for the hitting time of a given level for the Cox-Ingersoll-Ross process and thus be used in models coming from finance and neuroscience [17], http://hal.inria.fr/hal-00636056/en. This work is part of the ANR MANDy project.

- J. Inglis and E. Tanré studied with F. Delarue and S. Rubenthaler (Univ. Nice – Sophia Antipolis) the global solvability of a networked system of integrate-and-fire neurons proposed in the neuroscience literature. In the mean-field limit the equation resembles a McKean-Vlasov equation, but is highly non-standard and previous attempts at rigorous analysis were not satisfactory. They here bridge this gap, and shed light on a surprisingly complicated problem [35], http://hal.inria.fr/hal-00747565.

- A. Lejay continued his long term investigations on probabilistic interpretations and Monte Carlo simulations of interfaces conditions, such as ones arising in discontinuous media. With G. Pichot (IRISA, Rennes), he has developed a series of tests and benchmarks regarding one-dimensional Monte Carlo methods, such as the ones proposed in [19], http://hal.inria.fr/hal-00649170. He has also developed a new family of stochastic diffusion processes, called the *snapping out Brownian motion*, in order to take into account an interface condition where the concentration of the fluid is proportional to its gradient. Finally, A. Lejay and S. Maire also proposed new methods and tested a few ones to deal with the locally isotropic case for multidimensional problems [18], http://hal.inria.fr/hal-00689581.

- With A. Kohatsu-Higa (Ristumeikan University) and K. Yasuda (Hosei university), A. Lejay has continued his work [25] on the simulation of SDE with a discontinuous drift. http://hal.inria.fr/hal-00670123

- With L. Coutin (University of Toulouse), A. Lejay has developed an appropriate framework to deal with linear rough differential equations, extending some results (Magnus formula, Dyson series...) to this case. Using theses properties, they have studied the sensitivites of solutions of rough differential equations with respect to the signal, the vector field or the starting point. They have provided new results such as the Hölder continuity of the derivative of the so called Itô map which transforms a rough path to the solution of a rough differential equation [34]. http://hal.inria.fr/hal-00722900

- S. Maire and C. Prissette (Univ. du Sud – Toulon – Var) have developed in [21] a stochastic algorithm to solve Sudoku puzzles using estimation of distribution coupled with restart techniques. http://hal.inria.fr/inria-00591852

- S. Maire and E. Tanré have generalised the spectral methods for elliptic PDEs developed in [42], [43] 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 [40]. By taking into account these additional properties, they still obtained a spectral matrix having a condition number converging to one [36]. http://hal.inria.fr/hal-00677529

- C. Graham (Ecole Polytechnique) and D. Talay wrote the first volume [27] of their series of books published by Springer on the Foundations of Stochastic Simulations. They started to write the second volume.

- D. Villemonais wrote with S. Méléard (École Polytechnique) a survey on quasi-stationary distributions and Q-processes for stochastic models of population dynamics. This survey also contains a detailed numerical study of the behaviour of classical models with extinction [23]. http://hal.inria.fr/hal-00653834

- D. Villemonais worked on the empirical distribution of Fleming-Viot type particle systems. Using couplings with reflected diffusion processes, he proved the uniform tightness of such empirical distributions and deduced the non-degeneracy of the law of diffusion processes conditioned not to hit a boundary [39]. http://hal.inria.fr/hal-00681601

- D. Villemonais proved in [38] a general approximation method for Markov processes conditioned not be killed. The method is based on a mean field interacting particles system which is easy to simulate. The study also details the particular case of time/environment dependent diffusion processes. http://hal.archives-ouvertes.fr/hal-00598085

6.1.2. Other works in progress

- N. Champagnat and D. Villemonais obtained criterions for existence and uniqueness of quasi-stationary distributions and Q-processes for general absorbed Markov processes. A quasi-stationary distribution is a stationary distribution conditionnally on non-absorbtion, and the Q-process is defined as the original Markov process conditionned to never be absorbed. The criterion that they obtain also ensures exponential convergence of the conditionned t-marginal of the process conditionned not to be absorbed at time t to the quasi-stationary distribution and the exponential ergodicity of the Q-process. This work is currently being written.

- N. Champagnat and D. Villemonais work on time-reversal of absorbed processes, which allow to characterize the path to extinction in extinct populations which are known to be non-extinct at some time in the past. They plan to apply these results on practical ecological situations.

- 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 finite-horizon optimal control problem on a branching-diffusion process. In addition, he is working on modelling of a pH-mediated cancer treament.

- M. Deaconu and S. Herrmann continue the study of the hitting times for Bessel processes in the situation of noninteger dimensions and also in the application of this method to the simulation of the Brownian hitting time,

- M. Deaconu starts a collaboration with L. Beznea (Simion Stoilow Institute of Mathematics of the Romanian Academy) on coagulation-fragmentation models and their connection with branching processes.

- M. Deaconu studies in collaboration with F. Nobile and F. Tesei (EPFL) a pollution model by using hitting times of stochastic processes.

- 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 neuroscience. This work is part of the ANR MANDy project.

- S. Larnier joined the team in September as a post-doctoral researcher and began working with A. Lejay on data assimilation in order to predict the ocean wave energy from the knowledge of near-shore incoming waves. They started a collaboration on video data with R. Almar (LEGOS, Toulouse) and R. Cienfuegos (Pontificia Universidad Católica de Chile).

- S. Maire works with M. Simon (Mainz Univ.) on electrical impedance tomography problems using new Monte Carlo schemes that deal with Robin and transmission boundary conditions.

- S. Maire develops with I. Dimov (Bulgarian academy of sciences) a Monte Carlo method called the walk on equations to solve linear systems of algebraic equations.

- S. Niklitschek has continued his PhD. work under the supervision of D. Talay. They were able to extended their first work in which they gave a probabilistic interpretation of a parabolic equation with discontinuous drift and proved the weak rate of convergence of the Euler method using the accurate pointwise estimates obtained for the derivatives of the solution, to the case in which both drift and diffusion coefficients are discontinuous. Both results are consistent with each other, and also with the results obtained by M. Martinez and D. Talay in [22].

- N. Perrin continued his PhD. on stochastic methods in molecular dynamics under the supervision of M. Bossy, N. Champagnat and D. Talay. This year, he studied a stochastic interpretation of parabolic PDEs with divergence form operators involved in the Poisson-Boltzmann PDE of molecular dynamics, and the associated numerical Monte Carlo method. He also continued his study of 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.

- P. Guiraud (University of Valparaiso) and E. Tanré study the effect of noise in the phenomenon of spontaneous synchronisation in a network of full connected integrate-and-fire neurons. They detail cases in which the phenomenon of synchronization persists in a noisy environment, cases in which noise permits to accelerate synchronization, and cases in which noise permits to observe synchronization while noiseless model does not have synchronization.

- P. Orio (Centro Interdisciplinario de Neurociencia de Valparaiso) and E. Tanré work on the comparison of global properties of the solution of mathematical models and the associated measurements obtained by experiments.

- 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 (MOISE team, Inria Sophia Antipolis – Méditerranée, Montpellier). Laurent studied this year the rate of convergence of the Nadaraya-Watson conditional estimator for "linear" kinetic processes. He is currently working on the rate of convergence of the particle approximation of kinetic conditional McKean-Vlasov 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.

- D. Talay is working with J. Bion-Nadal (Ecole Polytechnique) on applications of risk measures to the calibration of stochastic models, with N. Touzi (Ecole Polytechnique) on the stochastic control of stochastic differential equations with weighted local times, and with O. Bardou (GDF) on Edgeworth expansions for the Central Limit Theorem for Brownian martingales whose integrands depend on ergodic diffusion processes.

6.2. Financial Mathematics

Participants: Mireille Bossy, Paul Charton, Dalia Ibrahim, Denis Talay, Etienne Tanré.

- Mireille Bossy, in collaboration with H. Quinteros (Univ. Chile) worked on the rate of convergence of non Lipschitz diffusion processes discretized with the symetrized Milstein scheme. Under the same kind of hypotheses than in [41] on the symetrized Euler scheme, they obtained the expected improvement of the strong rate of convergence, when the diffusion coefficient is of the form $\sigma(x) = x^\alpha$, with $\alpha \in [1/2, 1[$.

  A preprint is being written.

- P. Charton continued his PhD. under the supervision of M. Deaconu and A. Lejay. He studied some storage strategies for wind farms.

- **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. She studied the performances of this indicator in a modified Black-Scholes model such that the volatility is equal to σ_0 up to a random time τ, independent of the Brownian motion governing the prices. After τ, the volatility is equal to σ_1. She proved that Bollinger Bandwidth indicator can detect the time change (at which the volatility changes its value), in the case of small and large volatilities. She has also exhibited a mathematical optimal allocation strategy, by decomposing the initial allocation problem into an allocation problem before the change time τ and an allocation problem after τ, in order to circumvent some technical problems brought from the change of volatility.

  This work is part of the contract with FINRISK.

- In collaboration with C. Michel (CA-CIB) and V. Reutenauer (Citi), D. Talay and E. Tanré worked on the

  – the study of the liquidity risk in the interest rate options market;

  – the minimization of the hedging error in interest rates Gaussian models by means of strategies designed in an effective way by using stochastic optimization algorithms.

- P. Protter (Columbia University) and D. Talay continue to work on bubbles time evolution models, which leads them to try to extend Feller's results on explosion times for stochastic differetential equations.

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

5. New Results

5.1. Analysis of structures resulting from meristem activity

5.1.1. *Acquisition and design of plant geometry*

Participants: Chakkrit Preuksakarn, Mathilde Balduzzi, Frédéric Boudon, Christophe Pradal, Christophe Godin, Christian Fournier.

This research theme is supported by RTRA project 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 [48], [42]. 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, laser scanner, video and direct sketching.

- *Automated reconstruction of plant architecture* (Chakkrit Preuksakarn, Mathilde Balduzzi, Julien Diener, Frédéric Boudon, Jean-Baptiste Durand, Christophe Godin, Bernard Mourrain [Inria, Galaad], Franck Hetroy [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 developping 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 a second step, a skeleton procedure uses a Space Colonization Algorithm [47] 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 and a comparison with alternative methods has been carried out. Publication of this work is in progress.

  An automated reconstruction pipeline is also developed for processing 2D images of root system architecture (RSA) in the context of the Rhizopolis project. The analysis of these data is currently a major challenge in understanding root development. Existing tools either focus on specific applications, on simple structures (for example one root segment) or require long manual work. Here, we develop a processing pipeline that takes as an input 2D high resolution images of petri plates containing root systems. The pipeline makes it possible to extract from the images the whole architecture of root systems, with minimal or no user intervention. In order to obtain this result, the problem was decomposed in several steps: filter and label the input image, extract the image skeleton as a general graph structure and then convert it into a tree structure representing the visualised RSA, using a priori knowledge to solve inconsistencies. The pipeline has been added to the OpenAlea platform, thus allowing resulting data to be directly processed by other advanced high-level computational or statistical tools. The developed pipeline is currently being tuned and tested on several databases of 2D images with varying complexities of both arabidopsis and rice.

Figure 3. Reconstruction of a cherry tree. Left: photograph of the original tree. Right: 3D reconstruction from a laser scan rendered and integrated on the same background.

Additionnally, we also investigate the reconstruction of tree foliage from laser scanners in the context of Mathild Balduzzi's PhD thesis. Such elements are crucial to study the interaction of the plant with its environment. However, laser scans contain outliers on the silhouette of the scans that make the meshing of the point set difficult. New generation of laser scanners provides intensity of the laser reflected) on the surface of scanned objects. This intensity is dependant of the distance to the object, its optical property and the incidence angle. A first work on this topic shows that after correcting the distance effect, the incidence angle can be deduced from the intensity. From this result, we develop a reconstruction technique using the scan intensities and based on Shape-From-Shading approaches. The idea is to generate a new point set from the intensities and a set of seed points. This new point set has the property of being smooth but is not necessarily the exact representation of the scanned object. To consolidate the reconstruction, we are working on merging it with the orignal noisy point set coming from the scans using Kalman filtering. As a result, a final point set will be obtained without noise and with outiers naturally removed.

- *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 Imagine (ex-Evasion) we developed a method based on the design of plants from silhouettes [50]. 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 powerfull 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. It has been published Eurographics Symposium on Sketch-Based Interfaces and Modeling [32].

- *Reconstruction from video.* (Frédéric Boudon, Jérome Guenard [IRIT, Toulouse], Géraldine Morin [IRIT, Toulouse], Pierre Gurdjos [IRIT, Toulouse], Vincent Charvillat [IRIT, Toulouse])

Even if mature computer vision techniques allow the reconstruction of challenging 3D objects from images, due to high complexity of plant topology, dedicated methods for generating 3D plant models must be devised. We propose an analysis-by-synthesis method which generates 3D models of a plant from both images and a priori knowledge of the plant species.

Our method is based on a skeletonisation algorithm that allows to generate a possible skeleton from a foliage segmentation. Then, a 3D generative model, based on a parametric model of branching systems that takes into account botanical knowledge is built. This method extends previous works by constraining the resulting skeleton to follow hierarchical organisation of natural branching structure.

3D models are then generated. A reprojection of the models can be compared with the original image to assess the visual accuracy of the reconstruction. We optimise the parameter values of the generative model based on the reprojection criterion. Realistic results are obtained on different species of plants, in particular vineyards. Publication of this work is in progress.

- *Reconstruction of virtual fruits from pictures.* (Mik Cieslak, Nadia Bertin [Inra, Avignon], Frédéric Boudon, Christophe Godin, Michel Genard [Inra, Avignon], Christophe Goz-Bac [Université Montpellier 2])

This research theme is supported by the Agropolis project Fruit3D.

The aim of this work is to provide methods for generating fruit structure that can be integrated with models of fruit function and used to investigate such effects. To this end, we have developed a modeling pipeline in the OpenAlea platform that involves two steps: (1) generating a 3D volumetric mesh representation of the entire fruit, and (2) generating a complex network of vasculature that is embedded within this mesh. To create the 3D volumetric mesh, we use reconstruction algorithms from the 3D mesh generation package of the Computational Geometry Algorithms Library (CGAL). To generate the pattern of vasculature within this volumetric mesh, we use a Space Colonisation Algorithm that populates the volume of the fruit by simulating competition for space of the vasculature. We have applied our modeling pipeline to generate the internal and external geometry of a cherry tomato fruit using Magnetic Resonance Imaging data as input. These studies demonstrate the possibility to create species-specific models of fruit structure with relatively low effort [26]. These volumetric meshes are then combined with models of function to form integrative computational fruit models, which will help to investigate the effects of fruit structure on quality (see section 5.3.2).

- *Reconstruction of gramineous leaves.* (Christian Fournier, Christophe Pradal)

This research theme is supported by the Agropolis project OpenAlea.

Unlike trees, the 3D architecture of gramineous plants is much more related to the shapes of its leaves than the arrangement of its branches. Many modeling efforts have thus concentrated on correctly capturing its complex shape at different stages and use them as scalable geometric primitives. Still, additional control of such objects is needed in the context of Functional Structural Modeling. The objective of this work is to propose a plastic and dynamic 3D leaf model that is well suited for such uses, still able to capture a variety of observed static shapes. Leaf shape is modeled by a parametric surface describing leaf midrib curvature, leaf width variation, undulation of leaf margins and twist along the midrib. Meshes can be generated form these surfaces, and reduced using a decimation algorithm. The model can be fed with data or with curves drawn by user interaction. Morphological operators are defined and allows for plastic deformation of the control curves. The dynamics of shape acquisition can also be specified, and combined with morphological operators to simulate various scenarii of evolution and responses to stresses. The capabilities of the model are demonstrated through several cases of use. Future directions of research are thought to be a better integration of mechanical or physiological constraints that would reduce the model plasticity but avoid user-induced unrealistic simulation. [28].

5.1.2. *Modeling the plant ontogenic program*

Participants: Christophe Godin, Yann Guédon, Evelyne Costes, Jean-Baptiste Durand, Anaëlle ambreville, Pierre Fernique, Christophe Pradal, Jean Peyhardi, Catherine Trottier, Yassin Refahi, Etienne Farcot.

This research theme is supported by two PhD programs.

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. [37], [43], [44], [41], 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

program" [45], "age state" [40] or "physiological age" [38]. 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 [40], [38]. Here we develop computational methods to decipher these rules.

- *Relating branching structure to the shoot properties* (Jean Peyhardi, Yann Guédon, Evelyne Coste, Catherine Trottier, Yves Caraglio [AMAP], Pierre-Eric Lauri [AGAP, AFEF team]) Shoot branching structures often take the form of a succession of homogeneous branching zones and have been analyzed using segmentation models such as hidden semi-Markov chains. Axillary meristem fates are influenced by local properties of the parent shoot such as for instance its growth rate or local curvature. The objective of this work, which is part of the PhD subject of Jean Peyhardi, is to develop statistical models that generalize hidden semi-Markov chains with the capability to incorporate explanatory variables that vary along the parent shoot (e.g. maximum growth rate of the leaf, surface of the leaf, length of the internode, local curvature of the parent shoot). More precisely, the simple multinomial distributions that represent the axillary productions observed in the different branching zones are replaced by multinomial generalized linear models (GLMs). Since the two classical categories of multinomial GLMs that correspond either to nominal or ordinal categorical response variables were not appropriate in our setting, we chose to develop a new family of multinomial GLMs called multi-step multinomial GLMs that enable to tackle partially ordered categorical response variables. Typically, we need to distinguish different timing of branching (e.g. immediate shoot, one-year-delayed shoot and latent bud), different categories of offspring shoots (e.g. among one-year-delayed shoots, vegetative short shoot, vegetative long shoot and flowering shoot) and to specialize the explanatory variables for certain categories of offspring shoots (e.g. the growth of the parent shoot influence the immediate offspring shoots but not the one-year-delayed offspring shoots). The resulting integrative models are called semi-Markov switching multi-step multinomial GLMs and are applied to different data sets corresponding mainly to fruit tree branching structures.

- *Genetic determinisms of the alternation of flowering in apple tree progenies.* (Jean-Baptiste Durand, Jean Peyhardi, Baptiste Guitton [AGAP, AFEF team], Yan Holtz [AGAP, AFEF team] Catherine Trottier, Evelyne Costes, Yann Guédon)

  The aim of this work was to characterize genetic determinisms of the alternation of flowering in apple tree progenies. Data were collected at two scales: at whole tree scale (with annual time step) and a local scale (annual shoot or AS, which is the portions of stem that were grown during the same year). Two replications of each genotype were available.

  Indices were proposed for early detection of alternation during the juvenile phase. They were based on a trend model and a quantification of the deviation amplitudes and dependency, with respect to the trend. This allows for quantifying alternation from the yearly numbers of inflorescences at tree scale.

  However, phenotyping subsamples of AS sequences is more realistic in the framework of genotype selection. To model alternation of flowering at AS scale, a second-order Markov tree model was built. Its transition probabilities were modeled as generalized linear mixed models, to incorporate the effects of genotypes, year and memory of flowering for the Markovian part, with interactions between these components. Asynchronism of flowering at AS scale was also assessed using an entropy-based criterion.

  This work started during the PhD's work of Baptiste Guitton. It was then extended in 2012 by Yan Holtz during this Master 2 internship, supervised by Evelyne Costes and Jean-Baptiste Durand. New progenies were considered, as well as the performance of approximating the descriptors at whole tree scale with those at AS scale. These descriptors allowed the identification of QTL zones involved in the control of flowering in apple trees.

  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 (see next item in Section 5.1.2).

- *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 exploring the space of possible chain graph models, which required defining a notion of neighbourhood of these graphs. A parametric distribution was associated with each graph. It was obtained by combining families of univariate and multivariate distributions or regression models. These were chosen by selection model procedures among different parametric families.

This work was carried out in the context of Pierre Fernique's first year of PhD (Montpellier 2 University and CIRAD). 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 (see previous item in Section 5.1.2). It was also applied to the analysis of the connections between cyclic growth and flowering of mango trees. This work will be continued during Pierre Fernique's PhD thesis, with extensions to other fruit tree species and other parametric discrete multivariate families of distributions, including covariates and mixed effects.

- *Analyzing fruit tree phenology* (Anaëlle Dambreville, Jean-Baptiste Durand, Pierre Fernique, Yann Guédon, Christophe Pradal, Pierre-Eric Lauri [AFEF team, AGAP], Frédéric Normand, 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. We are now investigating jointly structure development and phenology of mango using statistical models for trees in particular hidden Markov tree models and multitype branching processes.

- *Integrative developmental growth stages of shoots* (Anaëlle Dambreville, Yann Guédon, Pierre-Eric Lauri [AFEF team, AGAP], Frédéric Normand) Growth and development are often studied as two separated processes. Our aim is to investigate the coordination between growth and development in mango shoots. We considered three types of organ, namely the shoot axis, its attached leaves and the inflorescence. Two types of data were collected during the shoot and inflorescence follow-up: developmental stages determined in an expert way and organs sizes determined from measurements. To give an integrative view of the shoot and inflorescence growth and development, we adopted the following strategy. For a given cultivar, we first built a multi-state model on the basis of absolute growth

rate sequences deduced from the measurements. Using these models, we computed growth stages. These growth stages highlighted growth asynchronisms between two topologically-connected organs: the axis and its leaves. Then, we compared these growth stages with the developmental ones and we obtained strong matches between them. The integrated developmental growth stages emphasized that the developmental stages are markedly related to growth rates an can be interpreted in terms of physiological (hydraulics, carbohydrates partitioning) and developmental (organs preformation versus neoformation) processes.

- *Self-nested structure of plants.* (Christophe Godin, Farah Ben Naoum) In a previous work [6], we designed a method to compress tree structures and to quantify their degree of self-nestedness. This method is based on the detection of isomorphic subtrees in a given tree and on the construction of a DAG, equivalent to the original tree, where a given subtree class is represented only once (compression is based on the suppression of structural redundancies in the original tree). In the compressed graph, every node representing a particular subtree in the original tree has exactly the same height as its corresponding node in the original tree. This method thus compresses a tree in width, but not in height. In this new work, we designed an extension of this compression method in which a tree is compressed in both width and height. The method is based on the detection of so-called *quasi-isomorphic paths* in a tree and on the compression of these paths in height. A paper describing the corresponding algorithms is being written.

5.1.3. *Analyzing the influence of the environment on the plant ontogenic program*

Participants: Frédéric Boudon, Jean-Baptiste Durand, Christophe Godin, Yann Guédon, Jean Peyhardi, Pierre Fernique, Maryline Lièvre, Christine Granier, Evelyne Costes, Pascal Ferraro, Catherine Trottier.

This research theme is supported by three PhD programs.

The ontogenetic program 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 program 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.

- *Influence of environment conditions and horticultural practices on the branching and axillary flowering structures of fruit tree shoots.* (Yann Guédon, Evelyne Costes [AFEF Team, AGAP], Ted DeJong [UC Davis], Claudia Negron [UC Davis]).

  In the context of a collaboration with Claudia Negron and Ted DeJong, we studied the influence of water availability and pruning practices on the branching and axillary flowering structures of different categories of almond shoots Stochastic models (hidden semi-Markov chains) were built for the branching and axillary flowering structures of different categories of almond shoots corresponding to different genetic backgrounds, levels of irrigation and pruning practices.

- *Analyzing growth components in trees.* (Yann Guédon, Yves Caraglio [AMAP], Olivier Taugourdeau [AMAP])

  In a forest ecology context, we identified robust indicators that summarize the balance between tree ontogeny and environmental constraints (mainly related to light environment). In this context, tree growth data typically correspond to the retrospective measurement of annual shoot characteristics (e.g. length, number of branches) along the main stem. We applied segmentation models (hidden Markov and semi-Markov chains) that enable to identify tree growth phases. This statistical modeling approach was applied to both deciduous (sessile oak and Persian walnut) and evergreen (Corsican pine and silver fir) tree species growing in contrasted conditions ranging from managed forest stands

to unmanaged understoreys. The growth phase duration distributions estimated within these segmentation models characterize the balance between tree ontogeny and the environmental constraints in tree development at the population scale. These distributions had very contrasted characteristics in terms of shape and relative dispersion between ontogeny-driven and environment-driven tree development. The characteristics of growth phase duration distributions may change over tree life reflecting changes in tree competition.

- *Investigating the influence of geometrical traits on light interception efficiency of trees and grass* (Liqi Han [AFEF Team, AGAP], Christophe Pradal, Frédéric Boudon, Christophe Godin, David Da Silva [UC Davis], Evelyne Costes, Philippe Balandier [PIAF], André Marquier [PIAF], Gaëtan Louarn [URP3F], Didier Combes [URP3F], Christian Fournier)

Light availability in forest understory is essential for many processes. It controls for instance the growth potential of species and individuals in plant communities. It is, therefore, a valuable information regarding forest and crop management. However, the effects of competition for light on short term vegetation dynamics are still poorly understood. This is in part due to a lack of tractacle and precise methods to estimate light resource within a canopy. To alleviate this difficulty, models can be used to compute light interception. At a detailed scale, they often require a lot of field data to accurately predict light distribution, particularly in the case of heterogeneous canopies.

To investigate this issue, we first analyzed the deterioration of the prediction quality of light distribution to the reduction of inputs by comparing simulations to transmitted light measurements in forests of increasing complexity in three different locations [17]. With a full set of parameters to describe the tree crown (i.e., crown extension in at least eight directions, crown height and length), the model accurately simulated the light distribution. Simplifying crown description by a geometric shape with a mean radius of crown extension led to deteriorated but acceptable light distributions. Allometric relationships used to calculate crown extension from trunk diameter at breast height seriously reduced light distribution accuracy.

We also studied the light interception of herbaceous plants with contrasting architectures (monocultures and binary mixtures) grown at high or low density and sought to determine the important architectural features necessary to account for light partitioning among individual plants [21]. It was shown that the studied plant populations were typical of a wide range of competition intensities, ranging from sparse plants to dense size-structured populations. Plant representations using whole plant envelopes with homogeneous leaf area density (LAD) were not reliable to estimate light partitioning, irrespective of the accuracy of envelope definition. Accounting for heterogeneous LAD within plants helped to solve this problem in both sparse and dense canopies. The relative importance of traits however changed with competition intensity and was different from reports made on isolated plants. Simple envelope-based reconstructions were finally shown robust enough to support parameterisation from a tractable set of traits measured in the field provided that height and vertical LAD gradient were characterised.

Using virtual growth simulation tools, a detailed analysis was also carried out to study more precisely the influence of architectural variability of apple trees on their light interception efficiency [30]. For this we used MAppleT, an in silico functional-structural plant model that has been built for simulating architectural development of apple trees. The STAR, namely the silhouette to total area ratio, of leaves, was chosen to evaluate the level of such efficiency. The strategy was to integrate MAppleT with the light interception model provided by the fractalysis module of the VPlants software library. Target values of four major traits (internode length, leaf area, branching angle and top shoot diameter), are varied in range previously observed in a segregating population of apple hybrids. A sensitivity analysis based on polynomial and generalised additive models was performed for highlighting the most influential trait on light interception and suggesting the optimal combination(s) of traits leading to the highest STAR. The contribution of stochastic processes that pilot tree topology in MAppleT is also investigated in the sensitivity analysis. This study not only

provides a time- and resource-saving alternative for data collection, but also sets a methodology for ideotype definition and further genetic improvement of apple trees.

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, Christophe Pradal, Vincent Mirabet [RDP, ENS], Jan Traas, Grégoire Malandain, Jean-Luc Verdeil [PHIV, AGAP].

This research theme is supported by the iSam and Morphogenetics projects.

- *Improvement of the MARS-ALT pipeline robustness* Meristem, laser microscopy, image reconstruction, cell segmentation, automatic lineaging
  Participants: Léo Guignard, Christophe Godin, Grégoire Malandain, Jan Traas, Pradeep Das [RDP, ENS], Vincent Mirabet [RDP, ENS].

  The MARS-ALT (Multi-Angles Registration and Segmentation - Automatic Lineage Tracking) software pipeline automatically performs a segmentation at cell resolution from 3D or 2D voxel images where the membranes/walls are marked (by a die for example) and makes it possible to follow the lineage of these cells through time [5]. A new version of this pipeline is currently being developed. MARS-ALT Version 2 is based on the same algorithms and methods and is intended to improve the overall robustness of the pipeline (protocol, noise in the input image) and automate completely the process. To test the new pipeline, we use different acquisition protocols and different organisms (floral and apical meristems and the early stages of development of a marine animal *Phallusia mammillata*). The segmentation is corrected a posteriori to deal with imaging artifacts due to uncertainties of acquisition. The image data set on which we develop the methods consists of :

  - *Arabidopsis thaliana* shoot apical meristem and primordia with around 6000 cells. The organ is captured from three different angles every 20 hours 3 or 4 times with a confocal microscope (Collaboration Sainsbury lab, Cambridge)

  - *Arabidopsis thaliana* flower meristems with around 2000 cells. The organ is also captured from three different angles every 20 hours 3 to 5 times with a confocal microscope (Collaboration RDP Lyon)

  - *Phallusia mammillata* and *Ciona intestinalis* embryos with from 32 cells to around 1000 cells. The organism is captured from four different angles every 2 minutes during 2 to 3 hours with a SPIM (Single Plane Illumination Microscope) (Collaboration CRBM Montpellier / EMBL Heidelberg)

  The pipeline provides as an output segmented images on which metrics for each cells can be extracted such as volume, principal components, convex hull and so on. A new non-linear registration algorithm developed by G. Malandain (MORPHEME team, Inria Sophia-Antipolis) is now available and will lead to an improvement of ALT algorithm. Redesign and improvement of the lineage tracking pipeline will be the next step.

- *Design of 3D virtual atlases for specifying gene expression patterns* (Jérôme Chopard, Christophe Godin, Jan Traas, Françoise Monéger [RDP, ENS])

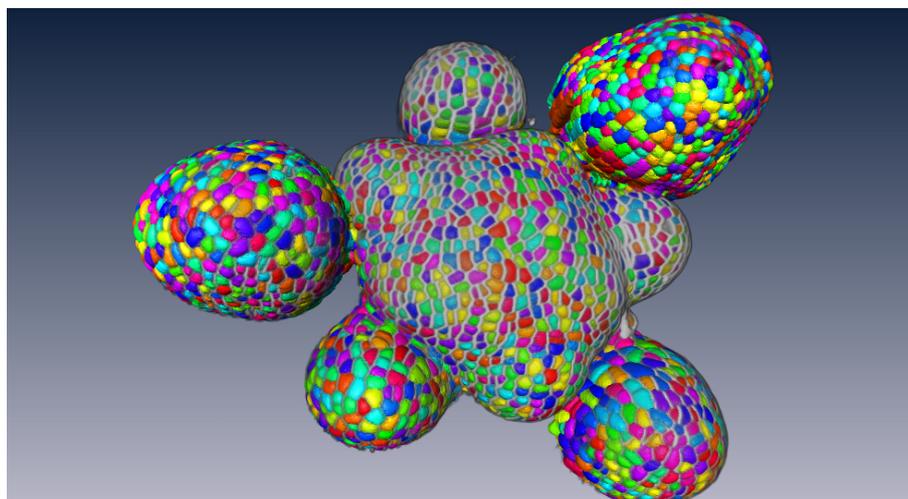 *This research theme is supported the ANR GeneShape and iSam projects.*

Figure 4. Superimposition of an automatic cell segmentation of an arabidopsis flower meristem using the new MARS pipeline with the original confocal image stack where the membranes are marked.

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 atlas is mainly descriptive. However, like for classical databases, specific tools make it possible to explore the virtual atlas 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 atlas.

A prototype version of the 3D virtual atlas was built last year [8]. Further developments of this tool will rely on the segmented images produced from microscopy, as presented in the previous section. In particular, a common underlying data structure has to be developed transversally to these two scientific developments. The definition of this data structure has been initiated last year through several team meetings, and should lead to a revised implementation next year.

5.2.2. *Shape analysis of meristems*

(Jonathan Legrand, Clémence Hatt [BURST, AGAP], Jean-Baptiste Durand, Frédéric Boudon, Christophe Godin, Yann Guédon, François Mankessi [BURST, AGAP], Olivier Monteuuis [BURST, AGAP], Jean-Luc Verdeil [PHIV, AGAP])

Plants that grow several forms or type of leaves along a shoot, depending on age or shoot length, are called heteroblastic.The influence of heteroblasty 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 heigh (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. A paper presenting these results has been published in Trees [20]

At cellular resolution, we studied the organization of cells in the meristems. The MARS-ALT pipeline provides rich spatio-temporal data sets for analyzing the development of meristems. A first step consisted of designing a dedicated graph for efficiently representing the spatial (adjacency between cells) and temporal (cell division) relationships between cells. Various variables can be attached either to the vertices (e.g. cell volume, inertia axes) or the edges (e.g. wall surface, distance between cell centroids). This graph may be augmented by new variables resulting from various spatial or temporal filtering (e.g. cell volumetric growth). We are now designing models and algorithms for finding patterns in time courses of meristems. In particular, we are investigating spectral clustering methods in order to define homogeneous regions in term of cell identities.

5.2.3. *Transport models*

Participant: Michael Walker.

This research theme is supported by the ANR GeneShape and ERASysBio+ iSAM projects and Morphogenetics.

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 [46] [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 different patterns of PIN allocations. In the last two years, we developed a unifying model showing that a unique flux-based model could be sufficient to explain PIN patterns in both L1 and L2 [25]. Contrary to our previous study [11], here no change in the model parameters is needed for this. Our approach is based on inherent topological and geometrical differences between the L1 and L2, specifically their dimensionality and the distribution of sources and sinks.

In a different perspective, another study on auxin transport models have been submitted this year. In this work, a generic, adimensional flux-based model of auxin transport was studied using a combination of analytic and numeric approach. The steady-states with uniform auxin distribution where characterised for arbitrary tissues, and some of their bifurcations (loss of stability and Hopf) were described. This work, initiated during an "Explorateur" project funded by Inria during the period October 2012-January 2013, was submitted by E. Farcot and Y. Yuan (Memorial University of Newfoundland, Canada) in September and is still under review.

5.2.4. *Mechanical model*

Participants: Jérôme Chopard, Olivier Ali, 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 together with the Inria project Morphogenetics and the ERC from Jan Traas.

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 changes of shape in plant tissues during organogenesis, we need a mechanical model of tissue development at cellular resolution. 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 cell wall synthesis that is triggered when wall deformation exceeds a particular threshold. 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 how cell walls should yield to mechanical stresses. In this way, different regions in the tissue with different cell identities can have different growth properties. The final shape of the tissue results from the integration of all these mechanical properties and stresses at organ level under the growth force due to turgor pressure at tissue scale.

A paper describing the mechanical model and its application to model primorium formation in the shoot apical meristem is currently being written. Additionally, a redesign of our mechanical model using the SOFA framework is in progress.

5.2.5. *Gene regulatory networks*

Modeling gene activities within cells is of primary importance since cell identities correspond to stable combination of gene expression. This fact is becoming more and more acknowledged, and has led this year to the publication of several review articles including members of Virtual Plants as authors [22], [23], [24], [13].

- *The auxin signaling pathway* (Etienne Farcot, Yann Guédon, Christophe Godin, Yassin Refahi, Jonathan Legrand, Jan Traas, Teva Vernoux, Stéphane Robin [AgroParisTech], Jean-Benoist Leger [AgroParisTech])

  The auxin signalling network involves about 50 potentially interacting factors. We applied a graph clustering method [12] 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 study 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 new model evidences different behaviors between the ARF+ and the Aux/IAA factors.

  Extensions of this ODE model are necessary to better understand this system in more general contexts than the shoot apical meristem development, which was the framework of our previous study. This work involves defining and studying mathematically a series of distinct network topologies for the auxin signalling pathway. This was the topic of a Master's internship for the studend Cyril Lavedrine, from September to December 2012, co-supervised by E. Farcot and T. Vernoux. This work led to qualitative predictions which remain to be tested experimentally.

- *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 long lasting collaboration between E. Farcot and R. Edwards (Univ. Victoria, Canada), with more 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 has appeared this year [18].

For the same class of piecewise-linear models, it is in general very difficult to entirely characterize the attractors of a given system. In an attempt to improve our ability on this question, a probabilistic approach has been proposed in [15], in which it is shown that a Markov chain can built as an approximation of a given piecewise-linear system, and actually used to make predictions about its periodic attractors.

At a higher scale, we have also continued the study of gene regulation in meristematic tissues. In the context of Y. Refahi's post-doc between Virtual Plants and the group of Henrik Jönsson in Cambridge (Sainsbury Laboratory), we have continued a work that was initiated in Y. Refahi's thesis. This work is motivated by recent biological results, indicating that gradient-like patterns originating from the external layers of meristems may play a decisive role in the specification of the pool of stem cells in a central position. Using the methods in [5], and their on-going improvements, we have acquired new 3D and 4D images that were then segmented. These structures will be used in the next few months to investigate generic patterning properties of gradient like morphogen patterns. This will require a thorough analysis of free diffusion in realistic geometries, as made possible by the newly acquired images. As a preliminary work, we are also currently investigating the formation of gradient patterns in idealised tissues, allowing for deeper analytic treatment than the complex structures obtained by microscopy.

5.2.6. *Model integration*

Participants: Mikaël Lucas [IRD], Michael Walker, Jérôme 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 of 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. This work is developed in collaboration with the RDP grou at ENS-Lyon [23] and the CPIB group in Nottingham, UK [13].

- *Development of a computer platform for the 'programmable tissue'.* (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. In the past year, progress has been made in defining a generic tissue data structure that could be used in this platform, through several group meetings along the year. A redesign of the structure is in progress.

- *Design of a genetic model of inflorescence development.* (Etienne Farcot, Christophe Godin, François Parcy)

  We studied the regulatory network that control flower development during morphogenesis. To overcome the network complexity and integrate this regulation during ontogenesis, 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 repeatedly 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 mixed model has led us to make different hypotheses about gene interactions and hormonal regulation. First predictions about gene actors controling the passage to flower could be verified. However, a complete integrated picture of flower development could not be reached yet. After several unsuccessful attempts, further experiments are currently being made to verify the scenario predicted by the model.

5.3. Multi-scale models and analysis: from cells to plant architecture (and back)

5.3.1. *Transport model in roots*

Participants: Mikaël Lucas [IRD], Christophe Pradal, Christophe Godin, Christophe Maurel [BPMP].

This research theme is supported by the ANR project HydroRoot.

A model of Arabidopsis thaliana root hydraulics at the cellular level was developed in the OpenAlea modeling platform. The model relies on the integration throughout root architecture of elementary hydraulic components. Each component integrates local radial and axial water flows. Axial hydraulic conductivity is calculated according to Poiseuille's law, based on local size of xylem vessels. Radial hydraulic conductivity is determined in part by aquaporin activity and was set constant throughout root architecture in the first model versions. In its current state, the model is parameterized using architectural, tissular and physiological data that were experimentally determined in the Aquaporin group at BPMP. The architectural reconstruction of the root system is based on a tridimensional multi-scale tree graph (MTG). The current model is capable of predicting the water flow that is transported by a root system in the standard experimental conditions used in the Aquaporin group. This model was used to perform sensitivity analyses and determine the respective contributions to root hydraulic dynamics of various biological parameters (axial and radial hydraulic conductivites, root architecture). One major finding is that the root hydraulic conductivity (Lpr) computed from the model is highly dependent on root architecture. This is due to the limiting role of axial (xylem) conductance, one feature that had been neglected in previous representations of root water transport. The radial hydraulic conductivity may primarily be limiting in conditions of Lpr inhibition, since its increase from values in control roots has marginal effects on Lpr. A new set of experimental data including root diameter repartitions in wild-type plants, and xylem vessel diameters in mutants with altered xylem morphology (irx3, esk1) will be used to implement the model. Root cell hydraulic conductivities will also be measured in these and aquaporin mutant phenotypes. Our aim is to check whether, based on anatomical and morphological data, the model can properly predict the radial hydraulic conductivity of these genotypes.

5.3.2. *Transport in fruits*

Participants: Mik Cieslak, Nadia Bertin [Inra, Avignon], Frédéric Boudon, Christophe Godin, 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 results from 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 ignored to a large extent the shape and internal structure of the fruit.

Two essential functions in determining fruit quality are the transport and accumulation of water and dry matter to various fruit tissues. Since water and carbon are delivered to fruit tissues through a complex vasculature system, the internal fruit structure and pattern of vasculature may have a significant impact on their distribution within 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, a modeling pipeline has been developed that includes the following steps: creation of a 3D volumetric mesh of the internal fruit structure, including vasculature (see section 4). Based on previous compartment models of fruit physiology developed at Avignon, we have then developed models of water and carbon transport that have been coupled with the 3D model of fruit. In the 3D model, different equations are describing the transport between adjacent regions of the fruit represented as a 3D mesh. The integration through space and time is carried out using a standard integration scheme (Runge-Kutta of order 4).

This approach has been applied to study tomato fruit (Solanum lycopersicum) by constructing 3D volumetric meshes from different sources (images of perpendicular fruit slices and MRI data), and integrating water and carbon transport processes into these meshes. To illustrate the tomato model, a simulation of one season of the fruit's growth has been performed and its results compared with an already published process-based tomato fruit model. We first showed that the classical results of the abstract process-based models could be also captured by the more detailed spatialized model. However, our model provides additional information on the internal heterogeneity 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.
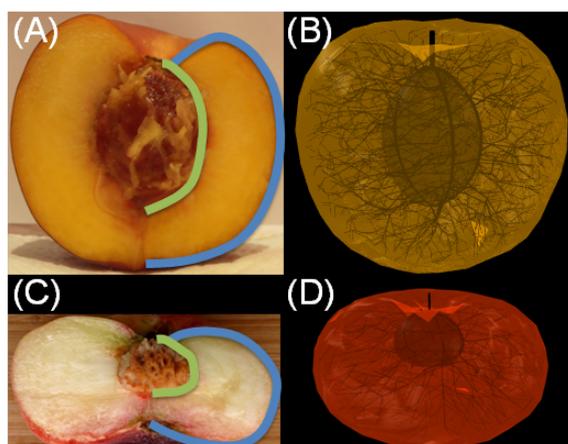

Figure 5. Virtual models of peaches reconstructed from images with simulated vasculatur to simulate carbon and water transport in the fruit

5.3.3. *Analazing shoot and leaf elongation*

Participants: Maryline Lièvre, Yann Guédon, Christine Granier.

The analysis of phenotyping data coming from automated platforms such as PHENOPSIS often focuses on the growth of a leaf at a given rank along the stem. We aim at developing a pipeline of methods for analyzing the growth of *arabidopsis* shoot at three scales:

1. tissular scale using a probabilistic model of endoreduplication for modeling the distribution of the leaf epidermis cell surfaces. Endoreduplication, which is a replication of the nuclear genome in the absence of cell division that leads to elevated nuclear gene content, strongly affects the leaf epidermis cells of *arabidopsis*.

2. organ scale using nonlinear regression model for analyzing the growth of each successive leaf.

3. shoot scale: The outputs of the analyses at the tissular and organ scales will be summarized as multivariate sequences along the shoots characterizing each successive leaf. These sequences will be augmented by supplementary morphological variables characterizing leaf shape and properties (e.g. presence/absence of trichomes). These sequences will be globally analyzed in order to take into account plant ontogeny and in particular the successive developmental stages before the floral transition for the wild type and selected mutants of *arabidopsis*.

5.3.4. *Analyzing perturbations in Arabidopsis thaliana phyllotaxis*

Participants: Christophe Godin, Yann Guédon, Yassin Refahi, Etienne Farcot.

This research theme is supported by iSAM.

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 thalianaahp6* 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 relies of combinatorial and statistical techniques. 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 organs 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. Two papers (one with biological aspects and the other about methodological developments) are currently in revision.

WIMMICS Team

6. New Results

6.1. Linked Data Access

Participants: Serena Villata, Luca Costabello, Fabien Gandon.

We designed and developed a context-aware access control framework for the Web of Data called Shi3ld [7]. The framework protects access to SPARQL endpoints, and it adopts Semantic Web languages only, as in the philosophy of the Web of Data. The innovative feature of the proposed framework consists in evaluating the accessibility to the data considering the attributes of the users. These attributes are defined following three main dimensions: user, device, and environment. The evaluation of the model shows that access control comes with a cost but it guarantees the protection of the data published on the Web of Data. The results of this research activity have been published in international conferences in the area of Artificial Intelligence (ECAI, [35]) and the Web (WWW [61], HT [34]).

On the same line, we have proposed a framework for attaching the licenses to the data resulting from a query on the Web of Data. The rationale is that the licenses associated to the data returned by the query are selected, and using a number of rules their compatibility is assessed. If the licenses are evaluated compatible, then they are composed into a composite license which is released to the user together with the data. The results of this research have been published in the COLD international workshop [74] and in the ISWC international conference [75] (best poster award). These two research lines have been performed in the context of the DataLift ANR project.

The PhD thesis of Luca Costabello, directed by F. Gandon and I. Herman (CWI and Semantic Web Activity Lead at W3C) investigates Web of Data interaction from mobile environments. Two main research activities have been carried out in 2012: i) PRISSMA [8], an adaptive rendering engine for RDF and ii) Shi3ld[9], a context-aware access control framework for Linked Data.

The goal of PRISSMA is delivering an *adaptive* rendering engine for Linked Data resources. PRISSMA tweaks RDF visualization to the context in which the resource consumption is performed. Work in 2012 has been mainly focused on determining the algorithm that selects the best RDF visualization according to the real, sensed context. The uncertain and incomplete nature of context data, led to investigate strategies that model the task as an inexact RDF instance matching operation.

The second main research line carried out in 2012 led to the creation of Shi3ld, an access control framework for Linked Data SPARQL endpoints. Shi3ld authorization procedure and Access Policies, defined using Semantic Web languages only, have been enriched with the notion of mobile context, thus enabling context-based access control (e.g. geo-temporal authorization policies).

A collateral research line has been carried out as the follow-up of the 9th edition of the Summer School on Ontology Engineering and the Semantic Web [10]. We contributed to an exploration work on the problem of spamming in Linked Data, providing a classification of potential spamming techniques and populating and publishing a dataset containing spammed triples[11]. The dataset is useful to train anti-spamming mechanisms.

6.2. ISICIL Platform

Participants: Nicolas Delaforge, Michel Buffa, Fabien Gandon, Alain Giboin.

---

[7]http://wimmics.inria.fr/projects/shi3ld/

[8]http://wimmics.inria.fr/projects/prissma

[9]http://wimmics.inria.fr/projects/shi3ld

[10]http://sssw.org/2012/

[11]http://www-sop.inria.fr/members/Luca.Costabello/spam/

ISICIL is an ANR project studying social networks and Semantic Web communities to support corporate intelligence.

First, ISICIL proposes a multidisciplinary design of a new form of corporate intelligence. The challenge of this project is to reconcile the new viral Web applications with formal representations of business processes and to integrate them into practical intelligence communities of actors working in the company. We designed, studied and experimented with new tools to support collaborative tasks in business intelligence by leveraging Web 2.0 interfaces (blog, wiki, social bookmarking) for interactions and Semantic Web technologies for interoperability and information processing. ISICIL also allowed to explore new scientific developments of the notion of epistemic cooperation (human interaction oriented toward the development and transmission of knowledge) and to identify usable technological solutions. An ergonomic approach, combining impregnation of ground truth data and freer technological inspirations from bibliographic and webographic sources, was proposed.

Secondly, ISICIL uses typed graphs as models underlying epistemic communities. The entire model relies on a unifying model based on RDF graphs to represent resources and community stakeholders. These models are integrated with bookmarking tools or "Web scraping" the results of which are tagged. The tags used are collected to form folksonomies and a semi-automatic thesaurus structure in these folksonomies. User feedback on this structure is captured when they use the search engine which offers tags related to their keywords and the user can accept, reject or adjust these suggestions. User profiles and links between them, considered as a network, are processed by a series of operators to propose a semantic analysis of social network, for example the computation of indicators of centrality parameterized by an ontology. Merged graphs of structured folksonomies and of social networks finally allow the detection and labeling of epistemic communities. Meanwhile we study how the analysis of user interactions can determine the trust and how to represent and control access to data and their semantics in a social network.

As part of the paradigm of Social Epistemology, ISICIL combined in an ergonomic approach, impregnation of ground truth data, a bibliometric study and technological inspirations to offer patterns and inferences exploiting the Semantic Web social networks to assist corporate intelligence. An open-source platform is available under CeCILL-C licence and was tested at ADEME and Orange. ISICIL is a proof of concept of the compatibility of the Semantic Web formalisms, practices and models of Web 2.0 and the philosophical framework of social epistemology.

The project has resulted in three PhD theses (Florie Bugeaud 2011, Guillaume Erétéo 2011 [98], Freddy Limpens 2010 [99]) and publications at the following conferences: ISWC 2009 IEEE / WIC / ACM Web Intelligence 2011, Web Science 2010, WWW 2011, ASWC 2009, COOP 2010, PRO-VE 2009, VSST 2010, EGC 2010, IC 2009 & 2011, Psycho Ergo Days 2010.

The ISICIL project participated to Colloque ANR STIC, January 4-6 2012.

6.3. Natural Language Processing

Participants: Elena Cabrio, Julien Cojan, Fabien Gandon, Maxime Lefrançois, Serena Villata.

We have proposed a combined framework of natural language processing and argumentation theory to support the users in their interactions within online debate platforms. The framework combines a natural language processing module which exploits the textual entailment approach and detects the arguments in natural language debates and the relationships among them, and an argumentation module which represents the debates as graphs and detects the accepted arguments. The results of this research have been published in one of the major conferences in the field of Artificial Intelligence (ECAI [30]), and in the areas of natural language processing (ACL [28]) and argumentation theory (COMMA [29]).

To enhance users interactions with the Web of Data, query interfaces providing an extensible mapping between natural language expressions, and concepts and relations in structured knowledge bases are becoming particularly relevant. As a result of the first year of her postdoc, Elena Cabrio (together with Julien Cojan) designed QAKiS (Question Answering wiKiframework-based System), that allows end users to submit a query in English to an RDF triple store and obtain the answer in the same language, hiding the complexity of the

non intuitive formal query languages involved in the resolution process. At the same time, the expressiveness of these standards is exploited to scale to the huge amounts of available semantic data.

In its current implementation, QAKiS addresses the task of Question Answering (QA) over structured knowledge bases (e.g. DBpedia) where relevant information is expressed also in unstructured form (e.g. Wikipedia pages). Its major novelty is to implement a relation-based match for question interpretation, to convert the user question into a query language (e.g. SPARQL). A demo of the system is available online [12]. The results of this research have been published as a demo paper in the main conference of Semantic Web, ISWC [57], and in the special issue of the journal Intelligenza Artificiale [14]. QAKiS has been evaluated with respect to state-of-the-art systems taking part into the QALD-2 (Question Answering over Linked Data) challenge at ESWC, obtaining satisfactory results [59].

In the PhD Thesis of Maxime Lefrançois, we are interested in bridging 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. In 2011 we introduced the ULiS project that aimed at designing a pivot-based NLP technique called Universal Linguistic System, 100% using the Semantic Web formalisms, and being compliant with the Meaning-Text theory [102].

We showed that neither Description Logics nor Conceptual Graphs suit our needs, so this Ph.D. now focuses on the formalization of the Unit Graphs mathematical framework that is conceived to fill the gap between the highly linguistically precise Explanatory Combinatorial Dictionaries of the Meaning-Text Theory and the Dependency Grammars, and the highly mathematically grounded model of the Conceptual Graphs.

Maxime finally joined the Multilingual-Web-LT W3C Working Group, and left it when the charter got revised.

6.4. Requirement Engineering

Participants: Zeina Azmeh, Isabelle Mirbel, Serena Villata.

Requirements engineering is an essential process of software engineering, during which the complete behavior of a software system can be defined. The success of this process plays a crucial role in the success of the whole software project. A key issue of requirements engineering is stakeholders participation, which is facilitated through the emergence of online collaborative working tools. These tools create new opportunities of practice regarding requirements elicitation. Nevertheless, they result in an information overload lacking structure and semantics. Consequently, requirements analysis and selection become more challenging.

Our current proposition is embodied in an approach based on Semantic Web languages as well as concept lattices to identify relevant communities of stakeholders depending on their past participation. These communities can be used to enable efficient decision-making and handling of requirements.

We exploited the idea of applying argumentation theory to deal with requirements engineering. In particular, the proposed framework detects consistent sets of goal-based requirements and maintains their consistency over time based on argumentation theory which allows to detect the conflicts among elements. More specifically, the framework relies on meta-argumentation, which instantiates abstract argumentation frameworks, where requirements are represented as arguments and the standard Dung-like argumentation framework is extended with additional relations between goal-based requirements. The results of this research have been published to the CLIMA international workshop [37].

6.5. Regulation Engineering

Participants: Khalil Bouzidi, Catherine Faron-Zucker, Olivier Corby.

Regulations in the Building Industry are becoming increasingly complex and involve more than one technical area, covering products, components and project implementations. They also play an important role in ensuring the quality of a building, and to minimize its environmental impact.

---

[12]http://dbpedia.inria.fr/qakis/

In a collaboration between CSTB and the I3S laboratory, we are carrying on research on the acquisition of knowledge from the technical and regulatory information contained in the REEF referential [13] and the automated processing of this knowledge with the final goal of assisting professionals in the use of these texts and the creation of new texts.

We are applying this work in CSTB to help industrials in the writing of Technical Assessments. The problem is how to specify these assessments and standardize their structure using models and adaptive semantic services.

A Technical Assessment (in French: Avis Technique ou ATec) is a document containing technical information on the usability of a product, material, component or element of construction, which has an innovative character. We chose this Technical Assessment as a case study because CSTB has the mastership and a wide experience in these kinds of technical documents.

In 2012, we were particularly interested in the modeling of the regulatory constraints derived from the Technical Guides used to validate the Technical Assessment. These Guides are regulatory complements offered by CSTB to the various industrials to enable easier reading of technical regulations. They collect execution details with a wide range of possible situations of implementations.

Our work aims to formalize the Technical Guides in a machine-processable model to assist the creation of Technical Assessments by automating their validation.

Our first contribution is the use of standard SBVR (Semantics of Business Vocabulary and Business Rules) and SPARQL to reformulate the regulatory requirements of guides on the one hand in a controlled language and on the other hand in a formal language

Second, our model incorporates expert knowledge on the verification process of Technical Documents. We have organized the SPARQL queries representing regulatory constraints into several processes. Each component involved in the Technical Document corresponds to an elementary process of compliance checking. An elementary process contains a set of SPARQL queries to check the compliance of an elementary component. A full complex process for checking a Technical Document is defined recursively and automatically built as a set of elementary processes relative to the components which have their semantic definition in the OntoDT ontology that we have designed.

Finally, we represent in RDF the association between the SBVR rules and SPARQL queries representing the same regulatory constraints. We use annotations to produce a compliance report in natural language to assist users in the writing of Technical Assessments.

As a result, we have designed a Semantic Web application to support and guide the process of writing Technical Assessment. The current version has allowed us to validate our approach. Also, we have developed a base of SBVR rules to describe business requirements of guides. This rule base is implemented in SPARQL.

6.6. Graph-based Knowledge Representation

Participants: Olivier Corby, Catherine Faron-Zucker, Fabien Gandon, Isabelle Mirbel, Adrien Basse, Oumy Seye.

We have designed a method to build pretty-printers for RDF Abstract Syntax Trees and Graphs, built on top of SPARQL Query Language. Pretty-print rules are written as SPARQL select-where queries. The *where* clause matches the target subtree to be printed and the *select* clause returns the pretty-printed statement using an external *kg:pprint* function. This function recursively calls the pretty printer, looking for appropriate pretty-print queries for the target subtrees.

We have designed a syntactic extension to SPARQL in order to ease writing pretty-printing rules. Below, an example of template for a SPIN like AST is shown:

```
template {
  "select " ?s
```

```
  "\n"
  "where "  ?w
}
where {
  ?in ast:select ?s ;
      ast:where  ?w
}
```

We were able to write a pretty-printer for a SPIN like complete SPARQL 1.0 AST with 19 templates.

We have designed an extension to our KGRAM SPARQL interpreter that enables to consider an RDF Graph directly as a Query Graph. This enables to compute RDF subgraph matching.

We have completed SPARQL 1.1 implementation with the final version of Property Path (PP) and federated queries (service & bindings). In order to be able to query a SPARQL 1.0 endpoint with PP, we have written a compiler that translates PP into basic graph patterns.

The work on KGRAM is published in [33].

Alban Gaignard from the I3S Modalis team has designed a distributed version of KGRAM to query remote triple stores, in the context of Semantic Federation of Distributed Neurodata. This work is published in [63], [64].

We have implemented a prototype of C-SET Commutative Replicated Data Type for RDF in Corese with Pascal Molli and Luis Ibáñez from U. Nantes [69].

6.6.1. *Extracting Graph Patterns to Characterize RDF Data Sources*

This work takes place in the PhD Thesis of Adrien Basse.

Many Semantic Web applications address the issue of integrating data from distributed RDF triple stores. There are several solutions for distributed query processing such as SPARQL 1.1 Federation, which defines extensions to the SPARQL Query Language to support distributed query execution. Such extensions make it possible to formulate a query that delegates parts of the query to a series of services, but one issue remains: how to automate the selection of RDF triple stores containing relevant data to answer a query. This is especially true in the context of the Linking Open Data project where numerous and very heterogeneous datasets are interlinked, allowing for interesting queries across several sources. To decompose and send queries targeting only relevant stores, we need a means to describe each RDF triple store, i.e. an index structure which provides a complete and compact description of the content of the RDF triple store.

To know the content of a RDF triple store, we proposed to use graph patterns as basic structures for index items. In this thesis we present an approach to extract these graph patterns from RDF triple store. For this purpose, we extend Depth-First Search coding (DFS) [104] to RDF labeled and oriented multigraphs and we provide a join operator between two DFS codes so as to sequentially build the different levels of the index structure.

Insertion or deletion of annotations in the triple store may cause changes to the index structure. To handle updates in triple store, we proposed a procedure to identify exactly the changes in the first level of the index structure and propagate them to the following levels. The DFSR (Depth First Search for RDF) coding makes it possible for us to efficiently manipulate graph patterns, but is difficult to read (succession of integer numbers). To facilitate the reading of our index structure, we propose a visualization user-interface and algorithms to turn a DFS code into a more legible format like RDF. Our algorithm relies on Corese/KGRAM [95]. We have tested our algorithm on many datasets. During the building of index structures we keep a set of data in order to help us to better understand the progress of our algorithm and improve it.

6.6.2. *Rules for the Web of Data*

This work takes place in the PhD Thesis of Oumy Seye.

We have characterized the subset of SPARQL that can be expressed in RIF and, conversely, we have searched for the maximal RIF dialect that can be expressed as SPARQL construct-where queries. This work is published in [71], [72] were we present the implementation of a RIF dialect with a SPARQL Rule Engine in Corese/KGRAM.

We have designed online services for RIF-BLD parsers for presentation syntax and XML syntax [14]. We have also done an online service for RIF-BLD translation into SPARQL and RDF [15].

6.7. Business Intelligence

Participants: Corentin Follenfant, Olivier Corby, Fabien Gandon.

This PhD Thesis is done with a CIFRE industrial grant from SAP Research.

Industrial Business Intelligence 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. Authoring these dynamic documents requires proficiency in technical domains like relational modeling and SQL for one to produce relevant content: end users therefore praise example-driven and information retrieval (IR) systems that help them reusing existing content. Such systems need common structured metadata to enable comparison, search, matching and recommendation of (parts of) documents.

As target data sources are mainly tabular or relational, queries executed to feed the dynamic documents are SQL or derivatives. In [62] we proposed to model these queries as RDF named graphs, and use the graphs as documents annotations. Queries are represented through their abstract syntax trees (AST) represented with RDF graphs. The SQL-specific modeling contribution can therefore be applied to any generic query language. We identified two desirable features for IR systems that deal with queries repositories: search and rewriting, the latter allowing further annotation as well as reconciliation of source relational entities against LOD (Linked Open Data) repositories. On this basis we evaluated SPARQL 1.1 to perform SQL query analysis, i.e. pattern-matching search or rewriting, using in particular property paths. Resulting SPARQL queries are intuitive and concise.

Next steps include a quantitative evaluation by extracting RDF representations from a repository of SQL-fed documents, the production of a library of SPARQL queries that perform generic IR operations against RDF-modelled SQL queries, a formalization of the modeling and operations to compare them with generic tree manipulation methods. In further work we plan to investigate rewriting queries from different languages modelled with language-specific abstract syntax trees to generic abstract syntax trees and experiment cross-language query comparison with SPARQL.

6.8. Fuzzy Knowledge Representation

Participant: Andrea Tettamanzi.

Andrea Tettamanzi has joined the Wimmics research team in September 2012, after winning a position as a full professor at the University of Nice-Sophia Antipolis. He got a PhD in Computational Mathematics and Operations Research in 1995 from the University of Milan with a thesis on evolutionary algorithms; he became assistant professor at the University of Milan in 1998 and associate professor at the same university in 2002.

His research interests focus on combining different methods of computational intelligence, namely evolutionary algorithms, fuzzy logic, and neural networks, to solve real-world problems in the presence of imprecision, noisy data, and ill-defined optimization criteria, but also on the management of vagueness and uncertainty in knowledge representation, the automatic extraction of knowledge from data, possibility theory and its application to belief revision and goal generation in cognitive agents.

---

[14] http://wimmics-ws.inria.fr/rifparser
[15] http://wimmics-ws.inria.fr/riftosparql

After joining Wimmics, Andrea Tettamanzi has continued work on previous collaborations with other members of the team, namely Serena Villata [76], and has begun exploring with the rest of the team several research axes that could benefit from his contribution.

6.9. Emotion Modeling

Participants: Franck Berthelon, Imen Tayari, Nhan Le Thanh, Peter Sander.

In the PhD Thesis of Imen Tayari, an algebraic vector representation model of emotional states was designed. This multidimensional model provides a powerful mathematical tools for the analysis and the processing of emotions. It permits to integrate information from different modalities(speech, facial expressions, gestures) in order to allow more reliable estimation of emotional states. Indeed our proposal aims at efficient recognition of emotional states even when they appear to be superposed or masked. Experiments show the efficiency of the proposed method in detecting basic emotion by giving hight recognition rate. This work is published in [39], [41], [43], [40], [42].

In the PhD Thesis of Franck Berthelon, we are working in the domain of affective computing to create an emotion sensitive system. Interaction between human and computer can be improved via such a system because emotion is so important in everyday communication. Our research focuses on serious gaming, particularly on enabling a user and a virtual character to "share" an emotion. The two main problems that arise are:

- How to detect a user's emotions given that the stimulus comes from a virtual environment?
- How to give feedback based on the user's current emotion?

We propose to model emotions as a complex system where data are retrieved from physiological sensors such as heart rate, EMG or EEG. We need to map the multi-sensor data back into a dimensional model of emotion space. Finally, we aim to have an effect on the user's emotional state by varying the stimulus received from the virtual environment. This puts the user into different emotional situations determined by the task to accomplish, with an accompanying effect on their ability to carry out the task.

We developed an application for experimentation purposes; it aims to implement our model using an EEG headset (Emotiv EPOC). This application allows us to generate an emotional map from a slide show of emotion annotated picture. Based on the created map and the real time EEG data, this application can compute a user's instantaneous emotion.

In addition to the first development, we reuse experimental data from MIT to validate our model in a more controlled way. We take the same data, features, signal processing and feature reduction algorithm but instead of using the k-nearest neighbors (KNN) classification algorithm we use our model to identify and annotate discontinuities that represents emotion state changes in accordance with Klaus R. Scherer hypothesis [103].

We are continuing work on validating our model with quantitative results and applying those results on a more realistic system with the application we have developed.

6.10. DBpedia in French

Participants: Julien Cojan, Fabien Gandon.

The purpose of the project *DBpedia in French* is to extract data from Wikipedia in French and publish it under structured format. Wikipedia content is mainly meant to be read by human and is not suited for use in applications. DBpedia publishes the data extracted from Wikipedia articles in RDF W3C standard for the Semantic Web [16] thus readily available for many applications. For instance, DBpedia is used to generate indexes for cultural resources (e.g. HdA-lab project [17]), it can also be used for mobile applications thanks to the geographic data it contains, or to answer natural language questions, etc.

---

[16]http://www.w3.org/RDF/
[17]http://hdalab.iri-research.org/hdalab/

The original version of DBpedia is focused on the English chapter of Wikipedia. Last versions also contain elements extracted from other chapters, but only when related to a page in English. Articles with no equivalent in English are skipped, leading to a significant number of pages being ignored and so a significant amount of data is lost. For instance, about 49 000 persons and 180 000 places described in the French chapter have no corresponding article in English and are then missing in the English DBpedia. Moreover, the description of the same topic can be different from one chapter to another, reflecting cultural diversity.

DBpedia in French publishes data extracted from the French Wikipedia in complement to the English DBpedia. Data are linked with the different chapters from the internationalization committee thus providing multilingual resources. In its release from October 2nd, DBpedia in French contains 130 million triples describing 1.3 million things, among them 260 000 places, 140 000 persons, 64 000 work pieces and 26 000 organizations.

This project is supported by the Semanticpedia collaboration platform [18] launched November 19th 2012 by Aurélie Filipetti, the French Ministry of Culture, Michel Cosnard, CEO of Inria, and Rémi Mathis, CEO of Wikimédia France. Inria currently hosts the project [19] and is the correspondent for the French chapter in DBpedia internationalization committee.

6.11. Co-Construction of Community Ontologies

Participants: Papa Fary Diallo, Isabelle Mirbel, Olivier Corby.

PhD Thesis on Co-Construction of Community Ontologies and Corpus in a Limited Technological Environment.

To refresh the memory of people and revive many stories that accompany the creation and daily life of different Africa territories, the establishment of an online sociocultural encyclopedia was conceived. It will be an online platform based on a Geographical Information System (GIS) enriched by a semantic layer allowing access to different information.

In the last decade, we have seen the rise of two visions of the Web: on one side the Social Web or Web 2.0, which places users at the heart of the Web, they are no longer spectator but become editor of the content of Web pages. On another hand, the Semantic Web proposes knowledge representations (ontologies) that allow machines (software agent) to better understand data on the Web. Both aims were often opposed, but there are a lot of work trying to combine these visions.

In our work, oriented in this direction, we will try to create a new point of view of Community concept within the Web. Community is a group of people who share a common set of values and interests. This shift of view allows us to address a specific community as an atomic entity and focus this time on the sharing of knowledge between communities. The second challenge is to combine Social Web and Semantic Web technologies. Using Semantic Web in our social network, we have a semantic layer that provides access to various information contained in the network. Furthermore, the Semantic Web opens up a semantic approach to social network analysis, which also allows extracting new knowledge.

In this thesis, we study the implementation of an online platform to build and share the collective memory of citizens and revive many stories by using a semantic layer. Semantic Social Network Analysis will allow us to present data in eye-catching way and in another view. This platform will be updated by the actors and the citizens of these territories, and share their history and heritage through their "social network".

6.12. Semantic Wiki

Participants: Pavel Arapov, Michel Buffa.

A Wiki is a Web site that lets users create and edit content collaboratively. A Wiki engine is a programming base to create Wiki sites. In this PhD Thesis, our approach to the creation of a Wiki engine is to use an application on the Wiki pages and semantic meta-data. Our vision for the Wiki pages is that a Wiki page is an application. We do not work with static data on the Internet neither in a Wiki, now it is a Web application that contains the source code of the application interface, as well as the data for display. Application is able to retrieve and update data based on Linked Data principles on each page load, updating their sources as needed and showing only relevant information. This work is published [51], [52].

6.13. Discovery Hub

Participants: Nicolas Marie, Fabien Gandon, Damien Legrand.

Nicolas Marie is PhD student in collaboration with Alcatel-Lucent Bell Labs (Cifre). He is the project leader of Discovery Hub: a discovery engine on the top of DBpedia using real-time spreading activation.

We continue the CRE and CIFRE PhD Thesis (2011-2013) initially on Social objects, object-centered sociality, and object-centered social networks to propose mobile context-based notification application in a semantic and pervasive Web. The work evolved toward exploratory search, discovery and recommendation. Web growth, both in size and diversity, and users' growing expectations increase the need for innovative search approaches and technologies. Exploratory search systems are built specifically to help user in cognitive consuming search tasks like learning or investigation. Some of these systems are built on the top of linked data and use its semantic richness to provide cognitively-optimized search experiences. This work addresses the question of real-time linked data processing for exploratory search purposes. This live aspect offers advantages in terms of query expressivity and data dynamicity-handling.

To achieve this goal we propose a real-time semantic spreading activation algorithm (RTSA) which process linked data on-the-fly. This live aspect offers advantages in data dynamicity handling and query expressivity. Approximation strategies, algorithm behavior study and user evaluation related to RTSA algorithm are currently performed. The work includes a study of its behavior on DBpedia and a validation of its relevance through a user evaluation. Finally we also implemented a real deployment introducing the Discovery Hub prototype. It is an exploratory search engine offering advanced querying, browsing and explanation strategies for discovery purposes.

This algorithm is deployed in the Discovery Hub prototype [20], a discovery engine interfaced with services. Discovery Hub aims to help users to make numerous discoveries starting from its interests. The application works on DBpedia data including local version like fr.dbpedia.org (hosted by Inria/Wimmics). It also makes extensive use of the Corese/KGRAM Semantic Web Factory. Application front-end was designed and developed by Damien Legrand during an internship.

The application was presented during the Semanticpedia day, organized by official French language delegation [21].

Nicolas Marie is also active in the Web Science community [36].

6.14. Semantic Newsfeed Aggregation

Participant: Christophe Desclaux.

Christophe was this year in his last year of engineering school at Polytech UNS in the KIS speciality. During his end of course project he worked on the RSS feed aggregation using Named Entities Recognition. He presented his research project to the student contest *Boost Your Code* organized by Inria. The aim of the contest is to offer to a junior engineer a one year full time contract to work on an innovating OpenSource project. Christophe won the contest and is now part of the Wimmics team since november 2012. The ZONE project [22] provides a specialized tool for monitoring domain. ZONE semantically increases news for a better classification for the user. Christophe will work in collaboration with the team on documents clustering, natural language processing and RDF datastores.

---

[20] http://semreco.inria.fr/
[21] http://www.dglflf.culture.gouv.fr/Actualites/Programme_Semanticpedia.pdf
[22] http://zone-project.org

6.15. Linked Justifications

Participants: Rakebul Hasan, Fabien Gandon.

Semantic Web applications use inferential capabilities and distributed data in their reasoning. Users often find it difficult to understand how these applications produce their results. Hence, users often do not trust the results these applications produce. Explanation-aware Semantic Web applications provide explanations of their reasoning. Explanations enable users to better understand reasoning of these application. Users can use this additional information about reasoning to make their trust decisions.

The emergence of Linked Data offers opportunities for large-scale reasoning over heterogeneous and distributed data. Explaining reasoning over Linked Data requires explaining how these distributed data were produced. Publishing also the explanation related metadata as Linked Data enables such explanations. Justifications are metadata about how a given piece datum is obtained. We introduce the concept of Linked Justifications and provide guidelines to publish justifications as Linked Data in [67]. We published the *Ratio4TA*[23] (inte**r**linked justific**atio**ns **fort**riple **a**ssertions) vocabulary to describe justifications. *Ratio4TA* extends W3C PROV Ontology[24] to promote interoperability.

In [89], [66], we analyze the existing explanation-aware Semantic Web systems. The existing systems inherit explanation features from explanation-aware expert systems. These explanations are targeted to expert users, such as knowledge engineers, with detailed information about all the execution steps of reasoners of these applications. Unlike the expert systems, users of Semantic Web applications have diverse background - from expert knowledge engineers who are interested in every details of the reasoning, to regular users who do not have any background in reasoning, logic, or ontologies. These non-expert users might feel overwhelmed with all the execution details of reasoners. To address this issue, we propose summarized and relevant explanations to users. Users can specify their explanation goals - types of information they are interested in. We take into consideration the explanation goals when we present explanations and summarize explanations. We use centrality and similarity matrices to summarize and provide relevant explanations.

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

6.16.1. *Analyzing and Modeling the Sharing and Articulation of Representations*
Participants: Alain Giboin, Gessica Puri.

Comparing and Bridging Models of Representation Sharing Processes

Context: Follow-up to the RefCom joint research action of the GDR CNRS Psycho-Ergo, in collaboration with Pascal Salembier (UTT, France).

We continued our work on comparing and bridging models of representation sharing processes (see Edelweiss activity report 2011) in order to achieve mutual intelligibility between researchers working on such models. We extended at the European level the test and application of the grid we elaborated for collaboratively comparing and bridging the conceptualizations [86], [65].

Methods and tools for articulating developers, domain experts, users and usage analysts' representations

Reconciling informal and formal representations through the ECCO collaborative ontology editor

Context: Follow-up to the ANR project e-WOK HUB, in collaboration with Priscille Durville (a former Inria expert engineer, currently engineer at Xerox), Sandrine Grataloup and Olivier Morel (BRGM), Michel Perrin (ENSMP)

---

[23]http://ns.inria.fr/ratio4ta/
[24]http://www.w3.org/TR/prov-o/

In a new publication [81], we reported the method we designed and proposed to geologists and knowledge engineers to help them jointly define domain ontologies from textual documents. The method is instrumented through a collaborative ontology editor (ECCO) which integrates two tools for automatic analysis of natural language. ECCO allows articulating the informal representations formulated by geologists in natural language and the knowledge engineers' fully formalized ontology-based representations that can be processed automatically by a semantic search engine like Corese.

Frameworks and Toolsets

Context 1: Capitalizing methods and tools developed in the Wimmics (formerly Edelweiss) team

We started to integrate into a general framework the methods and tools developed in the Wimmics (formerly Edelweiss) team to help developers, domain experts, users and usage analysts reconcile their views in order to design user-adapted social semantic applications. A preliminary presentation of the framework was given in the interdisciplinary seminar mentioned in Section "Invited Talks".

Context 2: PhD thesis of Gessica Puri and Wimmics projects related to visualization and manipulation of links.

We are currently developing a "design thinking" toolset (including a framework) for helping developers think in terms of a user's point of view when they design and evaluate link visualization and manipulation applications such as graph visualization applications [90]. A first version of the toolset is being validated by developers of the team.

6.16.2. Scenario Modeling and User Modeling for system design and evaluation

Participants: Alain Giboin, Gessica Puri.

In the context of different design projects, we applied, adapted or renewed some of our scenario and user modeling methods.

Context 1: ISICIL project. In collaboration with Rubiela Silva (UNS & Ademe), Claire Prendleloup (Ademe), Mylène Leitzelman (Telecom ParisTech)

In order to evaluate the usefulness and usability of the ISICIL platform (see 6.2) for one of the communities of potential users of the platform (the French Environment and Energy Management Agency), we adapted in particular: (a) a technique for modeling collectives and their related scenarios to prepare testing situations as close as possible to real situations met by the potential users; and (b) an existing set of collective heuristics, i.e., heuristics formerly designed to evaluate groupware; see, e.g. [93]. We also proposed the technique of *Online collaborative sessions* supported by the ISICIL communication functionalities (chat and comments) as a writing-based version of the *think out loud* protocol. These techniques are partly described in [91], [92].

Context 2: PAL project, in collaboration with David Daney (Coprin), Rémi Barraquand (Prima), Nadine Mandran (Pôle d'Ingénierie Multidisciplinaire du Laboratoire d'Informatique de Grenoble)

In this work we applied and adapted techniques for articulating system-oriented scenarios (coming from robotics and computer vision specialists) and user-oriented scenarios. These scenarios were intended to guide the design of useful and usable services improving the autonomy and quality of life for elderly and fragile persons. An output of this work has been the organization of workshop where PAL researchers were invited to specify the scenarios motivating the development of their services.

Context 3: Projects *Socio-cultural encyclopedia of Senegalese communities* and *Global Warming Platform* – Papa Fary Diallo (Wimmics PhD student), Fatou Kamara and Moussa Lo (Université Gaston-Berger, Saint-Louis, Sénégal)

We started to apply techniques for modeling groups of users and their related scenarios to the design of social semantic applications aimed at (a) communities wanting to adapt these applications to their own culture, and (b) at communities including persons from different specialties (geographers, mathematicians) and laypersons which want to collaborate on global warming issues.

Context 4: Discovery Hub project, PhD thesis of Nicolas Marie, in collaboration with Florentin Rodio (Alcatel Lucent); also related to the PhD thesis of Gessica Puri.

In order to perform a user-oriented evaluation of the Discovery Hub recommender system (see 6.13 and [100]), we used techniques allowing to define more realistic scenarios of interaction with the recommender system from the user's point of view, and to analyze users' cognitive processes when they interact with the system (e.g., when they select recommended items or when they assess the quality of a recommendation with and without explanations).

6.16.3. Exploring eye-tracking techniques for evaluating information organization aspects in Web applications

Participants: Valériane Dusaucy, Alain Giboin.

Context: collaboration with Valériane Dusaucy (PhD student, University of Aix-Marseille & CIFRE Société Ausy) and Franck Ferront (ergonome, Société Ausy)

We designed an experiment to explore the potentiality of eye-tracking techniques for evaluating information organization aspects in Web applications from a user's point of view, and to compare it to other evaluation techniques such as heuristic evaluation. The experiment, which takes place in the Ubiquarium of the I3S Laboratory, is in progress.

ZENITH Project-Team

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, error in data entry, etc.

One of the areas, in which uncertainty management is important, is the integration of heterogeneous data sources, in the sense where usually there may be an uncertainty in the possible mappings between the attributes of the sources. Usually the human interaction is demanded to help the system in choosing the correct mappings. In [30], we propose a pay-as-you-go data integration solution that aims at preforming the data integration in a fully automated way. Our solution takes advantage of attribute correlations by using functional dependencies, and captures uncertainty in mediated schemas using a probabilistic data model. It allows integrating a given set of data sources, as well as incrementally integrating additional sources, without needing to restart the process from scratch. We implemented our solution, and compared it with a baseline approach. The performance evaluation results show significant performance gains of our solution in terms of recall and precision compared to the baseline approaches.

Another problem that arises in many applications such as data integration systems is that of Entity Resolution (ER). ER is the process of identifying tuples that represent the same real-world entity. It has been well studied in the literature for certain data, but it has not been deeply investigated for uncertain data. Existing proposals for the ER problem are not applicable to the above examples since they ignore probability values completely and return the most similar tuples as the solution. Furthermore, the semantics of the solution for the ERUD problem has not been clearly defined in the literature. In [31], we address the ERUD problem. We adopt the well-known possible worlds semantics for defining the semantics for the ERUD problem, and propose a PTIME algorithm for a large class of similarity functions, i.e. context-free. For the rest of similarity functions, i.e. context-sensitive, we use Monte-Carlo randomization for approximating the answer. We propose a parallel version of our Monte-Carlo algorithm using the MapReduce framework. To the best of our knowledge, this is the first study of the ERUD problem that adopts the possible world semantics and the first efficient algorithm for implementing it.

Another important problem in uncertain data management is the efficient processing of probabilistic queries. We have continued the development of our probabilistic database prototype, called ProbDB (Probabilistic Database) that deals with large-scale probabilistic data sharing. ProbDB divides each probabilistic 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.

6.1.2. Metadata Integration

Participants: Zohra Bellahsène, Emmanuel Castanier, Duy Hoa Ngo, Patrick Valduriez.

Our work on metadata integration encompassed ontology matching and open data source integration.

The major focus of our work in 2012 was to deal with large scale ontology matching and scalability. To improve the matching quality of YAM++, we designed a new IR-based measure to deal with terminological heterogeneity in real world ontologies. To deal with large ontology matching, we designed a method based on indexing concepts from their labels and comments. Our approach aims at reducing the search space when comparing the concepts of the input ontologies. For this purpose, we designed three filters: Description Filter, Context Filter and Label Filter. These methods make use of the Lucene search engine for indexing and searching the context of entities in the input ontologies. Another contribution lies on the Fast Semantic Filtering method, which refines the discovered mappings in the ontology matching task. The aim of the Semantic Filter is to detect and reject inconsistent mappings by exploring semantic information of entities in the input ontologies [45]. The originality of our method is to use a new structural indexing technique and a heuristic to generate relative disjointness axioms. At the 2012 competition of the Ontology Alignment Evaluation Initiative (http://oaei.ontologymatching.org), YAM++ was one of the best matchers, with very good results in all tracks. It obtained the first postision in the Large BioMed Track [55].

Integrating open data sources can yield high value information but raises major problems in terms of metadata extraction, data source integration and visualization of integrated data. In [34], [33], we describe WebSmatch, a flexible environment for Web data integration, based on a real, end-to-end data integration scenario over public data from Data Publica. WebSmatch supports the full process of importing, refining and integrating data sources and uses third party tools for high quality visualization. We use a typical scenario of public data integration which involves problems not solved by currents tools: poorly structured input data sources (XLS files) and rich visualization of integrated data.

6.1.3. *High-dimensional data management*

Participants: Mohamed Riadh Trad, Alexis Joly, Saloua Litayem.

High dimensional data hashing is essential for scaling up and distributing data analysis applications involving feature-rich objects, such as text documents, images or multi-modal entities (scientific observations, events, etc.). In this first research track, we first investigated the use of high dimensional hashing methods for efficiently approximating K-NN Graphs [47], particularly in distributed environments. We highlighted the importance of balancing issues on the performance of such approaches and show why the baseline approach using Locality Sensitive Hashing does not perform well. Our new KNN-join method is based on RMMH, a hash function family based on randomly trained classifiers that we introduced in 2011. We show that the resulting hash tables are much more balanced and that the number of resulting collisions can be greatly reduced without degrading quality. We further improve the load balancing of our distributed approach by designing a parallelized local join algorithm, implemented within the MapReduce framework. In other work [43], we address the problem of speeding-up the prediction phase of linear Support Vector Machines via Locality Sensitive Hashing. Whereas the mainstream work in the field is focused on training classifiers on huge amount of data, less efforts are spent on the counterpart scalability issue: how to apply big trained models efficiently on huge non annotated collections ? In this work, we propose building efficient hash-based classifiers that are applied in a first stage in order to approximate the exact results and alter the hypothesis space. Experiments performed with millions of one-against-one classifiers show that the proposed hash-based classifier can be more than two orders of magnitude faster than the exact classifier with minor losses in quality.

6.2. Data and Process Sharing

6.2.1. *Hybrid P2P/cloud Architecture*

Participants: Esther Pacitti, Patrick Valduriez.

Zenith adopts a hybrid P2P/cloud architecture. P2P naturally supports the collaborative nature of scientific applications, with autonomy and decentralized control. Peers can be the participants or organizations involved in collaboration and may share data and applications while keeping full control over some of their data (a major requirement for our application partners). But for very-large scale data analysis or very large workflow activities, cloud computing is appropriate as it can provide virtually infinite computing, storage and networking resources. Such hybrid architecture also enables the clean integration of the users' own computational resources with different clouds.

In [24], we define Zenith's architecture with P2P data services and cloud data services. We model an online scientific community as a set of peers and relationships between them. The peers have their own data sources. The relationships are between any two or more peers and indicate how the peers and their data sources are related, e.g. friendship, same semantic domain, similar schema. The P2P data services include basic services (metadata and uncertain data management): recommendation, data analysis and workflow management through the Shared-data Overlay Network (SON) middleware. The cloud P2P services include data mining, content-based information retrieval and workflow execution. These services can be accessed through web services, and each peer can use the services of multiple clouds.

6.2.2. Social-based P2P Data Sharing

Participants: Reza Akbarinia, Emmanuel Castanier, Esther Pacitti, Didier Parigot, Patrick Valduriez, Guillaume Verger.

As a validation of the ANR DataRing project, we have developed P2PShare, a P2P system for large-scale probabilistic data sharing in scientific communities. P2PShare leverages content-based and expert-based recommendation. It is designed to manage probabilistic and deterministic data in P2P environments. It provides a flexible environment for integration of heterogeneous sources, and takes into account the social based aspects to discover high quality results for queries by privileging the data of friends (or friends of friends), who are expert on the topics related to the query.

Using the Shared-Data Overlay Network (SON), we have implemented a prototype of P2PShare that integrates three major DataRing services: ProbDB, a probabilistic database management service for relational data; WebSmatch, an environment for Web data integration; and P2Prec, a social-based P2P recommendation service for large-scale content sharing.

In [50], , we describe the demo of P2PShare's main services, e.g., gossiping topics of interest among friends, key- word querying for contents, and probabilistic queries over datasets.

6.2.3. View Selection in Distributed Data Warehousing

Participants: Zohra Bellahsène, 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 [23]. At large scale, a data warehouse can be distributed. We have examined the problem of choosing a set of views and a set of data warehouse nodes at which these views should be materialized so that the full query workload is answered with the lowest cost. To address this problem, we extended our view selection method that we proposed for the centralized case. Thus, we modelled the distributed view selection problem as a Constraint Satisfaction Problem (CSP). Furthermore, we introduced the distributed AND-OR view graph, which can be seen as an extensive form of the AND-OR view graph to reflect the relation between views and communication network within the distributed scenario. The experiment results show that our approach provides better performance compared with the genetic algorithm in term of the solution quality (i.e., the quality of the obtained set of materialized views). We demonstrated experimentally that our approach provides better results in term of cost savings when the view selection is decided under space and maintenance cost constraints [44].

6.2.4. Scientific Workflow Management

Participants: Ayoub Ait Lahcen, Jonas Dias, Didier Parigot, Patrick Valduriez.

Scientific experiments based on computer simulations can be defined, executed and monitored using Scientific Workflow Management Systems (SWfMS). Several SWfMS are available, each with a different goal and a different engine. Due to the exploratory analysis, scientists need to run parameter sweep (PS) workflows, which are workflows that are invoked repeatedly using different input data. These workflows generate a large amount of tasks that are submitted to High Performance Computing (HPC) environments. Different execution models for a workflow may have significant differences in performance in HPC. However, selecting the best execution model for a given workflow is difficult due to the existence of many characteristics of the workflow that may affect the parallel execution.

In [36], we develop a study to show performance impacts of using different execution models in running PS workflows in HPC. Our study contributes by presenting a characterization of PS workflow patterns (the basis for many existing scientific workflows) and its behavior under different execution models in HPC. We evaluated four execution models to run workflows in parallel. Our study measures the performance behavior of small, large and complex workflows among the evaluated execution models. The results can be used as a guideline to select the best model for a given scientific workflow execution in HPC. Our evaluation may also serve as a basis for workflow designers to analyze the expected behavior of an HPC workflow engine based on the characteristics of PS workflows.

This work was done in the context of the the CNPq-Inria project DatLuge and FAPERJ-Inria P2Pcloud project .

In the context of SON, we also proposed a declarative workflow language based on service/activity rules. In [27], [46], we present a formal approach that combines component-based development with well-understood methods and techniques from the field of Attribute Grammars and Data-Flow Analysis in order to specify the behavior of P2P applications, and then construct an abstract representation (i.e., Data-Dependency Graph) to perform analyzes on it. This formal approach makes it possible to infer a dependency graph for SON applications that provides for automatic parallelization.

6.2.5. *Plants identification and classification from social image data*

Participants: Hervé Goëau, Alexis Joly, Saloua Litayem.

This work is done in collaboration with the botanists of the AMAP UMR team (CIRAD) and with Inria team IMEDIA. Inspired by citizen sciences, the main goal of this trans-disciplinary 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 did continue working intensively on plants identification and classification [54], [37], [38], [26]. We first developed a new interactive method [37] for the visual identification of plants from social image data. Contrary to previous content-based identification methods and systems that mainly relied on leaves, or in few other cases on flowers, it makes use of five different organs and plant's views including habit, flowers, fruits, leaves and bark. Thanks to an interactive query widget, the tagging process of the different organs and views is as simple as drag-and-drop operations and does not require any expertise in botany. All training pictures used by the system were continuously collected during one year through a crowdsourcing application and more than 17K images are now integrated. System-oriented and human-centered evaluations of the application show that the results are already satisfactory and therefore very promising in the long term to identify a richer flora.

Besides, we did continue working on leaf-based identification notably through the organization of and participation to ImageCLEF plant identification evaluation campaign 2012 [54].

Finally we did apply one of our former work related to multi-source shared-nearest neighbors clustering to an original experiment aimed at evaluating if we were able to automatically recover morphological classifications built by the botanists themselves [38]. The results are very promising, since all clusters discovered automatically could be easily matched to one node of a morphological tree built by botanists.

6.3. Scalable Data Analysis

6.3.1. *StreamCloud*

Participants: Vincenzo Gulisano, Patrick Valduriez.

Recent years have witnessed the growth of a new class of data-intensive applications that do not fit the DBMS query paradigm. Instead, the data arrive at high speeds taking the form of an unbounded sequence of values (data streams) and queries run continuously returning new results as new data arrive. Examples of data streams are sensor data (e.g. in environmental applications) or IP packets (e.g. in a network monitoring application). The unbounded nature of data streams makes it impossible to store the data entirely in bounded memory. Current research efforts have mainly focused on scaling in the number of queries and/or query operators having overlooked the scalability with respect to the stream volume.

Current Stream Processing Engines do not scale with the input load due to single-node bottlenecks. Additionally, they are based on static configurations that lead to either under or over-provisioning. In [21], [22], we present StreamCloud, a scalable and elastic stream processing engine for processing large data stream volumes. StreamCloud uses a novel parallelization technique that splits queries into subqueries that are allocated to independent sets of nodes in a way that minimizes the distribution overhead. Its elastic protocols exhibit low intrusiveness, enabling effective adjustment of resources to the incoming load. Elasticity is combined with dynamic load balancing to minimize the computational resources used. We present the system design, implementation and a thorough evaluation of the scalability and elasticity of the fully implemented system.

6.3.2. *Mining Uncertain Data Streams*

Participants: Reza Akbarinia, Florent Masseglia.

Dealing with uncertainty by using probabilistic approaches has gained increasing attention these past few years. One of the main requirements for uncertain data mining is the ability to discover Probabilistic Frequent Itemsets (PFI). However, PFI mining, particularly in uncertain data streams, is very challenging and needs the development of new techniques, since approaches designed for deterministic data are not applicable in this context. In [29], we propose an efficient solution for exact PFI mining over data streams with sliding windows. Our proposal includes efficient solutions for updating frequentness probability of itemsets and thus fast extraction of PFI, whenever transactions are added or removed from the sliding window. To the best of our knowledge, this is the first efficient solution for data stream PFI mining. We have conducted an extensive experimental evaluation of our approach over synthetic and real-world data sets; the results illustrate its very good performance.

6.3.3. *Detecting Rare Events in Massive Datasets*

Participant: Florent Masseglia.

In this work, we consider that rare events are very small clusters typically representing less than 0.01% of the entire dataset. Finding these abnormal events allows to identify the emergence of pos- sible anomalies in their very early stages. Such a scenario is generally difficult to handle as it lies at the frontier between outlier detection and clustering and is characterized by a clear challenge to avoid false nega- tives. To address this challenge, we take a backward approach and pro- pose RARE, a framework that identifies and isolates the abnormal/rare regions. The dense regions are identified using a radius-limited density- driven variant of k-means and adjacent regions are merged to form new regions. These newly formed regions are gradually augmented as long as a density-driven condition is respected. When no more dense regions are observed, the remaining data is clustered and presented for further analysis to human experts. The framework is tested on a medical appli- cation and compared against human analysis. The experiments show that rare events that were missed during human analysis because of the multivariate character of the data can be discovered by our approach.

This work is funded by the labex NUMEV and a patent application involving Inria, CNRS, UM2 and INSERM has been filled.

6.3.4. *Highly Informative Feature Set Mining*

Participant: Florent Masseglia.

For many textual collections, the number of features is often overly large. As these features can be very redundant, it is desirable to have a small, succinct, yet highly informative collection of features that describes the key characteristics of a dataset. Information theory is one such tool for us to obtain this feature collection. In [48], we mainly contribute to the improvement of efficiency for the process of selecting the most informative feature set over high-dimensional unlabeled data. We propose a heuristic theory for informative feature set selection from high dimensional data. Moreover, we design data structures that enable us to compute the entropies of the candidate feature sets efficiently. We also develop a simple pruning strategy that eliminates the hopeless candidates at each forward selection step. We test our method through experiments on real-world data sets, showing that our proposal is very efficient.

6.3.5. Clustering Users with Evolving Profiles in Usage Streams
Participant: Florent Masseglia.

Existing data stream models commonly assume that users' records or profiles in data streams will not be updated once they arrive. In many applications such as web usage, however, the users' records/profiles may evolve along time. This kind of streaming transactions are referred to as bi-streaming data (*i.e.* the data evolves temporally in two dimensions, the flowing of transactions as with the traditional data streams, and the evolving of users' profiles inside the streams, which makes bi-streaming data different from traditional data streams). The two-dimensional evolving of bi-streaming data brings difficulties on modeling and clustering for exploring the users' behaviors. In [49], we propose three models to summarize bi-streaming data, which are the batch model, the Evolving Objects (EO) model and the Dynamic Data Stream (DDS) model. Through creating, updating and deleting user profiles, the models summarize the behaviors of each user as an object. Based on these models, clustering algorithms are employed to identify the user groups. The proposed models are tested on a real-world data set showing that the DDS model can summarize the bi-streaming data efficiently and effectively, providing better basis for clustering user profiles than the other two models.

6.3.6. Scalable Mining of Small Visual Objects
Participants: Pierre Letessier, Julien Champ, Alexis Joly.

Automatically linking multimedia documents that contain one or several instances of the same visual object has many applications including: salient events detection, relevant patterns discovery in scientific data or simply web browsing through hyper-visual links. Whereas efficient methods now exist for searching rigid objects in large collections, discovering them from scratch is still challenging in terms of scalability, particularly when the targeted objects are rather small. In this work [40], we formally revisit the problem of mining or discovering such objects, and then generalized two kinds of existing methods for probing candidate object seeds: weighted adaptive sampling and hashing based methods. We then introduce a new hashing strategy, working first at the visual level, and then at the geometric level. Experiments conducted on millions of images show that our method outperforms state-of-the-art.

This method was integrated within a visual-based media event detection system in the scope of a French project called the transmedia observatory. It allows the automatic discovery of the most circulated images across the main news media (news websites, press agencies, TV news and newspapers). The main originality of the detection is to rely on the transmedia contextual information to denoise the raw visual detections and consequently focus on the most salient trans-media events. This work was presented at ACM Multimedia Grand Challenge 2012 [39]. The movie presented during this event is available at http://www.otmedia.fr/?p=217.