informatics *mathematics*

Ínría

RESEARCH CENTER
Sophia Antipolis - Méditerranée

FIELD

Activity Report 2014

Section New Results

AOSTE Project-Team

6. New Results

6.1. Languages, Models of Computation and Metamodeling using logical clock constraints

Participants: Julien Deantoni, Robert de Simone, Frédéric Mallet, Marie Agnès Peraldi Frati.

A revised and updated version of our previous work on UML MARTE Time Model was written in survey textbook form for a larger audience, and published in [38]. Same was done for the more applied specific findings of the ARTEMIS PRESTO European project [39]. Also, a research report finalizing the denotational semantics of the logical clock constraint languages was issued for reference [44].

6.2. Experiments with Architecture and Application modeling

Participants: Robert de Simone, Émilien Kofman, Jean-Vivien Millo, Amine Oueslati, Mohamed Bergach.

We submitted for publication our theoretical results on formal mapping of an application written as a process network dataflow graph onto an abstract architecture model involving a network-on-chip and manycore processor arrays [24].

In the context of the *FUI Clistine* collaborative project (which aims at building a cheap supercomputer by assembling low-cost, general-purpose and network processors interconnected by a time-predictable, on-board network), we considered the issue of classifying general application types, in the fashion inherited from UC. Berkeley's 13 "dwarfs" [46]. Meanwhile,the modeling of desired architecture was slightly postponed due to hesitations from the main industrial partner (that will build the prototype itself). This work was the topic of Amine Oueslati's first year PhD. The classification, and the use of distinct type properties for efficient and natural encoding, was applied on typical application programs provided by partners (Galerkin methods for electromagnetic simulation by the Nachos Inria team, ray-tracing algorithms by the Optis/Simplysim SME design company).

In the context of Mohammed Bergach's CIFRE PhD contract with Kontron Toulon, we conducted an advanced modeling exercice on how to best fit large DFT (Discrete Fourier Transform) modules onto a specific processor architecture (first Intel Sandybridge, then Haswell) that offers computing compromise costs (in performance vs power) between regular CPUs and GPU hardware accelerators. There were two issues: first, how to best dimension the size of the largest FFT block that may be performed locally on a corresponding GPU compute block; second, how to distribute the many such optimal size FFT block needed in a typical radar application, using the GPU and CPU features at the best of their capacity, with account of the slow data transfer latencies across memory banks (to and from the GPU registers).
As a side-effect, people from Kontron are now using and distributing to their customers the FFT GPU libraries with ad-hoc FFT variants matching the GPU block memory sizes. The development, rather lenghty in the case of Sandybridge, was quickly adjusted and ported for Haswell. A new workshop paper is under submission.

6.3. Multiview modeling with performance and power aspects

Participants: Julien Deantoni, Ameni Khecharem, Robert de Simone, Emilien Kofman, Carlos Gomez.

In the context of the ANR HOPE project and The CIM PACA Design Platform, we continued our work on joint modeling and co-simulation of abstract architecture and application (use case scenario) models, together with non-functional aspect views such as performance, power and temperature. The goal of the HOPE project is to consider *hierarchical* {Power/Performance/Temperature} Management Units (MU), and our target is to connect our IDM modeling with dedicated tools such as Synopsys Platform Architect or Docea Power AcePlorer. The IP_Xact interface format for IP blocks is also aimed for compositional assembly representation, including non-functional properties and timing semantics constraints for co-simulation. This work is mostly continued from the former PhD thesis of Carlos Gomez to a new framework by Ameni Khecharem, as part of her PhD. Practical co-simulation trends are also investigated. Currents results were reported in [29]

6.4. Heterogeneous Languages Coordination with Concurrency and Time

Participants: Julien Deantoni, Matias Vara Larsen, Robert de Simone, Frédéric Mallet.

In the context of the ANR GEMOC project and in closely related to the mutiview approach of the team, we focused on how to deal with analysis and simulation of heterogeneous languages. Supporting coordinated use of heterogeneous domain specific languages leads to what we called the globalization of modeling language [22]. Concretely, we proposed to define a language behavioral interface to exhibit the concurrency and time aspects of the semantics of a language. The concurrency and time aspects are described by a formal extension of CCSL, named MoCCML (Model of Concurrency and Communication Modeling Language [45], [28]). Any models that conform such language exhibit a symbolic representation of all its acceptable schedules. Based on this, we shown that it is possible to coordinate heterogeneous models .To avoid redundant model coordination activities, we reified the know-how about model coordination in BCOoL (Behavioral Coordination Operator Language[34]), a language dedicated to language coordination. This work is mainly realized by Matias Vara Larsen, as part of his PhD. In this context, we organize the community around such subject for the second year in an international workshop [43] with an increasing number of participants.

6.5. Performance study of Massively Parallel Processor Array (MPPA) SoC architecture

Participants: Sid Touati, Franco Pestarini.

From a previous collaboration programme, we (Aoste Sophia) possess a MMPA manycore chip, designed and produced by the company Kalray, in Grenoble. The chip integrates 256 cores, composed of 16 clusters (themselves each with 16 cores), and a powerfull network-on-chip interconnect mesh structure. This architecture is oriented towards high performance embedded application, with real time constraints. The cores and NoC were designed to deliver predictable performance.

Our current project, during Franco Pestarini Inria International Intern period, was to test the performance of the NoC, trying to obtain better knowledge of its behavior. We put up a set of microbenchmarks to exercice the network under different specific scenarios (low overhead network traffic, high traffic), and analyzed the experimental results.

We produced a detailed deliverable report explaining under which conditions the NoC could deliver stable and predictable performances. We identified potential configurations where the network becomes unstable (leading to variable and impredictable performances and bandwith).

Meanwhile, the textbook on low-level code optimization, written between Sid Touati and Kalray CTO, appeared in published form [42]. Its content reports on some of the techniques used inside the MPPA compilation environment, and beyond.

6.6. Parametric and Non Parametric Statistics for Code Performance Analysis

Participant: Sid Touati.

This activity is conducted by Sid Touati in collaboration with Julien Worms, an associate professor in Mathematics at the university of Versailles Saint Quentin. It was started under the consideration that the performances of programs are hardly ever represented by a gaussian distribution. So, our purpose here is to study parametric statistics for analyzing the performances of programs. We are interested in modelling program performances by gaussian mixtures (using mixmod method). After a statistical modeling, we deduce multiple performance tests and performance criteria to decide with a high degree of confidence about the "best" program run version. This is still work-in-progress: we are implementing a free software for analysis based on our approach, and we are writing a rather complete research report prior to further publications in conferences and journals.

6.7. Uniprocessor Real-Time Scheduling

Participants: Falou Ndoye, Yves Sorel, Walid Talaboulma.

6.7.1. *Real-Time Scheduling with Exact Preemption Cost*

Previous years, we worked on schedulability analyses of dependent tasks, executed on a uniprocessor, which take into account the exact preemption cost and more generally the cost of the real-time operating system. Indeed, this cost is composed of the cost of the preemptions and the cost of the scheduler. Our approach is based on an offline real-time schedulability analysis, proved sustainable, that produces a scheduling table. This latter contains the next instants (activation and completion of tasks) when the scheduler will be called, beeing aware of the instants where tasks are preempted and then resumed. This approach allows the schedulability analysis to account preemption costs involved by other preemptions. The scheduling table contains also the address of the next task to execute preventing the scheduler to choose it in the ready tasks list, unlike with classical on-line scheduler. The theoretical results in the uniprocessor case, are given in the Falou Ndoye's PhD thesis [19] defended in April this year. This approach has been implemented through an offline scheduler that is triggered by a timer when this latter is equal to zero and loaded with the next instant contained in the scheduling table, according to a time trigger approach.

We carried out two kinds of implementations. Actually, the first one is a simulation since our time trigger offline scheduler is modelled as a high priority task running upon an existing operating system. We experimented this approach with Vanilla Linux (not modified) and Linux/Xenomai, real-time versions of the Linux operating system, with low latency caracteristics. Of course, these implementations were only able to show that the theoterical results were correct, but did not provide good real-time performances nor a robust way to measure time without influencing the usefull code. Therefore, we implemented our scheduler on a bare metal ARM968E-S processor based on an ARM9 architecture since it is widespread in the industry world, and we experimented this processor few years ago to determine the cost of classical online schedulers.

For this purpose we used a MCB2929 developpement board, from Keil, containing the LPC2929 SoC including the ARM968E-S, an accurate timer, and various peripherals. The scheduling table is generated offline for a set of tasks, and stored in the memory as an array of couples, each composed of the task to execute and the duration elapsing until the next scheduler call. This duration is used to set the timer counter. When it hits zero it triggers a high priority interruption that is serviced by calling again the scheduler to choose the next couple (task, duration), and so on up to the end of the scheduling table. This will repeat infinitely from the beginning of the scheduling table.

We tested different set of tasks with multiple preemption scenarios, that can yield to deadlines misses. We measured for a **12Mhz** CPU and timer clock frequencies a value of **28** μ**s** for the scheduler cost, and of **1** μ**s** for the preemption cost.

6.8. Multiprocessor Real-Time Scheduling

Participants: Aderraouf Benyahia, Laurent George, Falou Ndoye, Dumitru Potop-Butucaru, Yves Sorel, Meriem Zidouni.

6.8.1. *Multiprocessor Partitioned Scheduling with Exact Preemption Cost*

Since we chose a multiprocessor partitioned scheduling approach, we can take advantage of the results we obtained in the case of uniprocessor real-time Scheduling accounting for the cost of the real-time operating system, i.e. the cost of preemptions and of the scheduler. From the point of view of the off-line real-time schedulability analysis we only have to consider in addition to activation and completion instants, reception of data instants. This latter instant is determinated by supposing that the cost of every data transfer is known for every possible communication medium. Indeed, when two dependent tasks are allocated to two different processors, the consuming task will have to wait for the data sent by the producing task. The theoretical results in the multiprocessor case, are given in the Falou Ndoye's PhD thesis [19] defended in April this year. We chose the message passing protocol for interprocessor communications achieved through a switched ethernet network. In order to determine precisely the cost of data transfers, we started to investigate the possible approaches to synchronize the send and receive tasks located in two different processors and to schedule them with the other tasks allocated to the same processor. This synchronization protocol will be taken into account to determine the interprocessor communication costs. Concerning these communication costs, we consider FIFO and FIFO* schedulings in the switches, the later is a FIFO scheduling based on the release time of frames at their source node. We have first corrected the trajectory approach (recently shown to be optimistic for corner cases) with FIFO scheduling to compute worst case end-to-end communication costs. Then, we have extended the trajectory approach to FIFO* scheduling. We want to implement our off-line scheduler on every processor of a multiprocessor architecture composed of at least three processors, communicating through an ethernet switch.

Concerning the delay of communications, we consider FIFO and FIFO* schedulings in the switches, the later is a FIFO scheduling based on the release time of frames at their source node. We have first corrected the trajectory approach (recently shown to be optimistic for corner cases) with FIFO scheduling to computed worst case end-to-end communication delays. Then, we have extended the trajectory approach to FIFO* scheduling.

6.8.2. *Mutiprocessor Parallel Directed Acyclic Graph (DAG) scheduling*

We are interested in studying the hard real-time scheduling problem of parallel Directed Acyclic Graph (DAG) tasks on multiprocessor systems. In this model, a task is defined as a set of dependent subtasks that execute under precedence constraints. The execution order of these subtasks is dynamic, i.e., a subtask can execute either sequentially or in parallel with its siblings based on the decisions of the real-time scheduler. To this end, we analyze two DAG scheduling approaches to determine the execution order of subtasks: the Model Transformation and the Direct Scheduling approaches. We consider global preemptive multiprocessor scheduling algorithms to be used with the scheduling approaches, such as Earliest Deadline First (EDF) and Deadline Monotonic (DM).

6.8.3. *Gateway with Modeling Languages for Certified Code Generation*

This work was carried out in the P FUI project 8.2.2 We continued the work on the gateway between the P formalism and SynDEx, started the last two years. We have integrated in the gateway the IF and FOR blocks of Simulink that were missing in the functional specification, except for particular cases where the IF block is nested in the FOR block, or the opposite. The integration of the MERGE and MUX blocks are still to be done. We extended the P formalism with architectural elements that SynDEx needs to perform schedulability analyses on functional specifications. These architectural elements are hardware resources (processor, bus, shared memory, router) and timing characteristics (deadline, period, WCET, WCTT). We developed a new part in the gateway which transforms an architectural model described with the P formalism in the input format of SynDEx. We developed also a third part in the gateway which feedbacks the schedulability analysis results obtained with SynDEx (the scheduling table) and stores them into models described with the P formalism. Finally, we have collaborated with the industrial partners to test our gateway on their use cases.

6.8.4. *SynDEx updates*

The first tests on the alpha version of SynDEx V8, released last year, shown some bugs that we fixed. This first release did not include a code generator. Thus, we worked to interface the distributed real-time embedded code generator of SynDEx V7 with SynDEx V8.

6.9. Probabilistic Real-Time Systems

Participants: Liliana Cucu-Grosjean, Robert Davis, Adriana Gogonel, Codé Lo, Dorin Maxim, Cristian Maxim.

The advent of complex hardware, in response to the increasing demand for computing power in next generation systems, exacerbates some of the limitations of static timing analysis for the estimation of the worst-case execution time (WCET) estimation. In particular, the effort of acquiring (1) detail information on the hardware to develop an accurate model of its execution latency as well as (2) knowledge of the timing behaviour of the program in the presence of varying hardware conditions, such as those dependent on the history of previously executed instructions. These problems are also known as the timing analysis walls. The probabilistic timing analysis, a novel approach to the analysis of the timing behaviour of next-generation real-time embedded systems, provides answers to timing analysis walls. In [23] we have described the vision of FP7 IP PROXIMA, project that is interested in the introduction of randomization of the architectures at cache level. For this type of architecture static probabilistic timing analysis is possible [20] by providing bounds on the probabilistic execution time of a task. An industrial case study from avionics is detailed in [32]. Such distribution is then used as input for probabilistic scheduling as described in [37], [36].

This year we have also provided a complete state of the art of the probabilistic real-time systems in [17].

6.10. Off-line (static) mapping and WCET analysis of real-time applications onto NoC-based many-cores

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

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

In past years we have identified and compared the hardware mechanisms supporting precise timing analysis and efficient resource allocation in existing NoCs. We have determined that the NoC should ideally provide the means of enforcing a global communications schedule that is computed off-line and which is synchronized with the scheduling of computations on CPU cores. Furthermore, if in addition the computation and memory resources of the MPSoC have support for real-time predictability, then parallel applications can be developed that allow very precise WCET analysis of parallel code. WCET analysis of parallel code is joint work with Isabelle Puaut of Inria, EPI ALF.

This year we have completed our mapping (allocation and scheduling) and code generation technique and tool for NoC-based MPSoCs. NoCs pose significant challenges to both on-line (dynamic) and off-line (static) real-time scheduling approaches [25]. They have large numbers of potential contention points, have limited internal buffering capabilities, and network control operates at the scale of small data packets. Therefore, efficient resource allocation requires scalable algorithms working on hardware models with a level of detail that is unprecedented in real-time scheduling.

We considered a static (off-line) scheduling approach, and we targeted massively parallel processor arrays (MPPAs), which are MPSoCs with large numbers (hundreds) of processing cores. We proposed a novel allocation and scheduling method capable of synthesizing such global computation and communication schedules covering all the execution, communication, and memory resources in an MPPA. To allow an efficient use of the hardware resources, our method takes into account the specificities of MPPA hardware and implements advanced scheduling techniques such as pre-computed preemption of data transmissions[26] and pipelined scheduling [21].

Our method has been implemented within the Lopht tool presented in section 5.4 , and first results are presented in [26], [25], and in extenso in the PhD thesis of manel Djemal [18]. One of the objectives of the starting CAPACITES project is the evaluation of the possibility of porting Lopht and the WCET analysis technique for parallel code onto the Kalray MPPA platform.

6.11. Real-time scheduling and code generation for time-triggered platforms

Participants: Dumitru Potop Butucaru, Thomas Carle, Raul Gorcitz, Yves Sorel.

We have continued this year the work on real-time scheduling and code generation for time-triggered platforms. Much of this work was carried out as part of a bilateral collaboration with Airbus DS and the CNES, which fund the post-doctorate of Raul Gorcitz, and in our collaboration with the IRT SystemX, project FSF.

The objective is to facilitate the development of complex time-triggered systems by automating the allocation, scheduling, and code generation steps. We show that full automation is possible while taking into account all the specification elements required by a complex, real-life embedded control system. The main originality of our work is that it takes into account at the same time multiple complexity elements: functional specifications with conditional execution and multiple modes and various types of non-functional properties: real-time (release dates, deadlines, major time frame, end-to-end flows), ARINC 653 partitioning (which we can fully or partially synthesize), task preemptability, allocation. Our algorithms allow the automatic allocation and scheduling onto multi-processor (distributed) systems with a global time base, taking into account communication costs.

While the past years were mainly dedicated to the development of this scheduling and code generation technique, this year the technique and the associated tool have matured enough to allow the publication of the first results concerning the optimized scheduling algorithms [21] and its application on large case studies. Ongoing work by the post-doc Raul Gorcitz, funded by Airbus DS and the CNES aims at evaluating the applicability of our methods on embedded platforms that are being considered for the future european space launchers. The Lopht tool is also used in the IRT SystemX, project FSF as part of the proposed design flow. All extensions have been implemented in the Lopht tool. All this work has been presented *in extenso* in the PhD thesis of Thomas Carle [16].

GALAAD2 Team

6. New Results

6.1. Algebraic representations for geometric modeling

6.1.1. *A comparison of different notions of ranks of symmetric tensors*
Participants: Alessandra Bernardi, Jérome Brachat, Bernard Mourrain.

In [2], we introduce various notions of rank for a symmetric tensor, namely: rank, border rank, catalecticant rank, generalized rank, scheme length, border scheme length, extension rank and smoothable rank. We analyze the stratification induced by these ranks. The mutual relations between these stratifications, allow us to describe the hierarchy among all the ranks. We show that strict inequalities are possible between rank, border rank, extension rank and catalecticant rank. Moreover we show that scheme length, generalized rank and extension rank coincide.

6.1.2. *Dimensions and bases of hierarchical tensor-product splines*
Participant: Bernard Mourrain.

In [1], we prove that the dimension of trivariate tensor-product spline space of tri-degree (m,m,m) with maximal order of smoothness over a three-dimensional domain coincides with the number of tensor-product B-spline basis functions acting effectively on the domain considered. A domain is required to belong to a certain class. This enables us to show that, for a certain assumption about the configuration of a hierarchical mesh, hierarchical B-splines span the spline space. This paper presents an extension to three-dimensional hierarchical meshes of results proposed recently by Giannelli and Jüttler for two-dimensional hierarchical meshes.

Joint work with Dmitry Berdinsky, Taiwan Kim, Oh Min-Jae, Sutipong Kiatpanichgij (Department of Naval Architecture and Ocean Engineering, Seoul, South Korea), Cesare Bracco (Dipartimento di Matematica "Giuseppe Peano", Torino, Italy), Durkbin Cho (Department of Mathematics, Dongguk, South Korea).

6.1.3. *Bounds on the dimension of trivariate spline spaces: A homological approach*
Participant: Bernard Mourrain.

In [8], we consider the vector space of globally differentiable piecewise polynomial functions defined on a three-dimensional polyhedral domain partitioned into tetrahedra. We prove new lower and upper bounds on the dimension of this space by applying homological techniques. We give an insight of different ways of approaching this problem by exploring its connections with the Hilbert series of ideals generated by powers of linear forms, fat points, the so-called Fröberg–Iarrobino conjecture, and the weak Lefschetz property.

Joint work with Nelly Villamizar (RICAM - Johann Radon Institute for Computational and Applied Mathematics, Linz, Austria)

6.1.4. *High-quality construction of analysis-suitable trivariate NURBS solids by reparameterization methods*
Participants: André Galligo, Bernard Mourrain.

High-quality volumetric parameterization of computational domain plays an important role in three-dimensional isogeometric analysis. Reparameterization techniques can improve the distribution of isoparametric curves/surfaces without changing the geometry. In [10], using the reparameterization method, we investigate the high-quality construction of analysis-suitable NURBS volumetric parameterization. Firstly, we introduce the concept of volumetric reparameterization, and propose an optimal Möbius transformation to improve the quality of the isoparametric structure based on a new uniformity metric. Secondly, from given boundary NURBS surfaces, we present a two-stage scheme to construct the analysis-suitable volumetric parameterization: in the first step, uniformity-improved reparameterization is performed on the boundary surfaces to achieve high-quality isoparametric structure without changing the shape; in the second step, from a new variational harmonic metric and the reparameterized boundary surfaces, we construct the optimal inner control points and weights to achieve an analysis-suitable NURBS solid. Several examples with complicated geometry are presented to illustrate the effectiveness of proposed methods.

Joint work with Gang Xu (College of computer - Hangzhou Dianzi University, China), Timon Rabczuk (Bauhaus-Universität Weimar, Germany).

6.1.5. *Spline Spaces over Quadrangle Meshes with Complex Topologies*
Participants: André Galligo, Bernard Mourrain, Meng Wu.

Motivated by Magneto Hydrodynamic (MHD) simulation with isoparametric elements method, we pursue our work on new types of spline functions defined over a quadrangular mesh, that can follow isobaric curves with node singularities. The practicability of these splines is analyzed for different geometries related to MHD simulation.

This work is done in collaboration with Boniface Nkonga (Inria, EPI CASTOR and University of Nice).

6.1.6. *Parametric modeling for ship hull deformation*
Participant: Elisa Berrini.

The objective of the work is to develop a parametric modeler tool, allowing consistent ship hull deformations with respect to classic naval architecture design constraints. This work will be applied in automatic shape optimization process. Two scientific problematics are addressed : 1) The parametrization of the hull: the numerical representation of the shape from a defined set of parameters; 2) The deformations of curves and surfaces: getting a new shape by modifying chosen parameters from the parameterization set. The consistency with naval architecture constraints is essential.

To produce realistic models, we want to use methods similar to naval architects' ones. The approach under development is based on the extraction and deformation of skeletons curves.

6.2. Algebraic algorithms for geometric computing

6.2.1. *Resultant of an equivariant polynomial system with respect to the symmetric group*
Participants: Laurent Busé, Anna Karasoulou.

Given a system of n homogeneous polynomials in n variables which is equivariant with respect to the canonical actions of the symmetric group of n symbols on the variables and on the polynomials, it is proved that its resultant can be decomposed into a product of several smaller resultants that are given in terms of some divided differences. As an application, we obtain a decomposition formula for the discriminant of a multivariate homogeneous symmetric polynomial.

This work is submitted for publication [14].

6.2.2. *Delaunay Mesh Generation of NURBS Surfaces*
Participant: Laurent Busé.

We introduce a method for isotropic triangle meshing of NURBS surfaces. Based on Delaunay filtering and refinement, our approach departs from previous work by meshing in embedding space instead of parametric space. The meshing engine relies upon a novel line/surface intersection test, based on the matrix-based implicit representation of NURBS surfaces and numerical methods in linear algebra such as singular value and eigenvalue decompositions. A careful treatment of degenerate cases makes our approach robust to intersection points with multiple pre-images. In addition to ensure both approximation accuracy and mesh quality, our approach is seamless as it does not depend on the initial decomposition into NURBS patches, and is oblivious to the parameterization of the patches. Removing such dependencies provides us with a means to reliably mesh across patches with greater control over mesh sizing and shape of the elements.

This work was done in collaboration with Jingjing Shen and Neil Dodgson from Cambridge University and Pierre Alliez from TITANE.

6.2.3. *Toric Border Basis*
Participant: Bernard Mourrain.

In [11], we extend the theory and the algorithms of Border bases to systems of Laurent polynomial equations, defining "toric" roots. Instead of introducing new variables and new relations to saturate by the variable inverses, we propose a more efficient approach which works directly with the variables and their inverse. We show that the commutation relations and the inversion relations characterize toric border bases. We explicitly describe the first syzygy module associated to a toric border basis in terms of these relations. Finally, a new border basis algorithm for Laurent polynomials is described and a proof of its termination is given for zero-dimensional toric ideals.

Joint work with Philippe Trébuchet (LIP6 - UPMC).

6.2.4. *Border Basis relaxation for polynomial optimization*
Participants: Marta Abril-Bucero, Bernard Mourrain.

A relaxation method based on border basis reduction which improves the efficiency of Lasserre's approach is proposed to compute the optimum of a polynomial function on a basic closed semi-algebraic set. A new stopping criterion is given to detect when the relaxation sequence reaches the minimum, using a sparse flat extension criterion. We also provide a new algorithm to reconstruct a finite sum of weighted Dirac measures from a truncated sequence of moments, which can be applied to other sparse reconstruction problems. As an application, we obtain a new algorithm to compute zero-dimensional minimizer ideals and the minimizer points or zero-dimensional G-radical ideals. Experimentations show the impact of this new method on significant benchmarks. See [12].

6.2.5. *Flat extensions in ∗-algebra*
Participant: Bernard Mourrain.

The objective of this work is to develop a flat extension characterization on moment matrices in the non-commutative case. We give a flat extension theorem for positive linear functionals on $_*$-algebras. The theorem is applied to truncated moment problems on cylinder sets, on matrices of polynomials and on enveloping algebras of Lie algebras. See [17].

Joint work with Konrad Schmüdgen, University of Leipzig, Germany.

6.3. Symbolic-Numeric Analysis

6.3.1. *Cubatures, and related problems, with symmetry*
Participants: Mathieu Collowald, Evelyne Hubert.

We address the computation of cubature formulae as a moment problem. Symmetry by finite groups arise naturally for cubatures. We developed the algebraic results to use the symmetry in order to reduce the number of parameters and the size of the matrices involved in the flat extension.

6.3.2. *Quantitative Equidistribution for the Solutions of Systems of Sparse Polynomial Equations*

Participant: André Galligo.

For a system of Laurent polynomials $f_1, ..., f_n \in \mathbb{C}[x_1^{\pm 1}, ..., x_n^{\pm 1}]$ whose coefficients are not too big with respect to its directional resultants, we show in [6] that the solutions in the algebraic torus $(\mathbb{C}^*)^n$ of the system of equations $f_1 = \cdots = f_n = 0$, are approximately equidistributed near the unit polycircle. This generalizes to the multivariate case a classical result due to Erdös and Turán on the distribution of the arguments of the roots of a univariate polynomial. We apply this result to bound the number of real roots of a system of Laurent polynomials, and to study the asymptotic distribution of the roots of systems of Laurent polynomials over $\mathbb{Z}$ and of random systems of Laurent polynomials over $\mathbb{C}$.

Joint work with Carlos D'Andrea (DM-UBA - Departamento de Matemática, Spain), Martin Sombra (ICREA & Universitat de Barcelona, Spain).

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

6. New Results

6.1. Highlights of the Year

[10] was elected among the notable articles of 2013 by ACM and Computing Reviews (see http://computingreviews.com/recommend/bestof/notableitems_2013.cfm).

6.2. Mesh Generation and Geometry Processing

6.2.1. *A Surface Reconstruction Method for In-Detail Underwater 3D Optical Mapping*

Participant: Mariette Yvinec.

In collaboration with Pierre Alliez (EPI Titane), Ricard Campos (University of Girona), Raphael Garcia (University of Girona)

Underwater range scanning techniques are starting to gain interest in underwater exploration, providing new tools to represent the seafloor. These scans (often) acquired by underwater robots usually result in an unstructured point cloud, but given the common downward-looking or forward-looking configuration of these sensors with respect to the scene, the problem of recovering a piecewise linear approximation representing the scene is normally solved by approximating these 3D points using a heightmap (2.5D). Nevertheless, this representation is not able to correctly represent complex structures, especially those presenting arbitrary concavities normally exhibited in underwater objects. We present a method devoted to full 3D surface reconstruction that does not assume any specific sensor configuration. The method presented is robust to common defects in raw scanned data such as outliers and noise often present in extreme environments such as underwater, both for sonar and optical surveys. Moreover, the proposed method does not need a manual preprocessing step. It is also generic as it does not need any information other than the points themselves to work. This property leads to its wide application to any kind of range scanning technologies and we demonstrate its versatility by using it on synthetic data, controlled laser-scans, and multibeam sonar surveys. Finally, and given the unbeatable level of detail that optical methods can provide, we analyze the application of this method on optical datasets related to biology, geology and archeology. [23]

6.2.2. *A Transfer Principle and Applications to Eigenvalue Estimates for Graphs*

Participant: David Cohen-Steiner.

In collaboration with Omid Amini (ENS),

In this paper, we prove a variant of the Burger-Brooks transfer principle which, combined with recent eigenvalue bounds for surfaces, allows to obtain upper bounds on the eigenvalues of graphs as a function of their genus. More precisely, we show the existence of a universal constants C such that the k-th eigenvalue λ_k of the normalized Laplacian of a graph G of (geometric) genus g on n vertices satisfies $\lambda_k \leq C d_{max}(g+k)/n$ where d_max denotes the maximum valence of vertices of the graph. This result is tight up to a change in the value of the constant C. We also use our transfer theorem to relate eigenvalues of the Laplacian on a metric graph to the eigenvalues of its simple graph models, and discuss an application to the mesh partitioning problem. [44]

6.3. Topological and Geometric Inference

6.3.1. *Only distances are required to reconstruct submanifolds*

Participants: Jean-Daniel Boissonnat, Steve Oudot.

In collaboration with Ramsay Dyer (Johann Bernouilli Institute, University of Groningen, Pays Bas) and Arijit Ghosh (Max-Planck-Institut für Informatik, Saarbrücken, Germany).

In [45], we give the first algorithm that outputs a faithful reconstruction of a submanifold of Euclidean space without maintaining or even constructing complicated data structures such as Voronoi diagrams or Delaunay complexes. Our algorithm uses the witness complex and relies on the stability of *power protection*, a notion introduced in this paper. The complexity of the algorithm depends exponentially on the intrinsic dimension of the manifold, rather than the dimension of ambient space, and linearly on the dimension of the ambient space. Another interesting feature of this work is that no explicit coordinates of the points in the point sample is needed. The algorithm only needs the *distance matrix* as input, i.e., only distance between points in the point sample as input.

6.3.2. *Computing Persistent Homology with Various Coefficient Fields in a Single Pass*
Participants: Jean-Daniel Boissonnat, Clément Maria.

In [32], we introduce an algorithm to compute the persistent homology of a filtered complex with various coefficient fields in a single matrix reduction. The algorithm is output-sensitive in the total number of distinct persistent homological features in the diagrams for the different coefficient fields. This computation allows us to infer the prime divisors of the torsion coefficients of the integral homology groups of the topological space at any scale, hence furnishing a more informative description of topology than persistence in a single coefficient field. We provide theoretical complexity analysis as well as detailed experimental results.

6.3.3. *Recognizing shrinkable complexes is NP-complete*
Participants: Olivier Devillers, Marc Glisse.

In collaboration with Dominique Attali (Gipsa-lab, Grenoble), Sylvain Lazard (Inria Nancy - Grand Est)

We say that a simplicial complex is shrinkable if there exists a sequence of admissible edge contractions that reduces the complex to a single vertex. We prove [31] that it is NP-complete to decide whether a (three-dimensional) simplicial complex is shrinkable. Along the way, we describe examples of contractible complexes which are not shrinkable.

6.3.4. *Zigzag Zoology: Rips Zigzags for Homology Inference*
Participant: Steve Oudot.

In collaboration with Donald Sheehy (University of Connecticut)

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 [28] 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.

6.3.5. *Zigzag Persistence via Reflections and Transpositions*
Participants: Clément Maria, Steve Oudot.

We introduce [40] a simple algorithm for computing zigzag persistence, designed in the same spirit as the standard persistence algorithm. Our algorithm reduces a single matrix, maintains an explicit set of chains encoding the persistent homology of the current zigzag, and updates it under simplex insertions and removals. The total worst-case running time matches the usual cubic bound. A noticeable difference with the standard persistence algorithm is that we do not insert or remove new simplices "at the end" of the zigzag, but rather "in the middle". To do so, we use arrow reflections and transpositions, in the same spirit as reflection functors in quiver theory. Our analysis introduces a new kind of reflection called the "weak-diamond", for which we are able to predict the changes in the interval decomposition and associated compatible bases. Arrow transpositions have been studied previously in the context of standard persistent homology, and we extend the study to the context of zigzag persistence. For both types of transformations, we provide simple procedures to update the interval decomposition and associated compatible homology basis.

6.3.6. *Topological analysis of scalar fields with outliers*
Participants: Mickaël Buchet, Frédéric Chazal, Steve Oudot.

In collaboration with Tamal K. Dey (University of Ohio) Fengtao Fan (University of Ohio) Yusu Wang (University of Ohio)

We extend [57] the notion of the distance to a measure from Euclidean space to probability measures on general metric spaces as a way to do topological data analysis in a way that is robust to noise and outliers. We then give an efficient way to approximate the sub-level sets of this function by a union of metric balls and extend previous results on sparse Rips filtrations to this setting. This robust and efficient approach to topological data analysis is illustrated with several examples from an implementation.

6.3.7. *Efficient and Robust Persistent Homology for Measures.*
Participants: Mickaël Buchet, Frédéric Chazal, Steve Oudot.

In collaboration with Donald Sheehy (University of Connecticut)

In [34], we extend the notion of the distance to a measure from Euclidean space to probability measures on general metric spaces as a way to do topological data analysis in a way that is robust to noise and outliers. We then give an efficient way to approximate the sub-level sets of this function by a union of metric balls and extend previous results on sparse Rips filtrations to this setting. This robust and efficient approach to topological data analysis is illustrated with several examples from an implementation.

6.3.8. *Persistence-based Structural Recognition*
Participants: Frédéric Chazal, Maksims Ovsjanikovs.

In collaboration with Chunyuan Li (former intern in Saclay in 2013)

In [39] we present a framework for object recognition using topological persistence. In particular, we show that the so-called persistence diagrams built from functions defined on the objects can serve as compact and informative descriptors for images and shapes. Complementary to the bag-of-features representation, which captures the distribution of values of a given function, persistence diagrams can be used to characterize its structural properties, reflecting spatial information in an invariant way. In practice, the choice of function is simple: each dimension of the feature vector can be viewed as a function. The proposed method is general: it can work on various multimedia data, including 2D shapes, textures and triangle meshes. Extensive experiments on 3D shape retrieval, hand gesture recognition and texture classification demonstrate the performance of the proposed method in comparison with state-of-the-art methods. Additionally, our approach yields higher recognition accuracy when used in conjunction with the bag-offeatures.

6.3.9. *Convergence rates for persistence diagram estimation in Topological Data Analysis*
Participants: Frédéric Chazal, Marc Glisse, Bertrand Michel.

In collaboration with Catherine Labruère (University of Burgundy)

Computational topology has recently known an important development toward data analysis, giving birth to the field of topological data analysis. Topological persistence, or persistent homology, appears as a fundamental tool in this field. In [36], we study topological persistence in general metric spaces, with a statistical approach. We show that the use of persistent homology can be naturally considered in general statistical frameworks and persistence diagrams can be used as statistics with interesting convergence properties. Some numerical experiments are performed in various contexts to illustrate our results.

6.3.10. *Stochastic Convergence of Persistence Landscapes and Silhouettes*
Participant: Frédéric Chazal.

In collaboration with Brittany Fasy (Tulane University) Fabrizio Lecci (Carnegie Mellon University) Alessandro Rinaldo (Carnegie Mellon University) Larry Wasserman (Carnegie Mellon University)

Persistent homology is a widely used tool in Topological Data Analysis that encodes multiscale topological information as a multi-set of points in the plane called a persistence diagram. It is difficult to apply statistical theory directly to a random sample of diagrams. Instead, we can summarize the persistent homology with the persistence landscape, introduced by Bubenik, which converts a diagram into a well-behaved real-valued function. In [35], we investigate the statistical properties of landscapes, such as weak convergence of the average landscapes and convergence of the bootstrap. In addition, we introduce an alternate functional summary of persistent homology, which we call the silhouette, and derive an analogous statistical theory.

6.3.11. *Subsampling Methods for Persistent Homology*
Participants: Frédéric Chazal, Bertrand Michel.

In collaboration with Brittany Fasy (Tulane University) Fabrizio Lecci (Carnegie Mellon University) Alessandro Rinaldo (Carnegie Mellon University) Larry Wasserman (Carnegie Mellon University)

Persistent homology is a multiscale method for analyzing the shape of sets and functions from point cloud data arising from an unknown distribution supported on those sets. When the size of the sample is large, direct computation of the persistent homology is prohibitive due to the combinatorial nature of the existing algorithms. We propose to compute the persistent homology of several subsamples of the data and then combine the resulting estimates. We study the risk of two estimators and we prove that the subsampling approach carries stable topological information while achieving a great reduction in computational complexity.

6.3.12. *The observable structure of persistence modules*
Participant: Frédéric Chazal.

In collaboration with Vin de Silva (Pomona College) William Crawley-Boevey (University of Leeds)

In persistent topology, q-tame modules appear as a natural and large class of persistence modules indexed over the real line for which a persistence diagram is de- finable. However, unlike persistence modules indexed over a totally ordered finite set or the natural numbers, such diagrams do not provide a complete invariant of q-tame modules. The purpose of [59] is to show that the category of persistence modules can be adjusted to overcome this issue. We introduce the observable category of persis- tence modules: a localization of the usual category, in which the classical properties of q-tame modules still hold but where the persistence diagram is a complete isomorphism invariant and all q-tame modules admit an interval decomposition.

6.4. Data Structures and Robust Geometric Computation

6.4.1. *Efficiently Navigating a Random Delaunay Triangulation*
Participants: Olivier Devillers, Ross Hemsley.

In collaboration with Nicolas Broutin (EPI RAP)

Planar graph navigation is an important problem with significant implications to both point location in geometric data structures and routing in networks. Whilst many algorithms have been proposed, very little theoretical analysis is available for the properties of the paths generated or the computational resources required to generate them. In this work, we propose and analyse a new planar navigation algorithm for the Delaunay triangulation. We then demonstrate a number of strong theoretical guarantees for the algorithm when it is applied to a random set of points in a convex region [33]. In a side result, we give a new polylogarithmic bound on the maximum degree of a random Delaunay triangulation in a smooth convex, that holds with probability one as the number of points goes to infinity. In particular, our new bound holds even for points arbitrarily close to the boundary of the domain. [56]

6.4.2. *A chaotic random convex hull*
Participants: Olivier Devillers, Marc Glisse, Rémy Thomasse.

The asymptotic behavior of the expected size of the convex hull of uniformly random points in a convex body in $\mathcal{R}^d$ is polynomial for a smooth body and polylogarithmic for a polytope. We construct a body whose expected size of the convex hull oscillates between these two behaviors when the number of points increases [62]

6.4.3. *A generator of random convex polygons in a disc*
Participants: Olivier Devillers, Rémy Thomasse.

In collaboration with Philippe Duchon (LABRI)

Let $\mathcal{D}$ a disc in $\mathbb{R}^2$ with radius 1 centered at $\mathfrak{o}$, and $(x_1, \cdots, x_n)$ a sample of n points uniformly and independently distributed in $\mathcal{D}$. Let's define the polygon P_n as the convex hull of $(x_1, \cdots, x_n)$, and $f_0(P_n)$ its number of vertices. This kind of polygon has been well studied, and it is known, see [65], that

$$\mathbb{E}f_0(P_n) = c\ n^{\frac{1}{3}} + o(n^{\frac{1}{3}})$$

where $c > 0$ is constant. To generate such a polygon, one can explicitly generate n points uniformly in $\mathcal{D}$ and compute the convex hull. For a very large quantity of points, it could be interesting to generate less points to get the same polygon, for example to have some estimations on asymptotic properties, such as the distribution of the size of the edges. We propose an algorithm that generate far less points at random in order to get P_n, so that the time and the memory needed is reduced for n large. Namely [61], we generate a number of points of the same order of magnitude than the final hull, up to a polylogarithmic factor

6.4.4. *On the complexity of the representation of simplicial complexes by trees*
Participants: Jean-Daniel Boissonnat, Dorian Mazauric.

In [46], we investigate the problem of the representation of simplicial complexes by trees. We introduce and analyze local and global tree representations. We prove that the global tree representation is more efficient in terms of time complexity for searching a given simplex and we show that the local tree representation is more ecient in terms of size of the structure. The simplicial complexes are modeled by hypergraphs. We then prove that the associated combinatorial optimization problems are very dicult to solve and to approximate even if the set of maximal simplices induces a cubic graph, a planar graph, or a bounded degree hypergraph. However, we prove polynomial time algorithms that compute constant factor approximations and optimal solutions for some classes of instances.

6.4.5. *Building Efficient and Compact Data Structures for Simplicial Complexes*
Participant: Jean-Daniel Boissonnat.

In collaboration with Karthik C.S (Weizmann Institute of Science, Israël) and Sébastien Tavenas (Max-Planck-Institut für Informatik, Saarbrücken, Germany).

The Simplex Tree is a recently introduced data structure that can represent abstract simplicial complexes of any dimension and allows to efficiently implement a large range of basic operations on simplicial complexes. In this paper, we show how to optimally compress the simplex tree while retaining its functionalities. In addition, we propose two new data structures called Maximal Simplex Tree and Compact Simplex Tree. We analyze the Compressed Simplex Tree, the Maximal Simplex Tree and the Compact Simplex Tree under various settings.

6.4.6. *Delaunay triangulations over finite universes*

Participant: Jean-Daniel Boissonnat.

In collaboration with Ramsay Dyer (Johann Bernouilli Institute, University of Groningen, Pays Bas) and Arijit Ghosh (Max-Planck-Institut für Informatik, Saarbrücken, Germany).

The witness complex was introduced by Carlsson and de Silva as a weak form of the Delaunay complex that is suitable for finite metric spaces and is computed using only distance comparisons. The witness complex $\text{Wit}(L, W)$ is defined from two sets L and W in some metric space X: a finite set of points L on which the complex is built, and a set W of witnesses that serves as an approximation of X. A fundamental result of de Silva states that $\text{Wit}(L, W) = \text{Del}(L)$ if $W = X = \mathbb{R}^d$. In this paper we give conditions on L that ensure that the witness complex and the Delaunay triangulation coincide when $W \subset \mathbb{R}^d$ is a finite set, and we introduce a new perturbation scheme to compute a perturbed set L' close to L such that $\text{Del}(L') = \text{Wit}(L', W)$. The algorithm constructs $\text{Wit}(L', W)$ in time sublinear in $|W|$.

The only numerical operations used by our algorithms are (squared) distance comparisons (i.e., predicates of degree 2). In particular, we do not use orientation or in-sphere predicates, whose degree depends on the dimension d, and are difficult to implement robustly in higher dimensions. Although the algorithm does not compute any measure of simplex quality, a lower bound on the thickness of the output simplices can be guaranteed. Another novelty in the analysis is the use of the Moser-Tardos constructive proof of the general Lovász local lemma.

MARELLE Project-Team

6. New Results

6.1. Highlights of the Year

In June 2014, Yves Bertot received the ACM Software System award, as one of the main contributors to the Coq System, along with Gérard Huet, Thierry Coquand, Christine Paulin-Mohring, Bruno Barras, Jean-Christophe Filliâtre, Hugo Herbelin, Chet. Murthy, and Pierre Castéran.

6.2. Proof and computation

Participants: Laurent Théry [correspondant], Benjamin Grégoire.

We have been continuing our effort to improve the computing power of Coq. This has led to two "computational proof":

The Erdös conjecture for n = 2 was proved this year using a SAT solver. We succeeded to formally prove this instance in Coq independently checking the 3Gb trace of the SAT solver.

The weak Goldbach conjecture was proved last year by Harald Helfgott. This proof requires a computation that the conjecture holds for numbers less than 10^{28}. This is done in two stages. The first one is to verify Goldbach conjecture for numbers less than 10^{18}. The second one is to verify the weak Goldbach conjecture for numbers less than 10^{28} using a ladder with intervals 10^{18}. The second stage has been completely verified in Coq. We are currently working on improving the computation power of Coq to make it possible to perform the first stage in reasonable time.

6.3. Formal verification of automated proof algorithms

Participant: Laurent Théry [correspondant].

We have been interested in proving that the classic 2-Sat problem can be solved in linear time. This leads to proving two classic algorithms:

1. A version of Kosaraju's algorithm that computes the strongly connected components of a directed graph [21],
2. A more direct algorithm that solves the 2-Sat problem that is using unit propagation, proposed by Alvaro del Val [20].

6.4. Formal study of cryptography

Participants: Gilles Barthe [IMDEA], Sonia Belaid [THALES and ENS], François Dupressoir [IMDEA], Pierre-Alain Fouque [Université de Rennes 1 and Institut universitaire de France], Cédric Fournet [Microsoft Research], Benjamin Grégoire [correspondant], Benedikt Schmidt [IMDEA], Pierre-Yves Strub [IMDEA], Nikhil Swamy [Microsoft Research], Mehdi Tibouchi [NTT Secure Platform Laboratories], Santiago Zanella-Béguelin [Microsoft Research], Jean-Christophe Zapalowicz [Inria].

The goal of this work is to provide a friendly tool easily usable by cryptographers without knowledge of formal proof assistants. The idea is to use the techniques formally proved in Certycrypt and to call SMT-provers. We provide two differents tools EasyCrypt and ZooCrypt.

This year, we worked on the following topics:

- Relational program logics, as used in EasyCrypt, have been used for mechanizing formal proofs of various cryptographic constructions. In [15], we present rF*, a relational extension of F*, a general-purpose higher-order stateful programming language with a verification system based on refinement types. The distinguishing feature of rF* is a relational Hoare logic for a higher-order, stateful, probabilistic language.

- Fault Attacks are attacks in which an adversary with physical access to a cryptographic device, say a smartcard, tampers with the execution of an algorithm to retrieve secret material. In [13] we propose a new approach for finding fault attacks based on fault conditions. Using the method, we discover multiple fault attacks on RSA and ECDSA. Several of the attacks found by our tool are new. In [14], we propose a new counter measure to make RSA-PSS provably secure against non-random faults. We also prove the result using EasyCrypt.

- Many algorithms, particularly in cryptography, admit very efficient batch versions that compute simultaneously the output of the algoritms on a set of inputs. AutoBatch is a tool that computes highly optimized batch verification algorithms for pairing based signature schemes. In [12], we use EasyCrypt to formalise the methods used by AutoBatch and to automatically certify the result of the transformation performed by AutoBatch.

- We study the problem of automatically verifying higher-order masking countermeasures which is used to protect implementations where the attacker can observe intermediate computations (like in a smartcard). We propose an efficient method to check the correctness and the security of masked implementation. This work has been submitted to EuroCrypt 2015. We start the ANR BRUTUS on this subject.

6.5. Formalization of Bourbaki's sets and ordinals

Participant: José Grimm.

In previous years we developed a formal library describing the parts of the Bourbaki books on set theory, cardinals and ordinals. We completed it by adding the definition of real numbers using Dedekind cuts. The important properties we showed that **R** is an ordered Archimedean field, that every non-empty bounded subset has a least upper bound, that every Cauchy sequence has a limit, and that the intermediate value theorem holds.

It follows that every positive real number has positive square root. We give a pair of adjacent sequences that converges to this square root. For instance $\sqrt{2}$ is irrational, and we get a pair of rational adjacent sequences that converges to it. This produces an explicit order isomorphism $\mathbf{Q}^* \to \mathbf{Q}$. The number of such isomorphisms is equal to the power of the continuum (the cardinal of **R**) [18].

6.6. Stern-Brocot and Fibonacci sequences

Participant: José Grimm.

We constructed an explicit bijection $\mathbf{N} \to \mathbf{Q}$, first in the framework of the Bourbaki project (see above), then in Ssreflect. Every positive rational number x can uniquely be written as a quotient s_n/s_{n+1}. This result was established by Dijkstra who stated it in an obfuscated way. It was shown years before by Stern. It is possible to compute s_n/s_{n+1} without computing numerator and denominator separately, by considering the sequences of bits of n from left to right or from right to left. Truncating the binary expansion of n yields a sequence of approximations to s_n/s_{n+1} (this was studied by Brocot, and the so-called Stern-Brocot tree is an alternative representation of rational numbers). We implemented the work of Dijkstra and Stern in Coq [17].

We also studied how a number can be represented by a sequence of other numbers (for instance as a sum of distinct Fibonacci numbers, with or without constraints). The number of ways of writing n as a sum of powers of two, each power of two being used at most twice, is s_{n+1}. These results are presented in [17].

6.7. Formal proof that e and π are transcendental

Participants: Sophie Bernard, Laurence Rideau.

We constructed formal proofs that π is irrational, e is transcendental, and π is transcendental. These proofs share a common initial pattern, where rationality or algebraicity of the mathematical constants are shown to imply the existence of a sequence of positive integers that must decrease indefinitely.

This proof development is an opportunity to study the interplay between several existing libraries about algebraic structures and analysis: the ssreflect library for algebra and the Coquelicot library for calculus. Moreover, the proof that π is transcendental was an occasion to test the newly developed module on symmetric polynomials by P.-Y. Strub at IMDEA.

6.8. Fast computation of π

Participant: Yves Bertot.

In the previous year, we studied a proof that π could be approximated with a fast converging sequence based on arithmetic geometric means. This year we described a proof that rounding errors during this computation could be guaranteed as small as needed, based on a study of derivatives. This approach provides a fruitful alternative to interval-based approaches. The result was published in [16].

We also completed a journal paper on various ways to observe and compute the number π [7].

6.9. Decision procedures for polynomials

Participant: Yves Bertot.

Following up on the work in previous years around Bernstein Polynomials, we implemented a decision procedure for guaranteeing the sign of a polynomial function inside an interval, using Bernstein polynomials and dichotomy. In the long run, we hope to explore two approaches, one based on the off-line computation of certificates for sub-intervals (these certificates are easy to verify), and one based on implementing computational reflection. This approach should also generalize quite easily to multi-variate polynomials.

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

6. New Results

6.1. Source recovery problems

Participants: Laurent Baratchart, Sylvain Chevillard, Juliette Leblond, Christos Papageorgakis, Olga Permiakova, Dmitry Ponomarev.

The research in this section is partly joint work with Qian Tao (Univ. Macao).

It was proved in [38] that a vector field with $n + 1$ components on $\mathbb{R}^n$ can be expressed uniquely as the sum of (the trace on $\mathbb{R}^n$ of) a harmonic gradient in the upper half-space, of (the trace on $\mathbb{R}^n$ of) a harmonic gradient in the lower half-space, and of a tangential divergence free vector field on $\mathbb{R}^n$. This decomposition, that we call the *Hardy-Hodge* decomposition, is valid not only for L^p vector fields as mentioned in Section 3.3.1 , but in much more general distribution spaces like $W^{-\infty,p}$ which contains all distributions with compact support or $BMO^{-\infty}$ which contains all finite sums of derivatives of bounded functions. This year we extended the decomposition to smooth hypersurfaces, where divergence-free distributions may be defined as those annihilating tangential gradient vector fields. We also studied the case where the hypersurface is only Lipschitz smooth, and then we proved the decomposition in L^p provided that p is close enough to 2 (how close depends on the Lipschitz constant of the hypersurface).

The Hardy-Hodge decomposition was used in [38] to find the kernel of the planar magnetization operator, namely a potential of the form (1) with m supported in a plane generates the zero field above that plane if, and only if there is no harmonic gradient from below in the Hardy-Hodge decomposition of m. The above mentioned generalization is now to the effect that a magnetization supported on a bounded closed surface (*e.g.* a sphere) is silent in the unbounded component of the complement of that surface if, and only if there is no harmonic gradient from inside in its Hardy-Hodge decomposition. An article is being written on this topic.

We also considered the case where m is compactly supported in the bounded component of the complement of that surface. Then m is silent if and only if it is the sum of a divergence-free distribution and of finitely many derivatives of gradients of Sobolev functions having zero trace on the surface [41].

These results shed light on the indeterminacy of inverse source problems.

6.1.1. EEG

This work is conducted in collaboration with Maureen Clerc and Théo Papadopoulo from the Athena EPI, and with Jean-Paul Marmorat (Centre de mathématiques appliquées - CMA, École des Mines de Paris).

In 3-D, functional or clinically active regions in the cortex are often modeled by point-wise sources that must be localized from measurements of a potential on the scalp. Inside the cortex, identified to a ball after the cortical mapping step, the potential satisfies a Poisson equation whose right-hand side is a linear combination of gradients of Dirac masses (the sources in EEG). In the work [3] it was shown how best rational approximation on a family of circles, cut along parallel planes on the sphere, can be used to recover the sources when they are at most 2 of them. Later, results on the behavior of poles in best rational approximation of fixed degree to functions with branch points [6] helped justifying the technique for finitely many sources (see section 4.2).

The dedicated software FindSources3D (see section 5.6), developed, in collaboration with the team Athena and the CMA, dwells on these ideas. Functions to be approximated in 2-D slices turn out to have additional *multiple* poles at their branch points so that, in the rational approximation step, it is beneficial to consider approximants with multiple poles as well (for EEG data, one should consider *triple* poles). Though numerically observed in [9], there is no mathematical justification so far why these multiple poles are attracted more strongly than simple poles to the singularities of the approximated function. This intriguing property, however, definitely helps source recovery [28]. This year we used it to automatically estimate the "most plausible"

number of sources (numerically: up to 3 at the moment). Such enhancements were prompted by a developing collaboration with the BESA company, which is interested in automatic detection of the number of sources (which was left to the user until recently).

Soon, magnetic data from MEG (magneto-encephalography) will become available together with EEG data; indeed, it is now possible to use simultaneously the corresponding measurement devices. We expect this to improve the accuracy of our algorithms.

In relation to other brain exploration modalities like electrical impedance tomography (EIT, see [16]), we also consider identifying electrical conductivity in the head. This is the topic of the PhD of C. Papageorgakis, co-advised with the Athena project-team and BESA GmbH. Specifically, in layered models, we are concerned with estimating conductivity of the skull (intermediate layer). Indeed, the skull consists of a hard bone part, the conductivity of which is more or less known, and spongy bone compartments whose conductivities may vary considerably with individuals.

A preliminary question in this connection is: can one uniquely recover a homogeneous skull conductivity from a single EEG recording when the sources and the conductivities of other layers are known? And if sources are not known, which additional information do we need? These are issues currently under investigation. To put them into perspective, recall the famous Caldèron problem of deducing a bounded (nonconstant) conductivity from the knowledge of all possible pairs consisting of a potential and its current flux at the boundary. In dimension 3, when the conductivity is not smooth (less than $3/2$ of a derivative), it is unknown whether the problem is even injective (*i.e.* if two conductivities can have the same pairs of boundary potential and flux). A weaker, discrete version of this problem is: if the conductivity takes on finitely many values and the geometry of the level sets is known, does a finite set of pairs of boundary potential and flux allow one to recover it? This is a significant question to be tackled for the purpose of source recovery in EEG with known geometry but unknown conductivities inside the head.

6.1.2. *Inverse Magnetization problems*

This work is carried out in the framework of the "équipe associée Inria" IMPINGE, comprising Eduardo Andrade Lima and Benjamin Weiss from the Earth Sciences department at MIT (Boston, USA) and Douglas Hardin and Edward Saff from the Mathematics department at Vanderbilt University (Nashville, USA),

Localizing magnetic sources from measurements of the magnetic field away from the support of the magnetization is the fundamental issue under investigation by IMPINGE. The goal is to determine 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 (superconducting coil devices). Currently, rock samples are cut into thin slabs and the magnetization distribution is considered to lie in a plane, which makes for a somewhat less indeterminate framework than EEG because "less" magnetizations can produce the same field (for the slab has no inner volume). Note however that EEG data consist of both potential and current values at the boundary, whereas in the present setting only values of the normal magnetic field are provided to us.

Figure 5 presents a schematic view of the experimental setup: the sample lie on a horizontal plane at height 0 and its support is included in a rectangle. The vertical component B_z of the field produced by the sample is measured on points of a horizontal $N \times N$ rectangular grid at height h.

We set up last year a heuristic procedure to recover regularly spaced dipolar magnetizations, *i.e.* magnetizations composed of dipoles placed at the points of a regular rectangular $n \times n$ grid. The latter seems general enough a model class to approximate magnetizations commonly encountered in samples. However, for reasons of computational complexity, n is significantly smaller than N which limits the power of the model. Each dipole of the $n \times n$ grid is determined by the 3 components of its moment, thus the magnetization can be represented by a real $3n^2$-vector. If we denote by A the matrix of the operator that maps such a vector X to the vector b of measurements (which belongs to $\mathbb{R}^{N^2}$), we want to find X such that AX is close to b. For computational simplicity, we use a Euclidean criterion $\|AX - b\|_2$, which reduces the problem to a singular value decomposition of A. The inverse problem being ill-posed, A is poorly conditioned and we must resort to a regularization technique. The one we developed initially has been based on iteratively cropping the support of b, using a threshold on the intensity of the dipoles at each step, so as to reduce the number of active components
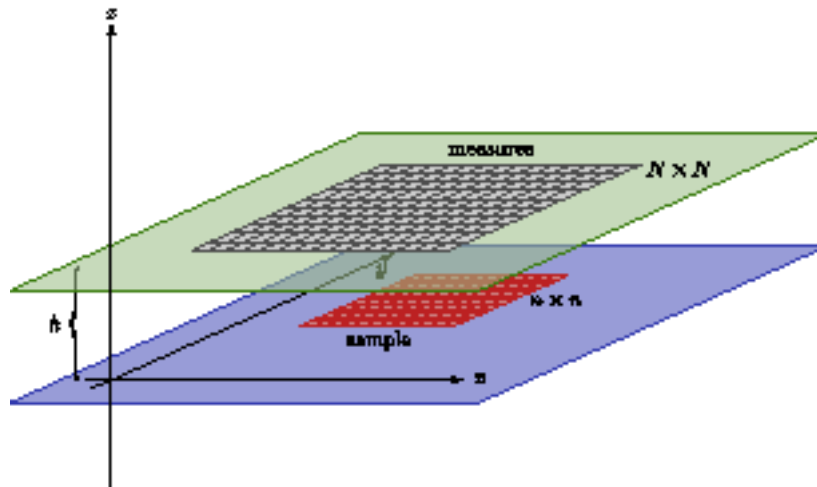
Figure 5. Schematic view of the experimental setup

in b. Preliminary experiments were performed last year on synthetic data and also on a real example (Lonar spherule).

This year, we performed more systematic experiments on real data (namely Allende chondrules and Hawaian basalt) provided by the SQUID scanning microscope at MIT lab. Cropping the support of b using thresholding has proved efficient to improve ill-conditioning for samples with localized support embedded in the slab (*e.g.*, chondrules). On the other hand, when the support of the sample is spread out (*e.g.*, Hawaian basalt), the reduction of active components of b was insignificant. We used this inversion procedure to estimate the net moment. The importance of the latter has been emphasized by the geophysicists at MIT for at least two reasons: firstly it yields important geological information on the sample in particular to estimate the magnitude of the ambient magnetic field at the time the rock was formed. Secondly, it may to some extent be measured independently, using a magnetometer, thereby allowing one to cross-validate the approach. A third, computational reason is that knowledge of the net moment should pave the way to a numerically stable reconstruction of an equivalent unidirectional magnetization. The support of the latter would provide us with valuable information to test for unidirectionality of the true magnetization, which is an important question to physicists.

When the support can be significantly shrunk while keeping the residue small (*i.e.*, explaining the data satisfactorily), estimates of the net moment based on the dipolar model obtained by inversion seem to be good. They apparently supersede the measurements by magnetometers as well as by dipole fitting procedures set up at MIT. It is interesting to notice that the magnetization obtained by our inversion procedure, either before or after shrinking the support, often does not resemble the true magnetization, even when it yields correct moment and field. This can be seen on synthetic examples and may be surmised on real data, thereby confirming that recovering the net moment and recovering the magnetization are rather different problems, the latter being considerably more ill-posed than the former.

One specific difficulty with chondrule-type examples has been to account for their thickness: they are indeed small spheres and their 3-D character cannot be completely ignored. In order to use the inversion procedure set up in the plane, we investigated the following question. Assume that the sample has some thickness, but small enough that the magnetization at a point $P = (x, y, z)$ inside the sample depends only on x and y (possibly weighted by some function that depends only on z), *i.e.* that it is of the form $m(x, y)\phi(z)$. If we consider a (truly) planar magnetization with the same distribution $m(x, y)$ but on a plane lying at some nonzero height

ε, how to choose ε so as to produce a field at height h which is closest to the field produced by the thick magnetization? This has been the object of the internship of Olga Permiakova who used local expansion of the dipole-to-field map (see her report [0]). An article is being written on this subject.

The case where the magnetization is flat but spread out on the sample is more difficult. First of all, the computational effort becomes significant and led us to use the cluster at Inria Sophia Antipolis. We succeeded in obtaining full inversions for the Hawaian basalt. The residue (approximation error) is moderate but not impressively small, which indicates that we reach the limit of modeling magnetizations by a regular grid of dipoles. However the computation of the moment compares favorably with estimates previously obtained by a different technique at MIT lab. Still, using a cluster and two days of evaluation to obtain a coarse estimate of the net moment of a sample is rather inefficient and calls for new investigations.

We also experimented an alternative regularization procedure, based on L^2 minimization under L^1 penalty as solved by the SALSA algorithm. Such methods are quite popular today for sparse recovery. However, the computational load, as well as the quality of the results, do not differ significantly from those obtained previously.

We now develop new methods in order to estimate the net moment of the magnetization, based on improvements of previously used Fourier techniques, and recently we reformulated the problem with the help of Kelvin transforms. It has been realized that the success of net moment recovery hinges on the ability to extrapolate the measurements. In particular, we managed to considerably improve previous estimates by means of data extension based on dipolar field asymptotics.

In the course of inverting the field map, we singled out magnetizations which are numerically (almost) silent from above though not from below. This illustrates how ill-posed (unstable) the problem, as theory predicts that no compactly supported magnetization can be exactly silent from above without being also exactly silent from below. Although such magnetizations seem to have small moment and therefore do not endanger the possibility of recovering the net moment, their existence is certainly an obstacle to inversion of the field map without extra measurements or hypotheses (*e.g.*, measuring from below or assuming unidirectionality).

In the course of the doctoral work by D. Ponomarev, the study of the 2D spectral decomposition of the truncated Poisson operator has been undertaken. It is a simplified version of the relation between the magnetization and the magnetic potential. We considered several formulations in terms of singular integral equations and matrix Riemann-Hilbert problems, and focused on finding closed form solutions for various approximations of the Poisson operator in terms of a the ratio between the distance h to the measurement plane and the sample support size.

Lately, Apics became a partner of the ANR project MagLune, dealing with Lunar magnetism, a in collaboration with the Geophysics and Planetology Department of Cerege, CNRS, Aix-en-Provence, see section 8.2.2 . The research is just starting, and will focus on computing net moments of lunar rock samples collected by NASA.

6.2. Boundary value problems

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

Collaboration with Laurent Bourgeois (ENSTA ParisTech, Lab. Poems), Elodie Pozzi (Univ. Bordeaux, IMB), Emmanuel Russ (Univ. Grenoble, IJF).

Generalized Hardy classes
As we mentioned in section 4.4 , 2-D diffusion equations of the form $\mathrm{div}(\sigma \nabla u) = 0$ with real non-negative valued conductivity σ can be viewed as compatibility conditions for the so-called conjugate Beltrami equation: $\overline{\partial} f = \nu \overline{\partial f}$ with $\nu = (1 - \sigma)/(1 + \sigma)$ [4]. Thus, the conjugate Beltrami equation is a means to replace the initial second order diffusion equation by a first order system of two real equations, merged into a single complex one. Hardy spaces under study here are those of this conjugate Beltrami equation: they are comprised of solutions to that equation in the considered domain whose L^p means over curves tending to the boundary

---

[0] http://www-sop.inria.fr/apics/IMPINGE/Documents/Report_Permiakova_Olga.pdf

of the domain remain bounded. They will for example replace holomorphic Hardy spaces in problem (P) when dealing with non-constant (isotropic) conductivity. Their traces merely lie in L^p ($1 < p < \infty$), which is suitable for identification from point-wise measurements, and turn out to be dense on strict subsets of the boundary. This allows one to state Cauchy problems as bounded extremal issues in L^p classes of generalized analytic functions, in a manner which is reminiscent of what we discussed for analytic functions in section 3.3.1 .

The study of such Hardy spaces for Lipschitz σ was reduced in [4] to that of spaces of pseudo-holomorphic functions with bounded coefficients, which were apparently first considered on the disk by S. Klimentov. Solutions factorize as $e^s F$, where F is a holomorphic Hardy function while s is in the Sobolev space $W^{1,r}$ for all $r < \infty$ (Bers factorization), and the analog to the M. Riesz theorem holds which amounts to solvability of the Dirichlet problem with L^p boundary data. The case of finitely connected domains was carried out in [14].

This year, we addressed in [25] the uniqueness issue for the classical Robin inverse problem on a Lipschitz-smooth domain $\Omega \subset \mathbb{R}^n$, with L^∞ Robin coefficient, L^2 Neumann data and isotropic conductivity of class $W^{1,r}(\Omega)$, $r > n$. The Robin inverse problem consists in recovering the ratio of the normal derivative and the solution (the so-called Robin coefficient) on a subset of the boundary, knowing them on the complementary subset. We showed that uniqueness of the Robin coefficient on a subset of the boundary, given Cauchy data on the complementary subset, does hold when $n = 2$ whenever the boundary subsets are of positive Lebesgue measure. We also showed that this no longer holds in higher dimension, and we gave counterexamples when $n = 3$. The subsets in these counterexamples look very bad, and it is natural to ask whether uniqueness prevails if they have interior points. This raises an interesting open issue on harmonic gradients, namely: can a nonzero harmonic function vanish together with its normal derivative on a subset of the boundary of positive measure, and still the Robin coefficient is bounded in a neighborhood of that set? This question is worth investigating

Best constrained analytic approximation
Several questions about the behavior of solutions to the bounded extremal problem (P) in section 3.3.1 , and of some generalizations thereof, are still under study by Apics.. We considered additional interpolation constraints on the disk in problem (P), and derived new stability estimates for the solution [24]. An article is being written on the subject. Ongoing work is geared towards applications of [24]. New insight leads us to relate these results to overdetermined boundary value problems for 2D Laplace equations on irregular boundaries. This has applications in set-ups where measurements are obtained from oddly distributed sensors. Treating some of the measurements as pointwise interpolation constraints seems a reasonable strategy in comparison with interpolation of the data along a geometrically complicated boundary. Such interpolation constraints arise naturally in inverse boundary problems like plasma shaping, when some of the measurements are performed inside the chamber of the tokamak, see section 4.4 .

6.3. Matching problems and their applications - De-embedding of filters in multiplexers

Participants: Laurent Baratchart, Martine Olivi, Sanda Lefteriu, David Martinez Martinez, 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 (Siae Microelettronica, Milan, Italy).

6.3.1. *Matching problems and their applications*

Filter synthesis is usually performed under the hypothesis that both ports of the filter are loaded on a constant resistive load (usually 50 Ohm). In complex systems, filters are however cascaded with other devices, and end up being loaded, at least at one port, on a non purely resistive frequency varying load. This is for example the case when synthesizing a multiplexer: each filter is here loaded at one of its ports on a common junction. Thus, the load is by construction non constant with the frequency, and not purely resistive either. Likewise, in an emitter-receiver, the antenna is followed by a filter. Whereas the antenna can usually be regarded as a resistive load at some frequencies, this is far from being true on the whole working band. A mismatch between

the antenna and the filter, however, causes irremediable power losses, both in emission and transmission. Our goal is therefore to develop a filter synthesis method that allows to match varying loads on specific frequency bands.
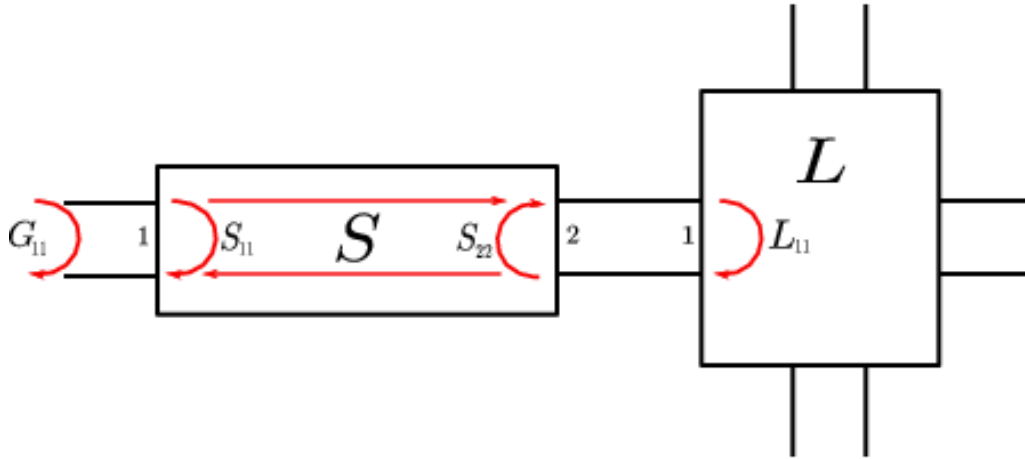

Figure 6. Filter plugged on a system with reflexion coefficient L_{11}

Figure 6 shows a filter with scattering matrix S, plugged at its right port on a frequency varying load with reflexion parameter $L_{1,1}$. If the filter is lossless, simple algebraic manipulations show that on the frequency axis the reflexion parameter satisfies:

$$|G_{1,1}| = \left| \frac{S_{2,2} - \overline{L_{1,1}}}{1 - S_{2,2}L_{1,1}} \right|.$$

The matching problem of minimizing $|G_{1,1}|$ amounts therefore to minimize the pseudo-hyperbolic distance between the filter's reflexion parameter $S_{2,2}$ and the load's reflexion $L_{1,1}$, on a given frequency band. For a broad class of filters, namely those that can be modeled by a circuit of n coupled resonators, the scattering matrix S is a rational function of McMillan degree n in the frequency. The matching problem appears therefore as a rational approximation problem in hyperbolic metric. When n is fixed, the latter is non-convex and led us to seek methods to derive good initial guesses for classical descent algorithms. To this effect, if $S_{2,2} = p/q$ we considered the following interpolation problem: given n frequency points $w_1 \cdots w_n$ and a transmission polynomial r, to find a unitary polynomial p of degree n such that:

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

where q is the unique monic Hurwitz polynomial of degree n satisfying the Feldtkeller equation

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

which accounts for the losslessness of the filter. This problem can be seen as an extended Nevanlinna-Pick interpolation problem, that was considered in [67] when the interpolation points w_j lie in the *open* left half-plane. The method in the last reference does not extend to imaginary interpolation point and we used rather different, differential-topological techniques to prove that this problem has a unique solution, which can be computed by continuation. In the setting of multiplexer synthesis, where this result must e applied recursively to each filter, we showed the existence of a fixed point for the tuning procedure, based on Brouwer's fixed point theorem. These results were presented at the MTNS [18], at the plenary of session of Ernsi workshop 2014, and they lie at the heart of the ANR Cocoram on co-integration of filters and antennas (8.2.1). Implementation of the continuation algorithm has been done under contract with CNES and yields encouraging results. Generalizations of the interpolation problem where the monic condition is relaxed are under study in the framework of co-integration of filters and antennas.

6.3.2. *De-embedding of multiplexers*

This work is pursued in collaboration with Thales Alenia Space, Siae Microelettronica, Xlim and under contract with CNES-Toulouse (see section 7.1).

Let S be the scattering parameters 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 7 . The de-embedding problem is to recover the F_k and it can be stated under various hypotheses. Last year we studied this problem when S and T are known [79] but no special structure for the F_k is assumed. It was shown that for generic T and for $N > 3$, the de-embedding problem has a unique solution. In practice, however, the junction's response is far from being generic, as it is usually obtained *via* assembly of T-junctions. This makes the problem extremely sensitive to measurement noise. It was also noticed that in practical applications, scattering measurements of the junction are hardly available.

It is therefore natural to consider the following de-embedding problem. Given S, and under the assumption that

- the F_k are rational of known McMillan degree,
- the coupling geometry of their circuital realization is known,

what can be said about the filter's response? Note that the above assumptions do not bear on the junction. Nevertheless, we showed that the filter's responses are identifiable up to a constant matrix chained at their nearest port to the junction [73]. It was proved also that the uncertainty induced by the chain matrix bears only on the resonant frequency of the last cavity of each filter, as well as on their output coupling. Most of the filters' parameters can therefore be recovered in principle. The approach is constructive and relies on rational approximation to certain scattering parameters, as well as on some extraction procedure similar to Darlington's synthesis. Software development is under way and experimental studies have started on data provided to us by Thales Alenia Space and by Siae Microelettronica. A mid-term objective is to extend Presto-HF (see Section 5.3) so as to handle de-embedding problems for multiplexers and more generally for multi-ports.

6.4. Stability of amplifiers

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

This work is performed under contract with CNES-Toulouse and the University of Bilbao. The goal is to help designing 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 detect instability before an amplifying circuit is physically built. Our goal is 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 in the small signal regime. Using this approximation, diodes appear as negative resistors and transistors as current sources controlled by the voltage at certain nodes of the circuit.
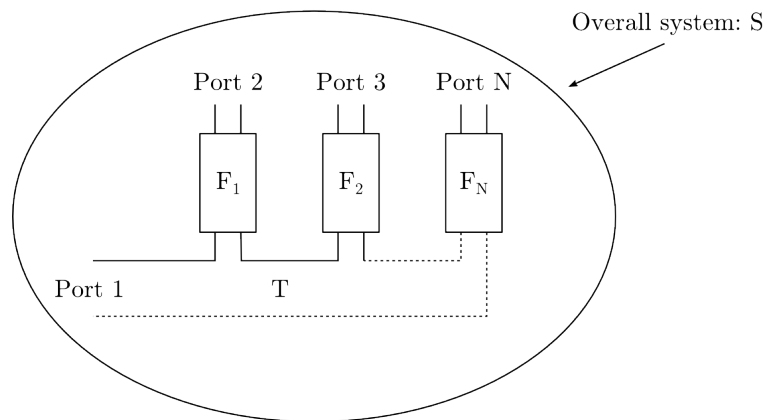
Figure 7. Multiplexer made of a junction T and filtering devices $F_1, F_2 \cdots F_N$

Over the last three years, we studied several features of transfer functions of amplifying electronic circuits:

- We characterized the class of transfer functions which can be realized with ideal components linearized active components, together with standard passive components (resistors, inductors, capacitors and transmission lines). It is exactly the field of rational functions in the complex variable and in the hyperbolic cosines and identity-times-hyperbolic sines of polynomials of degree 2 with real negative roots.

- We introduced a realistic notion of stability, by terming stable a circuit whose transfer function belongs to H^∞, as long a sufficiently high resistor is added in parallel to that circuit.

- We constructed unstable circuits having no pole in the right half-plane, which came as a surprise to our partners.

- In order to circumvent these pathological examples, we introduced a realistic hypothesis that there are small inductive and capacitive effects to active components. Our main result is that a realistic circuit without poles on the imaginary axis is unstable if and only if it has poles in the right half-plane. Moreover, there can only be finitely many of them.

This year, we were led to modify our definition of stability, taking a hint from scattering theory. We say that a transfer function Z is stable whenever $(R - Z)/(R + Z)$ belongs to H^∞ with uniformly bounded H^∞-norm for all R large enough. Equivalently, this means that the circuit can amplify signals but not require an unbounded amount of energy from the primary power circuit. This new definition is really about energy, hence is more natural. Also, it allows us a unified characterization in the corner case where instabilities are located on the imaginary axis. We obtained this way a nice characterization: Z is stable if and only if it has no pole in the open right half plane, while each pole it may have on the imaginary axis is simple and has a residue with strictly positive real part. We published a research report [23] and an article is being written to report on our results.

6.5. Approximation

Participant: Laurent Baratchart.

6.5.1. Orthogonal Polynomials

This is joint work with Nikos Stylianopoulos (Univ. of Cyprus).

We study the asymptotic behavior of weighted orthogonal polynomials on a bounded simply connected plane domain Ω. The n-th orthogonal polynomial P_n has degree n, positive leading coefficient, and satisfies

$$\int_\Omega P_n \overline{P}_k w \, dm = \delta_{n,k}$$

where w is an integrable positive weight and $\delta_{n,k}$ is the Kronecker symbol. When Ω is smooth while w is Hölder-continuous and non-vanishing, it is known that

$$P_n(z) = \left(\frac{n+1}{\pi}\right)^{1/2} \Phi^n \frac{\Phi'}{S_w(z)} \{1 + o(1)\},$$

locally uniformly outside the convex hull of $\overline{\Omega}$, where Φ is the conformal map from the complement of Ω onto the complement of the unit disk and S_w is the so-called Szegö function of the trace of w on the boundary $\partial\Omega$ [81]. If we compare it with classical exterior Szegő asymptotics, the formula asserts that P_n behaves asymptotically like the n-th orthogonal polynomial with respect to a weight supported *on $\partial\Omega$* (the trace of w), up to a factor $\sqrt{(n+1)/\pi}$.

When Ω is the unit disk, we proved this result under unprecedented weak assumptions on w, namely $w(re^{i\theta})$ should converge in $L^p(T)$ as $r \to 1$ for some $p > 1$ and its $\log^-$ should be bounded in the real Hardy space H^1. An article is being written on these findings and the case of a smooth domain Ω, more general than a disk, is under examination.

6.5.2. *Meromorphic approximation*

This is joint work with Maxim Yattselev (Purdue Univ. at Indianapolis, USA).

We proved in [6] that the normalized counting measure of poles of best H^2 approximants 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 (the normalized counting measure is the probability measure with equal mass at each pole). This result warrants source recovery techniques used in section 6.1.1 . Here we consider the corresponding problem for best uniform meromorphic approximants on the unit circle (so-called AAK approximants after Adamjan, Arov and Krein), in the case where the function may have poles and essential singularities. This year, we established a similar result when the function has finitely many essential singularities. The general case is still pending.

<center>**ECUADOR Project-Team**</center>

6. New Results

6.1. Resolution of linearised systems

Participants: Olivier Allain [Lemma], Gautier Brèthes, Alain Dervieux, Bruno Koobus, Stephen Wornom.

For Fluid Mechanics as well as for Structural Mechanics, implicit time-advancing is mandatory. It can be applied efficiently if the large systems involved are solved with a good parallel algorithm. In the 90's, a generation of solution algorithms was devised on the basis of Domain Decomposition Methods (DDM). For complex models such as compressible flows, Schwarz DDM were combined with quasi-Newton algorithms like GMRES. For example in the Aironum tool, we use Restrictive Additive Schwarz (RAS, developed by Cai and Farhat). RAS is an ancestor of the widely used class of Newton-Krylov-Schwarz (NKS) algorithms. NKS are, in some versions including RAS, almost scalable *i.e.* their convergence rate is independant of the number of processors. But scalability degrades over a thousand processors. During the ANR ECINADS, coordinated by Ecuador, a Coarse-Grid Deflated RAS was developed. The algorithmic scalability (iteration-wise) holds for all part, except for the coarse grid direct solver, which concerns a much smaller problem. Effective Convergence Scalability (ECS) was confirmed up to 2048 processors. After this level, the asymptotic complexity of the coarse-grid direct solver become predominant and ECS is lost. In other words, with a Coarse-Grid Deflated RAS, the size of the coarse grid problem must be limited in order to enjoy ECS.

In the thesis of Gautier Brèthes, we now study a further step towards super-massive scalability: we use a number of fine and medium grids in order to solve the target large system by a multi-mesh multigrid (MG) algorithm. An important novelty is that the complete FMG technology is applied, with a new stopping criterion avoiding useless cycling [12]. It is well-known that parallel MG is limited to "no-too-coarse" coarse levels due to an excessive ratio between communication and computation. Now our parallel MG can be complemented by the previous version of the solver (deflated RAS) for this no-too-coarse coarse level.

6.2. Algorithmic Differentiation of a CFD code

Participants: Valérie Pascual, Laurent Hascoët, Alain Dervieux.

This activity continues in collaboration with the partners of the FP7 project UMRIDA. The team is assisting Alenia-Aermacchi in the efficient differentiation of its Euler/Navier Stokes UNS3D code in tangent mode, using in particular a differentiable extension of the MPI library.

Inside a collaboration with EDF, Valérie Pascual is also applying Tapenade to produce various adjoint differentiations of the hydrographic code Mascaret.

6.3. Control of approximation errors

Participants: Gautier Brèthes, Eléonore Gauci, Alain Dervieux, Adrien Loseille [Gamma team, Inria-Rocquencourt], Frederic Alauzet [Gamma team, Inria-Rocquencourt], Stephen Wornom, Olivier Allain [Lemma], Anca Belme [University Paris VI].

A study of an interesting combination of Full Multigrid (FMG) and Anisotropic mesh Adaptation (AA) started last year, with the beginning of the thesis of Gautier Brèthes. FMG is one of the (very) few algorithm giving N results by consuming kN floats. Anisotropic adaptation produces approximation errors less than ε with $N = \varepsilon^{-\frac{1}{dim}}$ nodes, this for smooth and non-smooth solution fields. Anisotropic adaptative FMG may produce approximation errors less than ε by consuming $k\varepsilon^{-\frac{1}{dim}}$ floats. Moreover, theory and experiments show that FMG works better when combined with AA. A first AA-FMG platform has been developed. It combines several mesh-adaptation modules developed by Gamma and Distene. It is used for testing new adaptation criteria.

Third-order mesh adaptation was the main topic of last year in error control. The scheme is the ENO finite-volume formulation with quadratic reconstruction. An article describing our results for 2D applications is being written. A 3D version is developed in the Aironum CFD platform. A cooperation with Lemma is also running, with Eléonore Gauci, to apply the scheme to fluid-gas interfaces. Further studies of mesh adaptation for viscous flows are ongoing and an article in collaboration with Gamma3 and University Paris VI (Anca Belme) is being written.

An important novelty in mesh adaptation is the norm-oriented AA method. The method relies on the definition of ad hoc correctors. It has been developed in the academic FMG platform for elliptic problems. Another version is developed by Gamma, in collaboration with Ecuador, for the compressible flow models. The purpose is to devise a composite algorithm in which an approximation error norm can be specified by the user. The introduction of the norm-oriented idea considerably amplifies the impact of adjoint-based AA. The applied mathematician and the engineer now have methods when faced to mesh adaptation for the simulation of a complex PDE system, since they can specify which error norm level they wish, and for which norm. Eléonore Gauci starts a thesis, co-advised by Alain Dervieux and Frédéric Alauzet, on the norm-oriented criteria for CFD and coupled CSM-CFD systems. She also works on a new version of the mesh adaptive CFD demonstrator of Gamma3. This new version improves the resolution of curved features. A cooperation is also starting between Gautier Brèthes and Thierry Coupez (Ecole Centrale de Nantes) on discrete metrics.

These studies are supported by an European FP7 project UMRIDA which deals with the application of AA to approximation error modelling and control, and by ANR project MAIDESC coordinated by Ecuador, which deals with meshes for interfaces, third-order accuracy, meshes for boundary layers, and curved meshes.

6.4. Turbulence models

Participants: Emmanuelle Itam [University Montpellier II], Alain Dervieux, Bruno Koobus, Carine Moussaed [University Montpellier II], Maria-Vittoria Salvetti [University of Pisa], Stephen Wornom, Bruno Sainte-Rose [Lemma].

The purpose of our work in hybrid RANS/LES is to develop new approaches for industrial applications of LES-based analyses. This year, many experiments have validated the dynamic version of our VMS-LES. The quality of simulations is either comparable to non-dynamic, or better. In the applications targeted (aeronautics, hydraulics), the Reynolds number can be as high as several tenth millions, far too high for pure LES models. However, certain regions in the flow can be 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. An important difference between a real life flow and a wind tunnel or basin is that the turbulence of the flow upstream of each body is not well known. This year, we have started the study of multiple-body flows. A typical case is the interaction between two parallel cylinders, one being in the wake of the other. A recent workshop showed the rather disastrous predictions of LES models. Most hybrid models behave better, mainly for the first cylinder. We are progressively extending and validating our VMS-LES model and our hybrid ones ([11]).

6.5. AD tools infrastructure

Participants: Laurent Hascoët, Paul Hovland [Argonne National Lab. (Illinois, USA)], Sri Hari Krishna Narayanan [Argonne National Lab. (Illinois, USA)].

We have an ongoing partnership with Paul Hovland's team at Argonne National Lab, formalized by joint participation in the Inria-Illinois joint lab on petascale computing and with travels funded by the Partner University Fund (PUF) of the French embassy in the USA.

In this framework, we worked on the goal of blending our AD tool Tapenade with Argonne's tool OpenAD, buy developing bridges between their internal representations, through a common external representation of analyzed programs. This representation called XAIF is based on XML. We have developed running prototypes of these bridges in both directions, that run on a few examples and that need further development to allow each tool to take advantage of the other's analyses and models. This was supported by two visits of Krishna Narayanan to Inria and one of Laurent Hascoët to Argonne.

We also continued joint development of the Adjoinable-MPI library (AMPI) that provides efficient tangent and adjoint differentiation for MPI-parallel codes, independently of the AD tool used (now AdolC, dco, OpenAD, Tapenade).

We also extracted from Tapenade a standalone kernel (with documented API) for program parsing, analysis, and unparsing, which is not specific to AD and which could be used to develop other source-to-source code transformations. Paul Hovland's team and another Argonne team have shown interest for this library.

6.6. Algorithmic Differentiation and Dynamic Memory

Participants: Laurent Hascoët, Sri Hari Krishna Narayanan [Argonne National Lab. (Illinois, USA)].

In the same framework as in section 6.5 , we made progress in the development of the adjoint AD model for programs that use dynamic memory. Adjoint differentiated code obtained by source transformation (OpenAD, Tapenade...) consists of a forward sweep that essentially copies the original code, and a backward sweep that computes the derivatives, These two sweeps must have the same control-flow shape, only reversed. The allocation and deallocation of dynamic memory inside the forward sweep requires a similar pattern in the backward sweep. However, allocations do not always return the same memory chunk, and therefore all memory addresses must be updated to preserve their consistency in the backward sweep.

This problem can only be solved at run-time. A compile-time analysis simply cannot extract the information needed. Our approach is thus to design a library that encapsulates all calls to memory allocation primitives (`malloc, free`...) in order to register the allocated addresses and to restore consistency of pointers during the backward sweep. This strategy is similar to the one we use for MPI calls, *cf* 6.5 , and is actually an ingredient in our AMPI strategy.

This approach was tested with success on a medium-size industrial application in structural mechanics. For this unsteady simulation the C code allocates and frees memory repeatedly at each time step. The tangent and adjoint differentiated C codes, as produced by Tapenade, have been adapted by hand to run the new model, showing promising performance. Obviously, the next step is to update the Tapenade AD model to automate this hand adaptation.

6.7. Algorithmic Differentiation and Iterative Processes

Participants: Laurent Hascoët, Ala Taftaf.

Adjoint codes naturally propagate partial gradients backwards from the result of the simulation. However, this uses the data flow of the simulation in reverse order, at a cost that increases with the length of the simulation. In the special case of iterative Fixed-Point loops, it is clear that the first iterations operate on a meaningless "initial guess" state vector, and that reversing the corresponding data-flow is wasted effort. An adapted adjoint strategy for the iterative process should consider only the last or the few last iterations. Also the adjoint loop, which is itself a Fixed-Point iteration, must have its own stopping criterion and not merely run as many times as the direct Fixed-Point loop. We selected the strategy proposed by Bruce Christianson [17] and this year we implemented it in Tapenade. This strategy is triggered by differentiation directives that we defined. We tested this strategy with success on the medium-size testcase provided by Queen Mary University for the AboutFlow project.

Ala Taftaf presented her results at the WCCM congress during the Eccomas conference in Barcelona [13], july 21-25. Ala Taftaf did a two-months secondment for her Marie Curie PhD grant, with our partner team of Queen Mary University of London, during which she helped them take advantage of the latest developments in Tapenade and of her developments about Fixed-Point adjoints.

6.8. Multi-Activity specialized Differentiation

Participants: Laurent Hascoët, Ian Hueckelheim [Queen Mary University of London].

Up to this year, Tapenade did not allow for specialization of differentiated routines for different "activity" patterns. If a procedure must be differentiated once with respect to some of its arguments, and once with respect to another subset of arguments, then only one generalized differentiated procedure is created, with respect to the union of all subsets of active arguments. This incurs an efficiency penalty, but avoids a combinatorial explosion of the differentiated code.

However, there are cases where the efficiency penalty is high, and some users want more flexibility. Also the specialized adjoint for Fixed-Point iterations *cf* 6.7 uses two distinct activity patterns for the Fixed-Point loop body, and merging them looses some of the benefits of the approach. We have modified Tapenade to perform activity-specialized differentiation, giving the end-user a complete control through differentiation directives.

The experiments on a non-contrived industrial testcase of the AboutFlow project showed a non-negligible improvement between 5 to 10%. Work is still in progress to incorporate this new functionality into the mainstream distributed Tapenade. Ian Hueckelheim presented these results at the 16th EuroAD workshop in Jena, Germany, December 8-9.

MCTAO Project-Team

6. New Results

6.1. Optimal control for quantum systems and NMR

Participants: Bernard Bonnard, Mathieu Claeys [Imperial College, UK], Olivier Cots, Thierry Combot, Pierre Martinon [project team COMMANDS], Alain Jacquemard [Université de Bourgogne, IMB].

- The contrast imaging problem in nuclear magnetic resonance can be modeled as a Mayer problem, in the terminology of optimal control. The candidates as minimizers are selected among a set of extremals, solutions of a Hamiltonian system given by the Pontryagin Maximum Principle; sufficient second order conditions are known; they form the geometric foundations of the HAMPATH code which combines shooting and continuation methods.

  In [4], based on these theoretical studies, a thorough analysis of the case of deoxygenated/oxygenated blood samples is pursued, based on many numerical experiments.

- We initiated more than a year ago a program to compare and study the complementarities between these methods based on the Pontryagin Maximum Principle are known as indirect methods,
  - with the so-called direct methods where optimal control is seem as a generic optimization problem, as implemented in the Bocop software, developed in the COMMANDS project-team,
  - and with LMI techniques used to obtain global bounds on the extremum;
  this was naturally done in collaboration with Pierre Martinon, an important contributor to Bocop and with Mathieu Claeys (LAAS CNRS, a PhD student supervised by J.-B. Lasserre, now with Imperial College). The results are very promising, and there is a gain, numerically, in using both direct and indirect methods while working towards global optimality (in the contrast problem there are many local optima and the global optimality is a complicated issue). This is presented in [3].

  This also led to use algebraic techniques to further analyse the equations and their dependance of the materials to be discriminated [10].

- For time minimal control of a linear spin system with Ising coupling (more complex than the model above), we also analysed *integrability* properties of extremal solutions of the Pontryagin Maximum Principle, in relation with conjugate and cut loci computations. Restricting to the case of three spins, as in [11], the problem is equivalent to analyze a family of almost-Riemannian metrics on the sphere S^2 , with Grushin equatorial singularity. The problem can be lifted into a SR-invariant problem on $SO(3)$, this leads to a complete understanding of the geometry of the problem and to an explicit parametrization of the extremals using an appropriate chart as well as elliptic functions. This approach is compared with the direct analysis of the Liouville metrics on the sphere where the parametrization of the extremals is obtained by computing a Liouville normal form. This is backed by an algebraic approach applying differential Galois theory to integrability.

6.2. Conjugate and cut loci computations and applications

Participants: Bernard Bonnard, Olivier Cots, Jean-Baptiste Caillau, Alessio Figalli [Univ. of Texas at Austin, USA], Thomas Gallouët [MEPHYSTO project-team], Ludovic Rifford.

- Many optimal control problems from mechanics or quantum systems (see [11] and the last paragraph of section 6.1) lead to studying some king of singular metrics, sometimes known as almost-Riemannian. This led us to consider, in [2], metrics on the two-sphere of revolution of the following find: they are Riemannian on each open hemisphere whereas one term of the corresponding tensor becomes infinite on the equator. Length minimizing curves can be computed and structure results on the cut and conjugate loci given, extending those in [25]. These results rely on monotonicity and convexity properties of the quasi-period of the geodesics; such properties are studied on an example with elliptic transcendency. A suitable deformation of the round sphere allows to reinterpretate the equatorial singularity in terms of concentration of curvature and collapsing of the sphere.

- It is known that convexity of the injectivity domain (the boundary of which is sent by the exponential map to the first cut locus) and the "Ma–Trudinger–Wang condition" (an positivity condition on the Ma–Trudinger–Wang tensor) both play a very important role in the continuity of solutions of optimal transport problems. This led to study these properties on their own, and it is still an open question to decide under which conditions the latter implies the former. In [13], it is proved that the MTW condition implies the convexity of injectivity domains on a smooth nonfocal compact Riemannian manifold. This improves a previous result by Loeper and Villani.

6.3. Averaging in control and application to space mechanics

Participants: Bernard Bonnard, Helen-Clare Henninger, Jana Němcová [Instritute of Chemical Tech, Prague, CZ], Jean-Baptiste Pomet, Jeremy Rouot.

As explained in sections 3.5 and 4.1 , control problems where the non controlled system is conservative and the control effect is small compared to the free dynamics lead to computing an average system. This computation may be explicit or numerical.

Even though it will not be always the case that an explicit expression is available, it is interesting to study that case thoroughly.

- In [23], [24], a smooth Riemannian metric was introduced to describe the energy minimizing orbital transfer with low propulsion. We have pursued a study of its deformation due to the standard perturbations in space mechanics, e.g. oblate spheroid shape of the Earth and lunar attraction. In [12], using Hamiltonian formalism, we describe the effects of the perturbations on the orbital transfers and the deformation of the conjugate and cut loci of the original metric. This is done using averaging with respect to both the proper frequency of the space vehicle and the moon frequency.

- The average system has the advantage of being more controllable (it has new virtual controls), but often displays singularities that were not present in the original system. It is the case when minimum time is considered instead of the quadratic energy criterium. We are conducted an analysis of this average minimum time Hamiltonian flow.

  In [6], we compare the two problems for planar transfers. While the energy case leads to analyze a 2D Riemannian metric using the standard tools of Riemannian geometry (curvature computations, geodesic convexity), the time minimal case is associated to a Finsler metric which is not smooth. Nevertheless a qualitative analysis of the geodesic flow is given in this article to describe the optimal transfers. In particular we prove geodesic convexity of the elliptic domain.

6.4. Applications of control methods to dynamical systems

Participants: Gonzalo Contreras, Alessio Figalli, Ayadi Lazrag, Ludovic Rifford, Raffael Ruggiero.

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.

This has been applied to closing geodesics and weak-KAM theory [39], [38].

Ayadi Lazrag's PhD also deals with such problems; applying techniques close to these in [61], he established a version of Francks' lemma for geodesic flows; one goal is to apply this to persistence problems. The approach relies on control theory results, with order 2 conditions. See [14] and [15], where a non trivial conjecture on generic hyperbolicity of the so-called Aubry set of a Hamiltonian is solved on compact surfaces and in the C^2 topology (for genericity).

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

6. New Results

6.1. Electromagnetic wave propagation

6.1.1. *Numerical study of the 1d nonlinear Maxwell equations*

Participants: Loula Fézoui, Stéphane Lanteri.

The system of Maxwell equations describes the evolution of the interaction of an electromagnetic field with a propagation medium. The different properties of the medium, such as isotropy, homogeneity, linearity, among others, are introduced through *constitutive laws* linking fields and inductions. In the present study, we focus on nonlinear effects and address nonlinear Kerr materials specifically. In this model, any dielectric may become nonlinear provided the electric field in the material is strong enough. As a first setp, we consider the one-dimensional case and study the numerical solution of the nonlinear Maxwell equations thanks to DG methods. In particular, we make use of an upwind scheme and limitation techniques because they have a proven ability to capture shocks and other kinds of singularities in the fluid dynamics framework. The numerical results obtained in this preliminary study gives us confidence towards extending this work to higher spatial dimensions.

6.1.2. *High order geometry conforming method for nanophotonics*

Participants: Stéphane Lanteri, Claire Scheid, Jonathan Viquerat.

Usually, unstructured mesh based methods rely on tessellations composed of straight-edged elements mapped linearly from a reference element, on domains which physical boundaries are indifferently straight or curved. Such meshes represent a serious hindrance for high order finite element (FE) methods since they limit the accuracy to second order in the spatial discretization. Thus, exploiting an enhanced representation of physical geometries is in agreement with the natural procedure of high order FE methods, such as the DG method. There are several ways to account for curved geometries. One could choose to incorporate the knowledge coming from CAD in the method to design the geometry and the approximation. These methods are called *isogeometric*, and have received a lot of attention recently. This naturally implies to have access to CAD models of the geometry. On the other hand, *isoparametric* usually rely on a polynomial approximation of both the boundary and the solution. This can be added fairly easily on top of existing implementations. In the present study we focus on the latter type of method, since our goal is first to envisage the benefit of curvilinear meshes for light/matter interaction with nanoscale structures.

6.1.3. *Numerical treatment of non-local dispersion for nanoplasmonics*

Participants: Stéphane Lanteri, Claire Scheid, Nikolai Schmitt, Jonathan Viquerat.

When metallic nanostructures have sub-wavelength sizes and the illuminating frequencies are in the regime of metal's plasma frequency, electron interaction with the exciting fields have to be taken into account. Due to these interactions, plasmonic surface waves can be excited and cause extreme local field enhancements (surface plasmon polariton electromagnetic waves). Exploiting such field enhancements in applications of interest requires a detailed knowledge about the occurring fields which can generally not be obtained analytically. For the numerical modeling of light/matter interaction on the nanoscale, the choice of an appropriate model is a crucial point. Approaches that are adopted in a first instance are based on local (no interaction between electrons) dispersive models e.g. Drude or Drude-Lorentz. From the mathematical point of view, these models lead to an additional ordinary differential equation in time that is coupled to Maxwell's equations. When it comes to very small structures in a regime of 2 nm to 25 nm, non-local effects due to electron collisions have to be taken into account. Non-locality leads to additional, in general non-linear, partial differential equations and is significantly more difficult to treat, though. In this work, we study a DGTD method able to solve the system of Maxwell equations coupled to a linearized non-local dispersion model relevant to nanoplasmonics. While the method is presented in the general 3d case, in this preliminary stdudy, numerical results are given for 2d simulation settings.

6.1.4. *Multiscale DG methods for the time-domain Maxwell equations*

Participants: Stéphane Lanteri, Raphaël Léger, Diego Paredes Concha [LNCC, Petropolis, Brazil], Claire Scheid, Frédéric Valentin [LNCC, Petropolis, Brazil].

Although the DGTD method has already been successfully applied to complex electromagnetic wave propagation problems, its accuracy may seriously deteriorate on coarse meshes when the solution presents multiscale or high contrast features. In other physical contexts, such an issue has led to the concept of multiscale basis functions as a way to overcome such a drawback and allow numerical methods to be accurate on coarse meshes. The present work, which has been initiated in the context of the visit of Frédéric Valentin in the team, is concerned with the study of a particular family of multiscale methods, named Multiscale Hybrid-Mixed (MHM) methods. Initially proposed for fluid flow problems, MHM methods are a consequence of a hybridization procedure which caracterize the unknowns as a direct sum of a coarse (global) solution and the solutions to (local) problems with Neumann boundary conditions driven by the purposely introduced hybrid (dual) variable. As a result, the MHM method becomes a strategy that naturally incorporates multiple scales while providing solutions with high order accuracy for the primal and dual variables. The completely independent local problems are embedded in the upscaling procedure, and computational approximations may be naturally obtained in a parallel computing environment. In this study, a family of MHM methods is proposed for the solution of the time-domain Maxwell equations where the local problems are discretized either with a continuous FE method or a DG method (that can be viewed as a multiscale DGTD method). Preliminary results have been obtained in the 2d case for models problems.

6.1.5. *HDG methods for the time-domain Maxwell equations*

Participants: Alexandra Christophe-Argenvillier, Stéphane Descombes, Stéphane Lanteri.

This study is concerned with the development of accurate and efficient solution strategies for the system of 3d time-domain Maxwell equations coupled to local dispersion models (e.g. Debye, Drude or Drude-Lorentz models) in the presence of locally refined meshes. Such meshes impose a constraint on the allowable time step for explicit time integration schemes that can be very restrictive for the simulation of 3d problems. We consider here the possibility of using an unconditionally stable implicit time integration scheme combined to a HDG discretization method. As a first step, we extend our former study in [20] which was dealing with the 2d time-domain Maxwell equations for non-dispersive media.

6.1.6. *HDG methods for the frequency-domain Maxwell equations*

Participants: Stéphane Lanteri, Liang Li [UESTC, Chengdu, China], Ludovic Moya, Ronan Perrussel [Laplace Laboratory, Toulouse].

In the context of the ANR TECSER project, we continue our efforts towards the development of scalable high order HDG methods for the solution of the system of 3d frequency-domain Maxwell equations. We aim at fully exploiting the flexibiity of the HDG discretization framework with regards to the adaptation of the interpolation order (p-adaptivity) and the mesh (h-adaptivity). In particular, we study the formulation of HDG methods on a locally refined non-conforming tetrahedral mesh and on a non-confoming hybrid cubic/tetrahedral mesh. We also investigate the coupling between the HDG formulation and a BEM (Boundary Element Method) discretization of an integral representation of the electromagnetic field in the case of propagation problems theoretically defined in unbounded domains.

6.2. Elastodynamic wave propagation

6.2.1. *Sesimic wave interaction with viscoelastic media*

Participants: Nathalie Glinsky, Stéphane Lanteri, Fabien Peyrusse [Department of Mathematics, Purdue University].

This work is concerned with the development of high order DGTD methods formulated on unstructured simplicial meshes for the numerical solution of the system of time-domain elastodynamic equations. These methods share some ingredients of the DGTD 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. A recent novel contribution is the numerical treatment of 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 3d.

6.2.2. *DG 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.

6.2.3. *HDG method for the frequency-domain elastodynamic equations*
Participants: Hélène Barucq [MAGIQUE-3D project-team, Inria Bordeaux - Sud-Ouest], Marie Bonnasse-Gahot, Julien Diaz [MAGIQUE-3D project-team, Inria Bordeaux - Sud-Ouest], Stéphane Lanteri.

One of the most used seismic imaging methods is the full waveform inversion (FWI) method which is an iterative procedure whose algorithm is the following. Starting from an initial velocity model, (1) compute the solution of the wave equation for the N sources of the seismic acquisition campaign, (2) evaluate, for each source, a residual defined as the difference between the wavefields recorded at receivers on the top of the subsurface during the acquisition campaign and the numerical wavefields, (3) compute the solution of the wave equation using the residuals as sources, and (4) update the velocity model by cross correlation of images produced at steps (1) and (3). Steps (1)-(4) are repeated until convergence of the velocity model is achieved. We then have to solve $2N$ wave equations at each iteration. The number of sources, N, is usually large (about 1000) and the efficiency of the inverse solver is thus directly related to the efficiency of the numerical method used to solve the wave equation. Seismic imaging can be performed in the time-domain or in the frequency-domain regime. In this work which is conducted in the framework of the Depth Imaging Partnership (DIP) between Inria and TOTAL, we adopt the second setting. The main difficulty with frequency-domain inversion lies in the solution of large sparse linear systems which is a challenging task for realistic 3d elastic media, even with the progress of high performance computing. In this context, we study novel high order HDG methods formulated on unstructured meshes for the solution of the frency-domain elastodynamic equations. Instead of solving a linear system involving the degrees of freedom of all volumic cells of the mesh, the principle of a HDG formulation is to introduce a new unknown in the form of Lagrange multiplier representing the trace of the numerical solution on each face of the mesh. As a result, a HDG formulation yields a global linear system in terms of the new (surfacic) unknown while the volumic solution is recovered thanks to a local computation on each element.

6.2.4. Multiscale DG methods for the time-domain elastodynamic equations

Participants: Marie-Hélène Lallemand Tenkès, Frédéric Valentin [LNCC, Petropolis, Brazil].

In the context of the visit of Frédéric Valentin in the team, we have initiated a study aiming at the design of novel multiscale methods for the solution of the time-domain elastodynamic equations, in the spirit of MHM (Multiscale Hybrid-Mixed) methods previously proposed for fluid flow problems. Motivation in that direction naturally came when dealing with non homogeneous anisotropic elastic media as those encountered in geodynamics related applications, since multiple scales are naturally present when high contrast elasticity parameters define the propagation medium. Instead of solving the usual system expressed in terms of displacement or displacement velocity, and stress tensor variables, a hybrid mixed-form is derived in which an additional variable, the Lagrange multiplier, is sought as representing the (opposite) of the surface tension defined at each face of the elements of a given discretization mesh. We consider the velocity/stress formulation of the elastodynamic equations, and study a MHM method defined for a heterogeneous medium where each elastic material is considered as isotropic to begin with. If the source term (the applied given force on the medium) is time independent, and if we are given a arbitrarily coarse conforming mesh (triangulation in 2d, tetrahedrization in 3d), the proposed MHM method consists in first solving a series of fully decoupled (therefore parallelizable) local (element-wise) problems defining parts of the full solution variables which are directly related to the source term, followed by the solution of a global (coarse) problem, which yields the degrees of freedom of both the Lagrange multiplier dependent part of the full solution variables and the Lagrange multiplier itself. Finally, the updating of the full solution variables is obtained by adding each splitted solution variables, before going on the next time step of a leap-frog time integration scheme. Theoretical analysis and implementation of this MHM method where the local problems are discretized with a DG method, are underway.

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

6. New Results

6.1. Highlights of the Year

Paola Goatin was awarded the *"Prix Inria - Académie des sciences du jeune chercheur"*.

6.2. Mathematical analysis and control of macroscopic traffic flow models

6.2.1. *Vehicular traffic*

Participants: Enrico Bertino, Guillaume Costeseque, Maria Laura Delle Monache, Paola Goatin, Sheila Scialanga, Alexandre Bayen [UC Berkeley, CA, USA], Sebastien Blandin [IBM Research Collaboratory, Singapore], Christophe Chalons [LJLL, UP7].

In collaboration with UC Berkeley, and as part of the Associated Team ORESTE activity (see http://www-sop.inria.fr/members/Paola.Goatin/ORESTE/index.html), we have considered the System Optimal Dynamic Traffic Assignment problem with Partial Control (SO-DTA-PC) for general road networks with horizontal queuing. The goal of which is to optimally control any subset of the networks agents to minimize the total congestion of all agents in the network. We adopt a flow dynamics model that is a Godunov discretization of the Lighthill-Williams-Richards (LWR) partial differential equation with a triangular flux function and a corresponding multi-commodity junction solver. Full Lagrangian paths are assumed to be known for the controllable agents, while we only assume knowledge of the aggregate turn ratios for the non-controllable (selfish) agents. We solve the resulting finite horizon non-linear optimal control problem using the discrete adjoint method, see [75].

As part of our TRAM3 activity and in collaboration with C. Chalons (UVSQ), we designed a new finite volume conservative algorithm to track the trajectory of a bus in the surrounding traffic using a locally non-uniform moving mesh, see [70].

In collaboration with S. Blandin (IBM), we proved the existence and stability of entropy weak solutions of a scalar conservation law with non-local flux arising in traffic flow modeling. The result is obtained providing accurate L^∞, BV and L^1 estimates for the sequence of approximate solutions constructed by an adapted Lax-Friedrichs scheme.

In collaboration with the University of Mannheim and in the framework of the PHC Procope project "Transport Networks Modeling and Analysis", we studied how to manage variable speed limits combined with coordinated ramp metering within the framework of the LWR network model. Following a "first discretize then optimize" approach, we derived the first order optimality system and explained the switch of speeds at certain fixed points in time and the boundary control for the ramp metering as well. Sequential quadratic programming methods are used to solve the control problem numerically. For application purposes, we present experimental setups where variable speed limits are used as a traffic guidance system to avoid traffic jams on highway interchanges and on-ramps, see [71].

Finally, E. Bertino internship was devoted to uncertainty quantification in macroscopic traffic flow models.

6.2.2. *Crowd motion*

Participants: Aekta Aggarwal, Régis Duvigneau, Paola Goatin, Matthias Mimault, Rinaldo M. Colombo [Brescia University, Italy].

A. Aggarwal postdoc is devoted to the analytical and numerical study of systems of conservation laws with non-local fluxes in several space dimensions. In collaboration with R.M. Colombo, we presented a Lax-Friedrichs type algorithm to numerically integrate this class of systems. The convergence of the approximate solutions was proved, also providing the existence of solution in a slightly more general setting than in other results in the current literature. An application to a crowd dynamics model is considered. This numerical algorithm is then used to test the conjecture that as the convolution kernels converge to a Dirac δ, the nonlocal problem converges to its non-nonlocal analogue.

M. Mimault is working on scalar conservation laws with non-local flow in two space dimensions. These equations are meant to model crowd motion, where the movement direction of each pedestrian depends on a weighted mean of the crowd density around him. In particular, M. Mimault is implementing a finite volume numerical scheme which will be used for flow optimization purposes.

6.3. 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 disciplines, such as structural mechanics.

Our approach to *competitive optimization* is focused on the two-discipline problem. It 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 [65] (see also subsectionsubsect:helico).

Our approach to *cooperative optimization*, in theory, is not limited in number of objective functions. It is based on a result of convex analysis established for a general unconstrained multiobjective problem in which all the gradients are assumed to be known. The theorem [66] 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-stationary. 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 Pareto-stationary point, and in practice actual Pareto-optimality is verified a posteriori.

The two approaches have been combined to explore the Pareto front segment-wise as illustrated on Figure 2 .

6.3.1. *Multiple-Gradient Descent Algorithm (MGDA)*
Participants: Jean-Antoine Désidéri, Régis Duvigneau, Camilla Fiorini, Matteo Giacomini, Abderrahmane Habbal, Adrien Zerbinati.
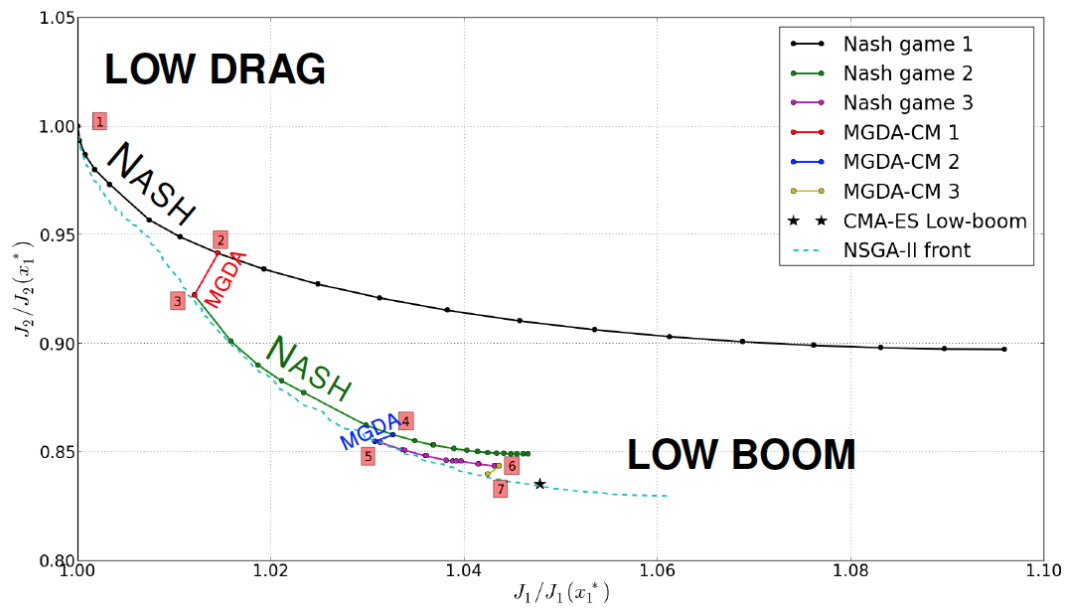
*Figure 2. **Two-discipline optimization of a generic geometry of a supersonic aircraft, for concurrent drag and sonic-boom reduction** (from A. Minelli's doctoral thesis, 2013). The wave drag is calculated by the ONERA elsA code in 3D finite-volume Eulerian flow mode over a 6M-node mesh and the sonic boom using a three-layer approach. The Nash-game paths have been devised by appropriate territory splitting in order to be tangent to the Pareto front, and they are interrupted whenever the Pareto-stationarity condition is judged excessively violated. The MGDA paths converge rapidly back to the front. The simulation demonstrates how the two algorithms complement each other and provide a potential for a piecewise description of the Pareto front, evaluated more economically than a stochastic algorithm operating on a large population.*

6.3.1.1. 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 [66].

The notion of Pareto-stationarity, originally established to be a necessary condition of optimality in differentiable multi-objective optimization of unconstrained problems, has been extended to problems subject to equality constraints. On this basis, we were able to establish that by augmenting, in a classical manner, the objective-functions of a penalty term equal to the square of the constraint violation, and applying the standard MGDA to it, would result in converged solutions that are Pareto-stationary in the extended sense. Numerical experimentation on this is on-going.

6.3.1.2. 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 several aerodynamic shape optimization problems : lift concurrently with drag optimization for transonic aircraft; drag (under lift constraint) concurrently with sonic boom reduction for 3D supersonic configuration (at ONERA); drag (under lift constraint) concurrently with mass reduction for transport aircraft (at ONERA) [56].

6.3.1.3. 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 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 [68].

6.3.1.4. Optimization of an unsteady system using a multiobjective formulation

An approach has been developed to solve optimization problems in which the functional that has to be minimized is time dependent. In the literature, the most common approach when dealing with unsteady problems, is to consider a time-average criterion. However, this approach is limited since the dynamical nature of the state is neglected. Our alternative consists in building a set of cost functionals by evaluating a single criterion at different sampling times. In this way, the optimization of the unsteady system is formulated as a multi-objective optimization problem, solved using an appropriate descent algorithm (MGDA). Moreover, we also consider a hybrid approach in which the set of cost functionals is built by doing a time-average operation over multiple intervals. These strategies have been illustrated and applied to a non-linear unsteady system governed by a one-dimensional convection-diffusion-reaction partial differential equation [67].

6.3.1.5. 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 multipoint problem by time-discretization of the time and parameter-dependent functional (as above); etc.

6.3.2. *Flow control*

Participants: Régis Duvigneau, Jérémie Labroquère, Emmanuel Guilmineau [Ecole Centrale de Nantes].

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, we develop a methodology to determine optimal control parameters by coupling the simulation of unsteady actuated flows 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 and apply an unsteady gradient-based approach[67] ;
- the optimization of control parameters using a statistical metamodel-based strategy [39].

In this perspective, unsteady Reynolds Averaged Navier-Stokes equations are solved, with some turbulence closures. Different models for synthetic jet have been implemented to simulate the actuation, and then validated for different turbulence closures .

Specific developments have be carried out in the metamodel-based optimizer to include a noise term into Gaussian Process model, which is used to filter errors arising from unsteady simulations. A systematic assessment of modeling and numerical errors has been archived for a backward facing step test-case, with the objective of controlling the re-attachment point location[46], [58].

This activity is conducted in collaboration with the CFD team of Ecole Centrale de Nantes.

6.3.3. *Adjoint-based mesh quality control*

Participants: Jean-Antoine Desideri, Maxime Nguyen-Dinh [ONERA doctoral student], Jacques Peter [Research Engineer, ONERA/DSNA], Renaud Sauvage [Airbus France], Mathieu Meaux [EADS IW].

In his doctoral thesis [29], Nguyen Dinh has studied mesh adaptation methods based on the total derivatives of aerodynamic outputs with respect to mesh coordinates by the discrete adjoint method. Firstly, mesh adaptation methods have been devised for Eulerian flows. Zones to be refined are detected using a sensor based on the total derivative, and numerical experiments confirmed the adequacy of the approach. Secondly, the method was extended to the Reynolds-averaged Navier equations (RANS) and thirdly demonstrated for 3D industrial configurations [53].

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

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

E. Roca Leon has conducted a CIFRE thesis at ONERA DAAP supported by Airbus Helicopter (Marignane) [34]. 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. Then, actual two-objective optimizations are conducted via our Nash game approach to competitive optimization with territory splitting based on reduced Hessian diagonalization. Successful optimization has been realized involving 16 geometrical parameters to reduce the power in forward motion while maintaining sub-optimality of the drag in hover [55] [64] [65].

6.4. Isogeometric analysis and design

Participants: Régis Duvigneau, Asma Gdhami, Bernard Mourrain [Galaad Project-Team], Bernd Simeon [Tech. Univ. of Kaiserslautern].

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 geometric modeling, computational domain description and solution basis functions. 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, a collaborative work has also been carried out with the Technical University of Kaiserslautern, concerning the computation of shape gradients for linear elasticity problems[42], [68]. Moreover, the doctoral thesis of Asma Gdhami, in collaboration with ENIT in Tunisia, has started and concerns the development of isogeometric schemes for hyperbolic systems.

6.5. Optimum design in structural mechanics

6.5.1. *Shape Optimization in Multidisciplinary Non-Linear Mechanics*

Participants: Aalae Benki, Jean-Antoine Désidéri, Abderrahmane Habbal, Gael Mathis [ArcelorMittal, CRAA].

In collaboration with the ArcelorMittal's Center for Research in Automotive and Applications (CRAA), 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.

Then, we considered the 3D problem of an automotive twist beam. In this 3D case, we aim to Pareto-optimal shapes for two objectives, the first being to minimize the Von-Mises strain to guarantee the formability of the twist beam, and the second being to maximize the stiffness. For solution with higher stiffness than the initial one, we could decrease the thickness to obtain a mass reduction with the same end-user properties.

We also introduced, to our knowledge for the first time in the structural optimization area, the notion of Kalai-Smorodinky equilibria which is aimed at the selection of equilibria among Pareto-optimal solutions. We applied this notion of equilibria to both industrial cases, and compared the results to Nash equilibria.

6.5.2. *Optimization of Addendum Surfaces in Stamping*

Participants: Fatima Zahra Oujebbour, Rachid Ellaia, Abderrahmane Habbal, Ziheng Zhao.

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 process parameters (two then four parameters). Then, we considered the more complex shape design of the initial blank. The initial blank design is a critical step in stamping design procedure, therefore it should be optimally designed. Our aim is to find the optimal initial blank shape that avoids or at least minimizes the springback and failure flaws. For this study, the geometry of the blank contour is described by parametric spline curves. Seven control points (P1,...,P7) are used to define the spline curves in order to have a wide variety of geometries. The exact computational evaluation of our criteria, springback and failure, is very expensive (the FE model request around 45 min to predict these two criteria) and the design space is of quite high dimension. Therefore, we considered the recourse to the sparse grid interpolation. Optimization process based on sparse grid interpolation is an optimal alternative in which criteria can be approximated with a suitable interpolation formula that needs significantly less points than the full grid. the obtained metamodel using sparse grid interpolation needs less than 1s to predict springback and failure on the same computation machine. To find the optimal initial blank shape, it was decided to perform the optimization process using the obtained metamodel. The construction of the sparse grid interpolant was based on the Chebyshev Gauss-Lobatto grid type and using the polynomial basis functions. This technique achieves a good accuracy with a competitive number of grid points. The comparison of the obtained fronts shows that we can capture Pareto solutions by NBI and NNCM with fewer points than NSGAII which requires a large number of populations and several generations to obtain the Pareto front. [48] [49] [50]

6.6. Application of shape and topology design to biology and medicine

6.6.1. *Assessing the ability of the 2D Fisher-KPP equation to model cell-sheet wound closure*

Participants: Abderrahmane Habbal, Hélène Barelli [Univ. Nice Sophia Antipolis, CNRS, IPMC], Grégoire Malandain [Inria, EPI Morpheme].

We address in this joint collaboration the ability of the widely used Fisher-KPP equations to render some of the dynamical features of epithelial cell-sheets during wound closure.

Our approach is based on nonlinear parameter identification, in a two-dimensional setting, and using advanced 2D image processing of the video acquired sequences. As original contribution, we lead a detailed study of the profiles of the classically used cost functions, and we address the "wound constant speed" assumption, showing that it should be handled with care.

We study five MDCK cell monolayer assays in a reference, activated and inhibited migration conditions. Modulo the inherent variability of biological assays, we show that in the assay where migration is not exogeneously activated or inhibited, the wound velocity is constant. The Fisher-KPP equation is able to accurately predict, until the final closure of the wound, the evolution of the wound area, the mean velocity of the cell front, and the time at which the closure occurred. We also show that for activated as well as for inhibited migration assays, many of the cell-sheet dynamics cannot be well captured by the Fisher-KPP model. Original unexplored utilizations of the model such as wound assays classification based on the calibrated diffusion and proliferation rate parameters is ongoing. [47]

6.7. Distributed Systems

6.7.1. *High-Performance manipulation and storage of e-Science data*

Participants: Benoit Lange, Toan Nguyen.

The work carried in previous years on distributed High-Performance Computing for e-Science workflows has enlightened the need for appropriate tools and methods to manage petabyte and exabyte volumes of data. This has been the focus of the work carried by Benoit Lange during his Post-Doc position in 2014. It was dedicated to the definition and prototyping of a large-scale HPC platform to support the execution of application solvers, efficient storage and management of large-volumes of data produced by the simulation applications and the visualization of their results on high-end graphics workstations. This platform also includes analytics software to produce specific results corresponding to the user queries. This is based on the Hadoop ecosystem [59]. Is is central for the communication between the dedicated HPC nodes running the solvers and the visualization nodes interfacing the end-users. It includes high-speed storage with dedicated file systems on specific nodes, and long-term storage for reference data using magnetic juke-boxes that store petabytes of application data. This work is supported by an FP7 project in which Inria is responsible for the Data Management work-package (Call FP7-2013-ICT-11, Grant 619439, 2014-2016). The partners of the project, named VELaSSCo (Visualization for Extremely Large Scale Scientific Computing), are : CIMNE (SP, coordinator), JOTNE and SINTEF (No), ATOS (SP), Fraunhofer IGD (D) and the University of Edinburg (UK).

<div align="center" style="color:red">TOSCA Project-Team</div>

6. New Results

6.1. Probabilistic numerical methods, stochastic modelling and applications

Participants: Mireille Bossy, Nicolas Champagnat, Julien Claisse, Madalina Deaconu, Benoît Henry, James Inglis, Antoine Lejay, Oana Valeria Lupascu, Sylvain Maire, Sebastian Niklitschek Soto, Denis Talay, Etienne Tanré, Denis Villemonais.

6.1.1. Published works and preprints

- M. Bossy and J.-F. Jabir (University of Valparaíso) [13] have proved the well-posedness of a conditional McKean Lagrangian stochastic model, endowed with the specular boundary condition, and further the mean no-permeability condition, in a smooth bounded confinement domain $\mathcal{D}$.

- M. Bossy, N. Champagnat, S. Maire and L. Violeau worked with H. Leman (CMAP, Ecole Polytechnique) and M. Yvinec (Inria Sophia, EPI GEOMETRICA) on Monte Carlo methods for the linear and non-linear Poisson-Boltzmann equations [12]. These methods are based on walk on spheres algorithm, simulation of diffusion processes driven by their local time, and branching Brownian motion. Their code for the linear equation can deal with bio-molecules of arbitrary sizes, based on computational geometry tools from the CGAL C++ Library developed by the GEOMETRICA team. The non-linear equation is solved using branching Brownian motion.

- M. Bossy, O. Faugeras (Inria Sophia, EPI NEUROMATHCOMP), and D. Talay have clarified the well-posedness of the limit equations to the mean-field N-neuron models proposed in [42] and proven the associated propagation of chaos property. They also have completed the modeling issue in [42] by discussing the well-posedness of the stochastic differential equations which govern the behavior of the ion channels and the amount of available neurotransmitters. See [29].

- N. Champagnat and D. Villemonais obtained criterions for existence and uniqueness of quasi-stationary distributions and Q-processes for general absorbed Markov processes [31]. A quasi-stationary distribution is a stationary distribution conditionally on non-absorbtion, and the Q-process is defined as the original Markov process conditioned to never be absorbed. The criterion that they obtain ensures exponential convergence of the conditioned t-marginal of the process conditioned not to be absorbed at time t, to the quasi-stationary distribution and also the exponential ergodicity of the Q-process.

- M. Deaconu and S. Herrmann continued and completed the study of the simulation of the hitting time of some given boundary for Bessel processes. They constructed an original approximation method for hitting times of a given threshold by Bessel processes with non-integer dimension. In this work, they combine the additivity property of the laws of squared Bessel processes with their previous results on the simulation of hitting times of Bessel processes with integer dimension, based on the method of images and on the connexion with the Euclidiean norm of the Brownian motion [33].

- M. Deaconu, S. Herrmann and S. Maire introduced a new method for the simulation of the exit time and position of a δ-dimensional Brownian motion from a domain. The main interest of this method is that it avoids splitting time schemes as well as inversion of complicated series. The idea is to use the connexion between the δ-dimensional Bessel process and the δ-dimensional Brownian motion thanks to an explicit Bessel hitting time distribution associated with a particular curved boundary. This allows to build a fast and accurate numerical scheme for approximating the hitting time [34].

- M. Deaconu and O. Lupaşcu worked with L. Beznea (Bucharest, Romania) on the construction and the branching properties of the solution of the fragmentation equation and properly associate a continuous time càdlàg Markov process. The construction and the proof of the path regularity of the Markov processes are based on several newly developed potential theoretic tools.

- J. Inglis, together with O. Faugeras (Inria NEUROMATHCOMP) finalized their article [18] on the well-posedness of stochastic neural field equations within a rigorous framework.

- J. Inglis and E. Tanré together with F. Delarue and S. Rubenthaler (Univ. Nice – Sophia Antipolis) finalized their article [16] on the global solvability of a networked system of integrate-and-fire neurons proposed in the neuroscience literature.

- J. Inglis and E. Tanré together with F. Delarue and S. Rubenthaler (Univ. Nice – Sophia Antipolis) completed their study of the mean-field convergence of a highly discontinuous particle system modeling the behavior of a spiking network of neurons, based on the integrate-and-fire model [17]. Due to the highly singular nature of the system, it was convenient to work with a relatively unknown Skorohod topology.

- J. Inglis and D. Talay introduced in [38] a new model for a network of spiking neurons that attempted to address several criticisms of previously considered models. In particular the new model takes into account the role of the dendrites, and moreover includes non-homogeneous synaptic weights to describe the fact that not all neurons have the same effect on the others in the network. They were able to obtain mean-field convergence results, using new probabilistic arguments.

- A. Lejay have worked with G. Pichot (EPI SAGE) on benchmarks for testing Monte Carlo methods to simulate particles in one-dimensional media, and applied this statistical methodology to four methods, including the exact method developed previously [45]. This work led also to empirical observations that should guide the design of new methods [24].

- S. Maire is working with the Bulgarian Academy of sciences on Monte Carlo algorithms for linear equations based on killed random walks. In a first work, with I. Dimov and J-M. Sellier [37], a new Monte Carlo method to solve linear systems of equations has been introduced. This method can either compute one component of the solution or all components simultaneously. In a second work, with Ivan Dimov and Rayna Georgieva, a new Monte Carlo method to solve Fredhom integral equations of the second kind is developed [36].

- D. Villemonais worked with P. Del Moral (Univ. Sydney) on the conditional ergodicity of time inhomogeneous diffusion processes [35]. They proved that, conditionally on non extinction, an elliptic time-inhomogeneous diffusion process forgets its initial distribution exponentially fast. An interacting particle scheme to numerically approximate the conditional distribution is also provided.

- D. Villemonais proved a Foster-Lyapunov type criterion which ensures the exponential ergodicity of a Fleming-Viot type particle system whose particles evolve as birth and death processes. The criterion also ensures the tightness of the sequence of empirical stationary distributions considered as a family of random measures. A numerical study of the speed of convergence of the particle system is also obtained under various settings [41].

6.1.2. Other works in progress

- M. Bossy and J-F. Jabir (University of Valparaíso) proved the validity of a particle approximation of a (simplified) Lagrangian Stochastic Model submitted to specular reflections at the boundary and satisfying the mean no-permeability condition. This work achieves to extend our previous study [43] to the multidimensional case.

- N. Champagnat and D. Villemonais obtained criterions for existence, uniqueness and exponential convergence in total variation of quasi-stationary distributions and Q-processes for general absorbed and killed diffusion processes. The criterion obtained is equivalent to the property that a diffusion on natural scale coming down from infinity has uniformly (w.r.t. the initial condition) bounded expectation at a fixed time t. A study of nearly critical cases allow to conjecture that this property is true for all diffusion processes on natural scale coming down from infinity. This work is currently being written.

- N. Champagnat and B. Henry worked on the long-time behavior of the frequency spectrum for the Splitting Tree models under the infinitly-many alleles model. They obtained, using a new method for computing the expectation of an integral with respect to a random measure, the asymptotic behavior

of the moments of the frequency spectrum. As an application, they derived the law of large number and a new central limit theorem for the frequency spectrum. This work is currently being written.

- J. Claisse defended his PhD. under the supervision of N. Champagnat and D. Talay on stochastic control of population dynamics. He completed a finite-horizon optimal control problem on branching–diffusion processes. He also created and studied a hybrid model of tumor growth emphasizing the role of acidity. Key therapeutic targets appear in the model to allow investigation of optimal treatment problems.

- J. Claisse and D. Talay in collaboration with X. Tan (Univ. of Paris Dauphine) extended their previous work on a pseudo-Markov property enjoyed by the solutions of controlled stochastic differential equations and its application to the proof of the dynamic programming principle. A paper is being finished.

- M. Deaconu and O. Lupascu are working with L. Beznea (Bucharest, Romania) on a stochastic model for avalanche phenomena involving rupture properties that occur in the physical and deterministic models for snow avalanches. This approach is based on their recent results on fragmentation processes by stochastic differential equation and branching processes.

- M. Deaconu and O. Lupascu are working on a numerical probabilistic algorithm for an avalanche-type process. The originality of this approach is to use a coagulation/fragmentation model to describe the avalanche phenomenon. More precisely, they consider a particular fragmentation kernel which introduces "rupture-type" properties of deterministic models for snow avalanches.

- An important issue in neuroscience is the modelling of spike trains of a single neuron. In this context, the membrane potential of a neuron can be described by using a simple stochastic differential equation with periodic input, that is reset to a rest potential each time it hits a certain threshold. J. Inglis, A. Richard, D. Talay, and E. Tanré study how the law of these hitting times is affected when one changes the white noise (in the SDE) into a correlated noise. Practically, they use a fractional Brownian motion, and since the computation of the hitting times of such a non-Markovian, non-semimartingale process is still an open question, they rather try to compute the deviations from the white noise model. This is expected to give insights on the relevance of models with memory and long-range dependence.

- J. Inglis started a collaboration with B. Hambly and S. Ledger at the University of Oxford, in which interacting mean-field particle systems with common noise are being studied. Such systems are representative of systems of spiking neurons or portfolio defaults. In previous studies each particle was driven by a noise that was assumed independent from particle to particle (i.e. intrinsic noise). By considering a common driving noise in addition to the intrinsic noise, it is possible to model the fact that the environment in which the particles live is also noisy. This leads to the study of a new type of conditional McKean-Vlasov equation.

- J. Inglis, in collaboration with J. Maclaurin (EPI NEUROMATHCOMP) and W. Stannat (Berlin), has begun working on a new framework to understand the effect of noise on neural field equations. Deterministic neural field equations exhibit traveling wave solutions, and so the effect of noise on these solutions is of great interest. The idea is to decompose the solution into various components, which allow one to see directly how the noise affects the solution in the direction of the moving wave front. In particular, the goal is to reconcile mathematically the previous works of P. Bressloff and W. Stannat on the same subject, and to obtain a large deviation principle.

- J. Inglis and D. Talay are in the process of studying the emergence of spatio-temporal noise starting from microscopic models of neuron conductance.

- A. Lejay continued his collaboration with S. Torres (Universidad de Valapraíso, Chile) and E. Mordecki (Universidad de la República, Uruguay) on the estimation of the parameter of the Skew Brownian motion. This work is related to the modelling of diffusion processes in media with interfaces and has potential applications in many domains, such as population ecology.

- Together with R. Rebolledo (Pontificia Universidad Católica, Santiago, Chile), A. Lejay continued his review work on the mathematical modelling of the Wave Energy Converter Called the Oscillating water column, within the framework of the CIRIC project.

- A. Lejay continued his work on the Snapping out Brownian motion to perform numerical tests for the computation of the mean residence time in a diffusive medium with semi-permeable membranes, such as the one encountered in the mathematical modelling of diffusion Magnetic Resonance Imaging.

- A. Lejay continued his collaboration with L. Coutin (Universté Paul Sabatier, Toulouse) on the sensitivity of rough linear differential equations, by providing general results on the derivatives of the solution of rough differential equations with respect to parameters or the starting point.

- S. Niklitschek Soto and D. Talay completed their stochastic analysis of diffraction parabolic PDEs with general discontinuous coefficients in the multidimensional case.

- P. Guiraud (University of Valparaíso) and E. Tanré study the effect of noise in the phenomenon of spontaneous synchronisation in a network of connected leaky 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 the noiseless model does not show synchronization. (Math Amsud program SIN)

- O. Faugeras (EPI NEUROMATHCOMP) and E. Tanré worked on an extension of [44] to a context of several populations of homogeneous neurons. They study the limit mean field equation of the membrane potential as the number of neurons increases in a network with correlated synaptic weights. A paper is in preparation.

- C. Graham (CMAP, Ecole polytechnique) and D. Talay are writing the second volume of their series published by Springer on the Mathematical Foundations of Stochastic Simulations.

- In collaboration with N. Touzi (CMAP, Ecole polytechnique), D. Talay is studying stochastic differential equations involving local times with stochastic weights, and extensions of classical notions of viscosity solutions to PDEs whose differential operator has discontinuous coefficients and transmission boundary conditions.

6.2. Financial Mathematics

Participants: Mireille Bossy, Nicolas Champagnat, Madalina Deaconu, Antoine Lejay, Khaled Salhi, Denis Talay, Etienne Tanré.

6.2.1. Published works and preprints

- In collaboration with N. Maïzi (CMA - Mines Paristech) and O. Pourtallier (COPRIN team, Inria Sophia Antipolis - Méditerranée), M. Bossy studied the existence of a Nash equilibrium between electricity producers selling their production on an electricity market and buying CO_2 emission allowances on an auction carbon market. The producers' strategies integrate the coupling of the two markets via the cost functions of the electricity production. The authors set out the set of Nash equilibria on the electricity market, that constitutes an equivalence class (same prices and market shares) from which they exhibit a dominant strategy. On the coupled markets, given a specific carbon market design (in terms of penalty level and allowances), they compute the bounds of the interval where carbon prices (derived from the previous dominant strategy) evolve. They specify the properties of the associated equilibria (see [30] and [14]).

- In their article [40], N. Champagnat, M. Deaconu, A. Lejay and K. Salhi have constructed a regime switching model for estimating the Value-at-Risk. This model classifies the states in crisis and steady regimes and constructs a mixture of power laws as a model for returns of financial assets.

- In collaboration with V. Reutenauer and C. Michel (CA-CIB), D. Talay and E. Tanré worked on a model in financial mathematics including bid-ask spread cost. They study the optimal strategy to hedge an interest rate swap that pays a fixed rate against a floating rate. They present a methodology using a stochastic gradient algorithm to optimize strategies. A paper has been submitted [39].

6.2.2. Other works in progress

- In collaboration with J. Bion-Nadal (Ecole Polytechnique and CNRS), D. Talay pursued the study of a new calibration methodology based on dynamical risk measures and stochastic control PDEs.

<div align="center">

ABS Project-Team

</div>

5. New Results

5.1. Highlights of the Year

In 2014, two achievements are worth noticing:

Analysis of large assemblies using native mass spectrometry data. Native mass spectrometry is about to revolutionize structural biology, since such experiments give access to the composition in terms of subunits of large macro-molecular assemblies, usually beyond reach for classical experimental techniques. In this context, we designed an algorithm to infer pairwise contacts within subunits of large macro-molecular assemblies – see section 5.3.1 . To the best of our knowledge, our algorithm is the only one whose performances can be precisely analyzed, the contenders being of heuristic nature.

Analysis and comparison of conformational ensembles and sampled energy landscapes. A key property governing the behavior of many biophysical systems is the classical enthalpy - entropy balance, which is the root of thermodynamics. Therefore, studying the way a protein folds or the way two proteins assemble requires unveiling properties of ensembles of conformations of the system scrutinized. In this context, we designed novel methods to analyze and compare collections of conformations and the associated energy landscape – see section 5.4.1 . The algorithms are based on state-of-the-art techniques from computational topology (Morse theory, Morse homology), and optimal transportation.

5.2. Modeling Interfaces and Contacts

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

The work undertaken in this vein in 2014 will be finalized in 2015.

5.3. Modeling Macro-molecular Assemblies

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

5.3.1. *Connectivity Inference in Mass Spectrometry based Structure Determination*

Participants: Frédéric Cazals, Deepesh Agarwal.

In collaboration with C. Caillouet, and D. Coudert, from the COATI project-team (Inria - I3S (CNRS, University of Nice Sophia Antipolis)).

Consider a set of oligomers listing the subunits involved in sub-complexes of a macro-molecular assembly, obtained e.g. using native mass spectrometry or affinity purification. Given these oligomers, connectivity inference (CI) consists of finding the most plausible contacts between these subunits, and minimum connectivity inference (MCI) is the variant consisting of finding a set of contacts of smallest cardinality. MCI problems avoid speculating on the total number of contacts, but yield a subset of all contacts and do not allow exploiting a priori information on the likelihood of individual contacts.

In this paper [14], we present two novel algorithms, MILP-W and MILP-WB. The former solves the minimum weight connectivity inference (MWCI), an optimization problem whose criterion mixes the number of contacts and their likelihood. The latter uses the former in a bootstrap fashion, to improve the sensitivity and the specificity of solution sets.

Experiments on three systems (yeast exosome, yeast proteasome lid, human eiF3), for which reference contacts are known (crystal structure, cryo electron microscopy, cross-linking), show that our algorithms predict contacts with high specificity and sensitivity, yielding a very significant improvement over previous work, typically a twofold increase in sensitivity.

The software accompanying this paper is made available, and should prove of ubiquitous interest whenever connectivity inference from oligomers is faced.

5.4. Modeling the Flexibility of Macro-molecules

Protein, flexibility, collective coordinate, conformational sampling dimensionality reduction.

5.4.1. *Conformational Ensembles and Sampled Energy Landscapes: Analysis and Comparison*
Participants: Frédéric Cazals, Tom Dreyfus, Christine Roth.

In collaboration with D. Mazauric (Inria Sophia Antipolis Méditerranée, Geometrica) and C. Robert (IBPC / CNRS, Paris).

In this work, we present novel algorithms and software addressing four core problems in computational structural biology, namely analyzing a conformational ensemble, comparing two conformational ensembles, analyzing a sampled energy landscape, and comparing two sampled energy landscapes [15]. Using recent developments in computational topology, graph theory, and combinatorial optimization, we make two notable contributions. First, we present a generic algorithm analyzing height fields. We then use this algorithm to perform density based clustering of conformations, and to analyze a sampled energy landscape in terms of basins and transitions between them. In both cases, topological persistence is used to manage (geometric) frustration. Second, we introduce two algorithms to compare transition graphs. The first is the classical *earth mover distance* metric which depends only on local minimum energy configurations along with their statistical weights, while the second incorporates topological constraints inherent to conformational transitions.

Illustrations are provided on a simplified protein model (BLN69), whose frustrated potential energy landscape has been thoroughly studied.

The software implementing our tools is also made available, and should prove valuable wherever conformational ensembles and energy landscapes are used.

5.5. Algorithmic Foundations

Computational geometry, Computational topology, Voronoi diagrams, α-shapes, Morse theory.

5.5.1. *Mass Transportation Problems with Connectivity Constraints*
Participant: Frédéric Cazals.

In collaboration with D. Mazauric (Inria Sophia Antipolis Méditerranée, Geometrica).

Given two graphs, the supply and the demand graphs, we analyze the mass transportation problem between their vertices, under connectivity constraints [16]. More precisely, for every subset of supply nodes inducing a connected component of the supply graph, we require that the set of demand nodes receiving non-zero flow from this subset induces a connected component of the demand graph. As opposed to the classical problem, a.k.a the earth mover distance (EMD), which is amenable to linear programming (LP), this new problem is very difficult to solve, and we make four contributions. First, we formally introduce two optimal transportation problems, namely minimum-cost flow under connectivity constraints problem (EMD-CC) and maximum-flow under cost and connectivity constraints problem (EMD-CCC). Second, we prove that the decision version of EMD-CC is NP-complete even for very simple classes of instances. We deduce that the decision version of EMD-CCC is NP-complete, and also prove that EMD-CC is not in APX even for simple classes of instances. Third, we develop a greedy heuristic algorithm returning admissible solutions, of time complexity $O(n^3 m^2)$ with n and m the numbers of vertices of the supply and demand graphs, respectively. Finally, on the experimental side, we compare the transport plans computed by our greedy method against those produced by the aforementioned LP. Using synthetic landscapes (Voronoi landscapes), we show that our greedy algorithm is effective for graphs involving up to 1000 nodes. We also show the relevance of our algorithms to compare energy landscapes of biophysical systems (protein models).

5.5.2. *Ciruvis: a web-based Tool for Rule Networks and Interaction Detection using Rule-based Classifiers*

Participant: Simon Marillet.

In collaboration with J. Komorowski and S. Bornelöv (Uppsala University).

The use of classification algorithms is becoming increasingly important for the field of computational biology. However, not only the quality of the classification, but also its biological interpretation is important. This interpretation may be eased if interacting elements can be identified and visualized, something that requires appropriate tools and methods.

We developed a new approach to detecting interactions in complex systems based on classification [12]. Using rule-based classifiers, we previously proposed a rule network visualization strategy that may be applied as a heuristic for finding interactions. We now complement this work with Ciruvis, a web-based tool for the construction of rule networks from classifiers made of IF-THEN rules. Simulated and biological data served as an illustration of how the tool may be used to visualize and interpret classifiers. Furthermore, we used the rule networks to identify feature interactions, compared them to alternative methods, and computationally validated the findings. Rule networks enable a fast method for model visualization and provide an exploratory heuristic to interaction detection. The tool is made freely available on the web and may thus be used to aid and improve rule-based classification.

ASCLEPIOS Project-Team

5. New Results

5.1. Highlights of the Year

- Nicholas Ayache was elected a member of the Académie des sciences on 18th Nov. 2014.
- Nicholas Ayache received the "Grand Prix Inria – Académie des sciences 2014" for his major contributions to Informatics and Computational Sciences at Inria.
- Nicholas Ayache taught the "Personalized Digital Patient" course at the Collège de France on the annual chair "Informatics and Computational Sciences".
- Hervé Lombaert was awarded and ranked 1st in computer science at the highly selective NSERC Postdoctoral Fellowship (Top funding agency in Canada).
- Nina Miolane and Bishesh Khanal won the first prize in the "Popular Vote Awards" at the MICCAI 2014 Educational Challenge for their video on "Statistics on Lie groups for Computational Anatomy".

BEST PAPER AWARD :

[12] **MICCAI Workshop on Abdominal Imaging – Computational and Clinical Applications**. C. AUDIGIER, T. MANSI, H. DELINGETTE, S. RAPAKA, V. MIHALEF, D. CARNEGIE, E. BOCTOR, M. CHOTI, A. KAMEN, D. COMANICIU, N. AYACHE.

5.2. Medical Image Analysis

5.2.1. *3D/2D Coronary Arteries Registration*

Participants: Thomas Benseghir [correspondent], Grégoire Malandain [Morpheme Team], Régis Vaillant [GE-Healthcare], Nicholas Ayache.

This work has been performed in colaboration with GE-Healthcare (Buc) and the Morpheme team at Inria SAM.

3D/2D Registration, Computed Tomography Angiography, X-ray Fluoroscopy, Coronary Arteries, Vascular Tree

Integrating vessel calcifications and occlusion information, extracted from pre-operative 3D CT angiography images into a live fluoroscopic 2D image can greatly improve the guidance of percutaneous coronary interventions. Such task requires a registration step that must provide relevant correspondences between these two complementary modalities. We are developing a framework aiming at preserving the topology of the vascular structures matched between both images.

The introduction of topology in the pairing procedure allows to decrease the mismatches with respect to geometrically-based pairing procedures (e.g. Iterative Closest Point), which, in turn, improves the success rate of the registration method. This is exemplified by Fig. 3 where the proposed pairing method is compared to ICP.

5.2.2. *Video Synchronization: An Approach Towards Endoscopic Re-localization*

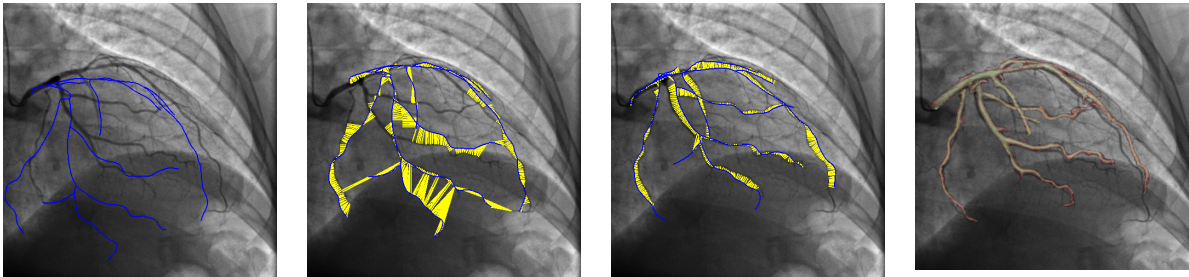**Participants:** Anant Suraj Vemuri [correspondent], Nicholas Ayache.

Figure 3. From left to right: 1. initial pose estimate (3D centerlines projection in blue); 2. iterative closest point algorithm registered position (point-pairings in yellow); 3: proposed method registered position; 4: resulting fusion between the two modalities

Endoscopy, Barrett's Esophagus, Re-localization, Electromagnetic Tracking

- Barrett's esophagus is the pre-malignant lesion for the majority of patients with esophageal adenocarcinoma. The evolution of the disease involves endoscopic surveillance for patients every 3-6 months, according to the Seattle protocol.

- The approach is labor-intensive and the primary problem is the inter-operative re-localization of these biopsy sites to guide the treatment.

- In an earlier work we had proposed a general framework for inter-operative biopsy site re-localization framework by introducing an Electro-magnetic tracking system (EMTS) into the loop and providing a way to inter-operatively register video sequences to provide a guided navigation in the esophagus.

- This work has been extended further to fit the operating room workflow. Two external landmarks have been added to the system setup as shown in Fig.4 , to obtain a complete reference frame with respect to the patient and to make the registration, patient specific [38]. The patient-localized reference frame allows the recovery of complete $SE(3)$ including the roll angle about the esophageal axis.
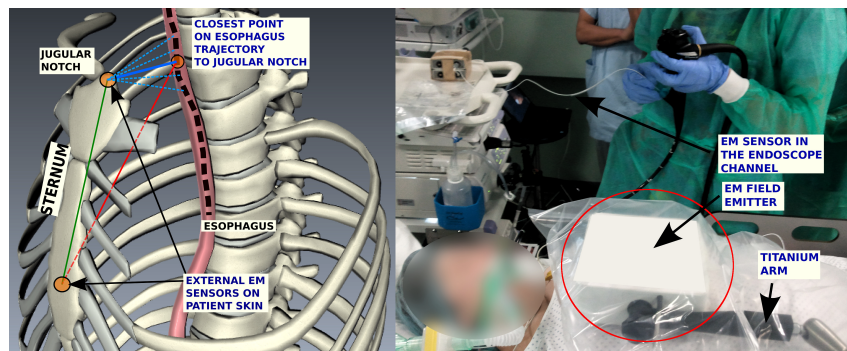

Figure 4. (Left) The orange circular markings indicate the position of the sensors taken as anatomical landmarks to form the reference frame for the patient. (Right) System setup in the operating room.

5.2.3. *A sparse Bayesian framework for non-rigid registration*

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

This work has been partly supported by Microsoft Research - Inria joint laboratory 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).

Registration, Automatic Relevance Determination, Uncertainty Quantification

We propose a sparse Bayesian framework for non-rigid registration. It provides a principled approach to efficiently find an optimal, sparse parameterization of deformations among any preset, widely overcomplete range of basis functions. It addresses open challenges in state-of-the-art registration, such as the automatic joint estimate of model parameters (e.g. noise and regularization levels). We have evaluated the feasibility and performance of our approach on cine MR, tagged MR and 3D US cardiac images, and show state-of-the-art results on benchmark datasets evaluating accuracy of motion and strain (see Fig.5). This work was presented during the MICCAI 2014 conference[20].
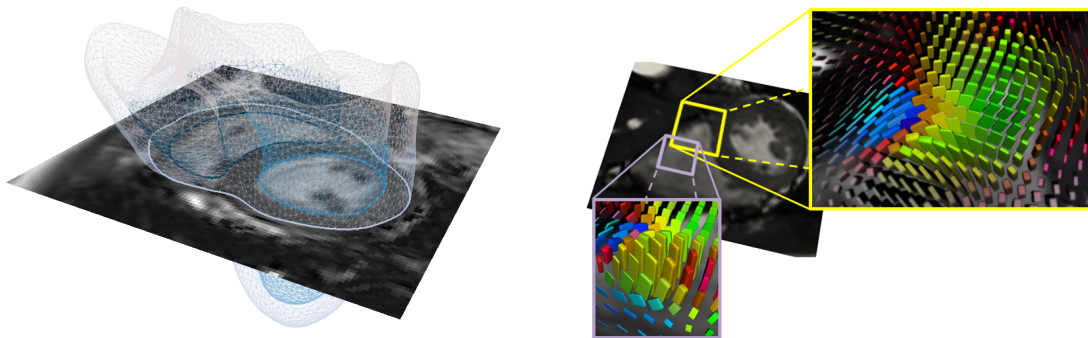

Figure 5. (Left) Mesh contour propagated to end systole via the registration output; (Right) Spatial uncertainty visualized as a tensor map.

5.2.4. *Segmentation and anatomic variability of the cochlea and other temporal bone structures from medical images*

Participants: Thomas Demarcy [correspondent], Hervé Delingette, Clair Vandersteen, Dan Gnansia [Oticon Medical], Nicholas Ayache.

This work is funded by a CIFRE grant involving Oticon Medical (Vallauris) and performed in collaboration with the IUFC Nice (Pr. Guevara) and CHU Nice (Pr. Raffaelli) .

image segmentation ; surgery planning ; shape modelling ; anatomic variability ; cochlear implant ; temporal bone

- We applied semi-automatic segmentation methods to extract anatomical structures on the inner ear on both micro-CT and CT scan images.
- μ-CT and CT images acquired on the same subject were fused with their segmentation.
- We designed a teaching tool[37] for advanced visualization of temporal bone structures (see Fig. 6).

5.2.5. *Understanding cardiac planes of acquisition*

Participants: Jan Margeta [correspondent], Nicholas Ayache, Daniel C Lee [Northwestern University], Antonio Criminisi [Microsoft Research Cambridge].
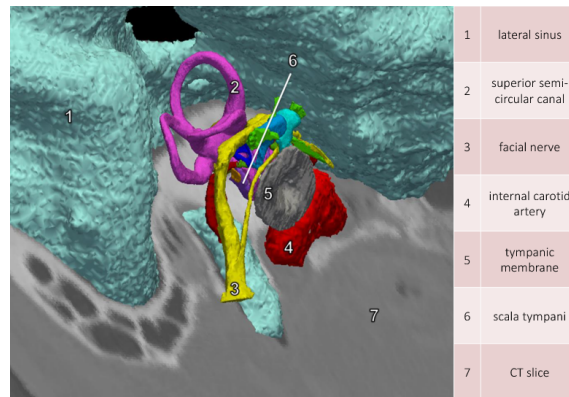
Figure 6. Virtual view of the posterior tympanotomy approach fused with a CT-scan

Cardiac imaging, Machine learning, Magnetic resonance, Data wrangling

DICOM image format defines several tags by which the images can be queried and filtered. Many useful tags are however not standardized and must be cleaned prior to any large scale analysis.

- We developed a machine learning method for automatic recognition of cardiac planes of acquisition (See Fig. 7 for sample predictions).
- Our image based method achieved state of the art performance.
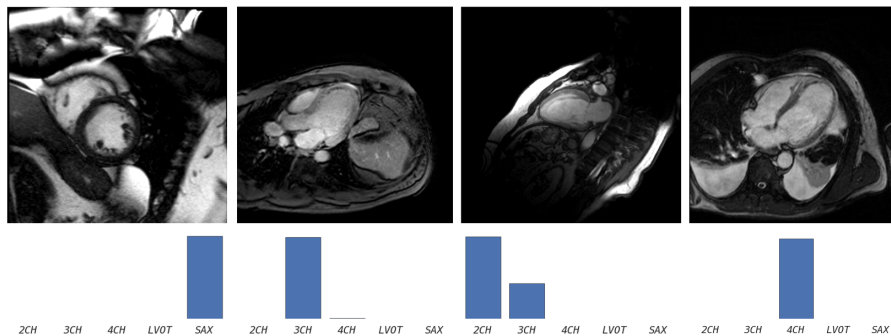- This work was presented[23] at the Medical Image Understanding and Analysis conference in London.

Figure 7. Examples of cardiac acquisition plane predictions and confidences

5.3. Computational Anatomy

5.3.1. Statistical Analysis of Diffusion Tensor Images of the Brain

Participants: Marco Lorenzi [Correspondent], Nicholas Ayache, Xavier Pennec.

Image non-linear registration, Longitudinal modeling, Alzheimer's disease

Alzheimer's disease is characterized by the co-occurrence of different phenomena, starting from the deposition of amyloid plaques and neurofibrillary tangles, to the progressive synaptic, neuronal and axonal damages. The brain atrophy is a sensitive marker of disease progression from pre-clinical to the pathological stages, and computational methods for the analysis of magnetic resonance images of the brain are currently used for group-wise (cross-sectional) and longitudinal studies of pathological morphological changes in clinical populations. The aim of this project is to develop robust and effective computational instruments for the analysis of longitudinal brain changes. In particular novel methods based on non-linear diffeomorphic registration have been investigated in order to reliably detect and statistically analyze pathological morphological changes [5] (see Fig.8). This project is also focused in the comparison of the trajectories of longitudinal morphological changes [31] estimated in different patients. This is a central topic for the development of statistical atlases of the longitudinal evolution of brain atrophy.
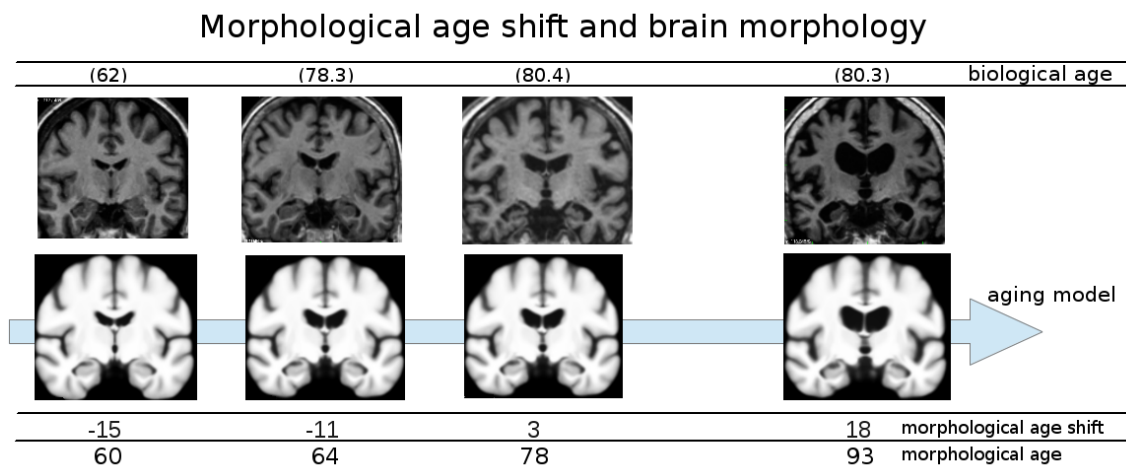

Figure 8. Modeled longitudinal brain changes in normal aging extrapolated from -15 to 18 years, and corresponding observed patient anatomies with estimated morphological age and age shift (biological age in parenthesis). Our modeling framework describes meaningful anatomical changes observed in clinical groups.

5.3.2. *Statistical Learning via Synthesis of Medical Images*

Participants: Hervé Lombaert [Correspondent], Nicholas Ayache, Antonio Criminisi.

Statistical learning, Synthesis

Machine learning approaches typically require large training datasets in order to capture as much variability as possible. Application of conventional learning methods on medical images is difficult due to the large variability that exists among patients, pathologies, and image acquisitions. The project aims at exploring how realistic image synthesis could be used, and improve existing machine learning methods.

First year tackled the problem of better exploiting existing training sets, via a smart modeling of the image space (Fig. 9), and applying conventional random forests using guided bagging [21]. Synthesis of complex data, such as cardiac diffusion images (DTI), was also done. Synthesis of complex shapes, using spectral graph decompositions, is currently on-going work.

The modeling of shapes also includes novel representations based on the spectral decomposition of images[4] which are more robust to large deformations when comparing multiple patients.
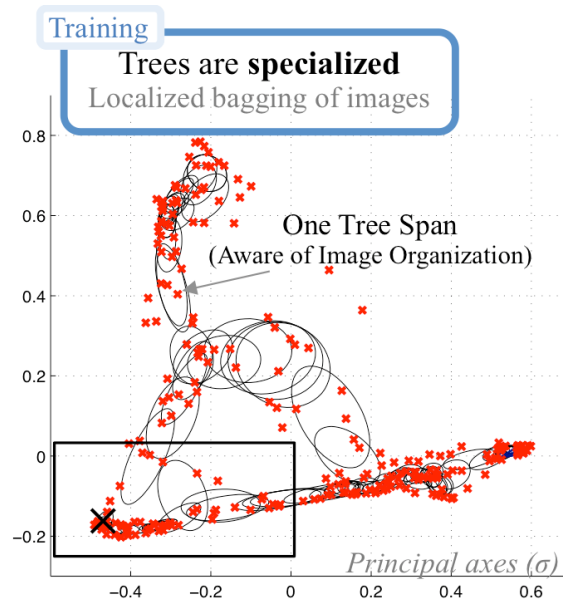

Figure 9. Laplacian Forest, where images are here represented as points, and where decision trees are trained using the spatial organization of these images on a reduced space.

5.3.3. *Statistical analysis of heart shape, deformation and motion*

Participants: Marc-Michel Rohé [correspondent], Xavier Pennec, Maxime Sermesant.

Statistical analysis, Registration, Reduced order models, Machine learning

The work aims at developping statistical tools to analyse cardiac shape, deformation, and motion. In particular, we are interested in developping reduced order models so that the variability within a population described by a complex model can be reduced into few parameters or modes that are clinically relevant. We use these modes to represent the variability seen in a population and to relate this variability with clinical parameters, and we build group-wise statistics which relate these modes to a given pathology. We focus on cardiomyopathies and the cardiovascular disease risk in obese children and adolescents.

5.3.4. *Geometric statistics for Computational Anatomy*

Participants: Nina Miolane [Correspondent], Xavier Pennec.

Lie groups, pseudo-Riemannian, Statistics, Computational Anatomy

Lie groups are widely used in mathematical models for Medical Imaging. In Computational Anatomy for example, an organ's shape can be modeled as the deformation of a reference shape, in other words : as an element of a Lie group. If one wants to analyze the variability of the human anatomy, e.g. to help diagnose diseases, one has to perform statistics on Lie groups. We investigate the geometric structures on Lie groups that enable to define consistent statistics. A Lie group G is a manifold with an additional group structure. Statistics on *Riemannian* manifolds have be studied throughout the past years. One may wonder if we could use the

theory of statistics on Riemannian manifolds for statistics on G. To this aim, we need to define a Riemannian metric on the Lie group that is *compatible with the group structure*: a so-called *bi-invariant* metric. However, it is known that most Lie groups do not admit any bi-invariant metric. One may wonder if we could generalize the theory of statistics on Riemannian manifolds to pseudo-Riemannian manifolds and use it for statistics on G. To this aim, we need to define a bi-invariant pseudo-metric on G. How many Lie groups do admit such a pseudo-metric and can we compute it? These investigations and their results (see Fig. 10) were presented at MaxEnt 2014 [24].

Structure of Lie groups with a bi-invariant metric

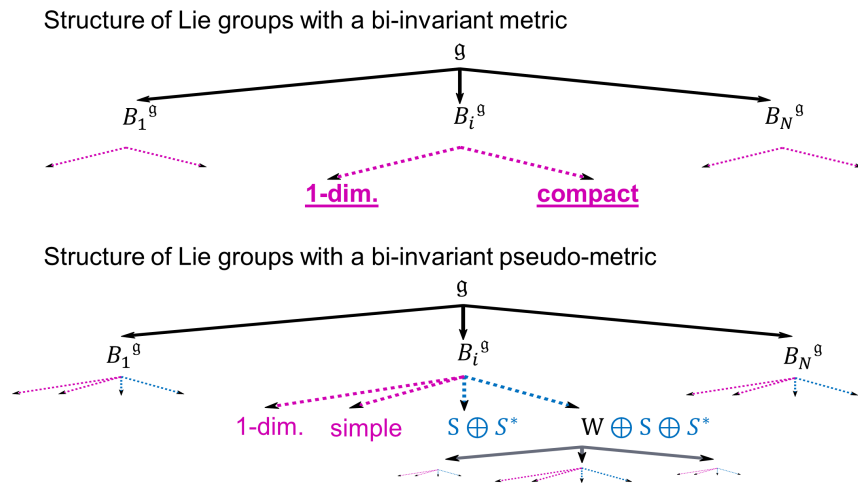Structure of Lie groups with a bi-invariant pseudo-metric

Figure 10. Structure of Lie groups on which one can define a bi-invariant metric or a bi-invariant pseudo-metric. The black levels of the tree represent the adjoint decomposition of the Lie algebra, the dashed lines represent the possible algebraic types of the substructures. Note the recursive construction in the pseudo-Riemannian case.

5.3.5. *Statistical Analysis of Diffusion Tensor Images of the Brain*

Participants: Vikash Gupta [correspondent], Nicholas Ayache, Xavier Pennec.

Population specific multimodal brain atlas for statistical analysis of white matter tracts on clinical DTI.

HIV virus can cross the hematoencephallic barrier and affect the neural connectivity in the human brain causing compromised motor controls, loss in episodic, long term memory and working memory, loss in attention/concentration and visual agnosia. These cognitive losses are characterized by the neuropsychological (NP) test scores and believed to be correlated with destruction of white matter (WM) integrity among the HIV patients. For quantifying the loss in WM integrity, the HIV subjects are compared against controls using a tract based spatial statistics (TBSS) routine. The standard TBSS routines uses univariate statistics using the fractional anisotropy (FA) maps. However, we improved on the existing routines using tensor based registration for normalizing the diffusion tensor images (DTI) followed by a multivariate statistics using the full tensor information. With the improved method it is possible to detect differences in WM regions which was not possible using the existing TBSS routines. For this study a population specific multimodal (T1 and DTI) brain atlas was developed from the population. The joint atlas also contains a probabilistic parcellation of WM regions in the brain which can be used for region of interest (ROI) based statistical studies (see Fig.11).

5.3.6. *Longitudinal Analysis and Modeling of Brain Development*

Participants: Mehdi Hadj Hamou [correspondent], Xavier Pennec, Nicholas Ayache.
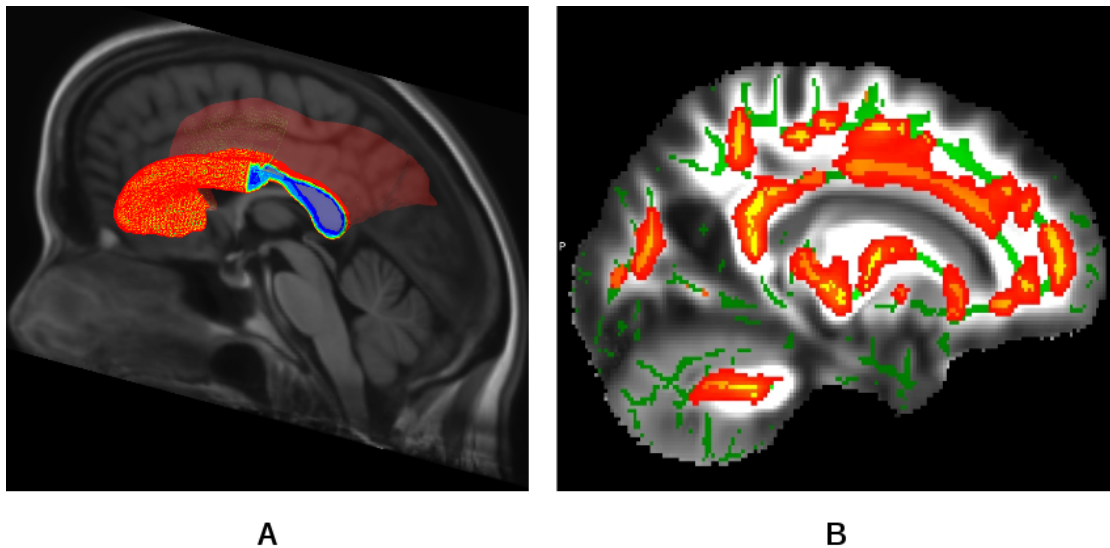
Figure 11. ***A:*** *Probabilistic parcellation of corpus callosum with blue and red being the maximum and minimum probability regions respectively.* ***B:*** *Multivariate statistics on white matter tracts. The red-yellow sections show statistically significant differences*

Brain development, adolescence, longitudinal analysis, non-rigid registration algorithm, extrapolation, interpolation

This work is divided into 2 complementary studies about longitudinal trajectories modeling:

- Diffeomorphic registration parametrized by Stationary Velocity Fields (SVF) is a promising tool already applied to model longitudinal changes in Alzheimer's disease. However, the validity of these model assumptions in faithfully describing the observed anatomical evolution needs to be further investigated. In this work, we thus analyzed the effectiveness of linear regression of SVFs in describing anatomical deformations estimated from past and future observations of the MRIs.

- Due to the lack of tools to capture the subtle changes in the brain, little is known about its development during adolescence. The aim of this project is to provide quantification and models of brain development during adolescence based on diffeomorphic registration parametrized by SVFs (see Fig.12). We particularly focused our study on the link between gender and the longitudinal evolution of the brain. This work was done in collaboration with J.L. Martinot et H. Lemaître (Inserm U1000).

5.4. Computational Physiology

5.4.1. Biophysical Modeling and Simulation of Longitudinal Brain MRIs with Atrophy in Alzheimer's Disease

Participants: Bishesh Khanal [correspondent], Nicholas Ayache, Xavier Pennec.
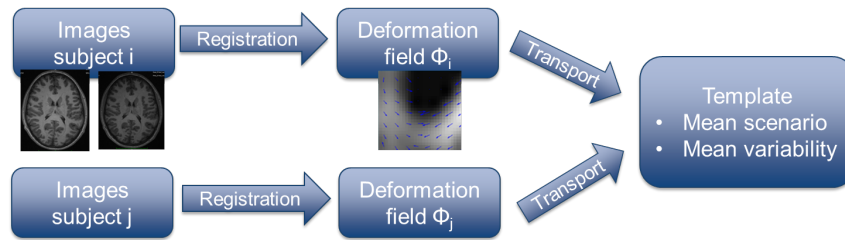
Figure 12. Pipeline for the longitudinal analysis of brain development during adolescence.

Alzheimer's Disease (AD), modeling brain deformation, biophysical model, simulation

- We propose a biophysical model of brain deformation due to atrophy in Alzheimer's Disease(AD) [17]. The model allows simulation of longitudinal brain MRIs with a desired level of atrophy in brain parenchyma. Here we enhanced our previous implementation to model brain parenchyma and cerebrospinal fluid (CSF) differently so that there is no need to prescribe atrophy in CSF region (see Fig.13).

- The model could be used to explore different possible hypotheses about evolution of atrophy in the brain and how it affects the brain shape changes.
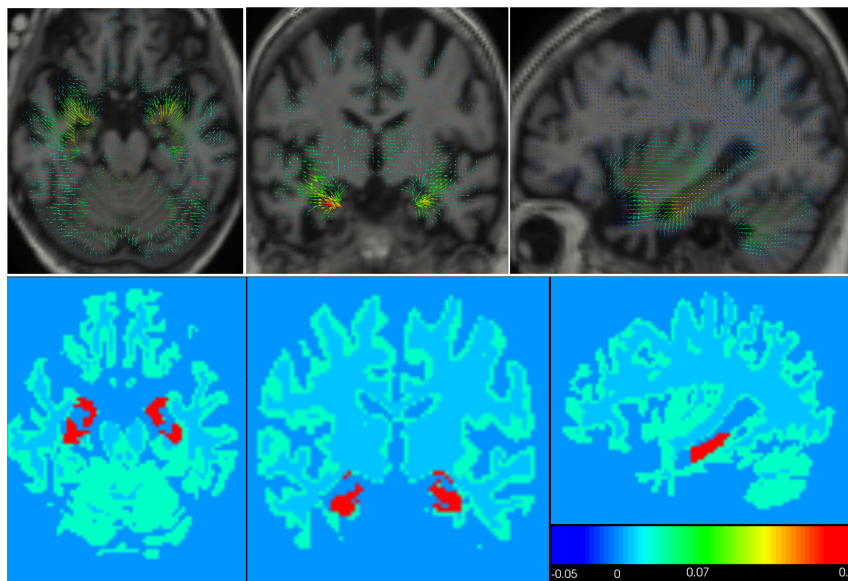

Figure 13. An example of obtained deformation field (top) from the model for the prescribed atrophy (bottom). From left to right: Axial, Coronal and Sagittal views.

5.4.2. Glioblastoma : Study of the vasogenic edema

Participants: Matthieu Lê [correspondent], Hervé Delingette, Jan Unkelbach [Massachussetts General Hospital], Nicholas Ayache.

This work is carried out between Ascelpios research group, Inria Sophia Antipolis, France and the Department of Radiation Oncology of the Massachusetts General Hospital, Boston, USA.

Glioblastoma, Vasogenic Edema, Radiotherapy, Target Delineation

- We studied the impact of anti-angiogenic treatment on the MRI appearance of glioblastoma.

- We studied how MRI extracted features could help distinguish between the vasogenic edema and the tumor infiltration[22].

- We analyzed the impact of excluding the vasogenic edema from the gross tumor volume during radiation therapy (see Fig. 14). Our approach leads to a dose more comformal to the underlying tumor cell density knowing that prescribing less dose might open the way for later re-irradiation.
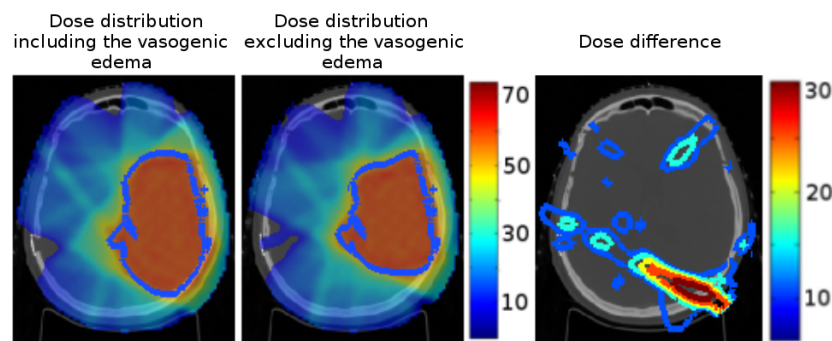

Figure 14. Comparison of the dose distribution including the vasogenic edema (clinical practice) and the excluding the estimated vasogenic edema (proposed method).

5.4.3. *Image-based Prediction of Cardiac Ablation Targets*

Participants: Rocio Cabrera Lozoya [correspondent], Maxime Sermesant, Nicholas Ayache.

Electrophysiology, ablation planning, machine learning

Ventricular radio-frequency ablation (RFA) can have a critical impact on preventing sudden cardiac arrest but is challenging due to a highly complex arrhythmogenic substrate. We aim at identifying local image characteristics capable of predicting the presence of local abnormal ventricular activities (LAVA). This could lead to pre-operatively and non-invasively improve and accelerate the procedure.

- We present the use of intensity and texture-based local imaging features in the vicinity of myocardial scar and grey zones towards the prediction of RFA target localisation (see Fig.15).

- We detail the uncertainty in the data and explore its impact on the classification results.

- A preliminary output with visual interpretation and potential use in a clinical environment was presented.

- The encouraging obtained results warrant further investigation and open up possibilities for non-invasive cardiac arrhythmia ablation planning. [13]

5.4.4. *Personalised Canine Electromechanical Model of the Heart*

Participants: Sophie Giffard-Roisin [correspondent], Maxime Sermesant, Hervé Delingette, Stéphanie Marchesseau, Nicholas Ayache.
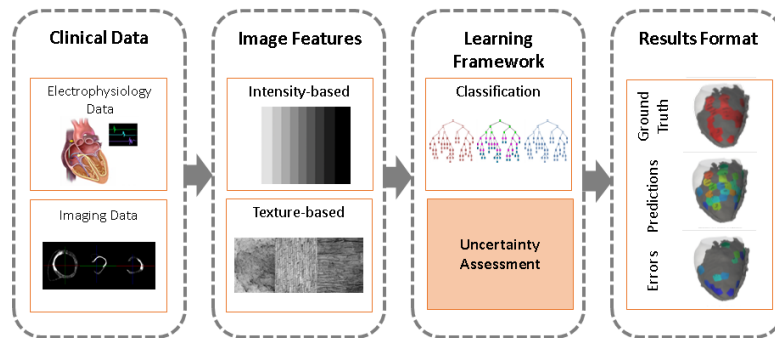
Figure 15. Pipeline showing the processing of our multimodal data. It includes an image feature extraction phase, followed by a classification with uncertainty assessment stage. The rightmost panel shows the preliminary output result format for a clinical environment.

Cardiac Modelling, Personalised Simulation, Electrical and Mechanical Simulation

- We studied the coupled electro-mechanical modelling of the heart, where the mechanics is handled by the Bestel-Clement-Sorine model while the electrophysiological phenomena is driven by an Eikonal model (see Fig. 16).

- We participated to the STACOM'2014 LV Mechanics Challenge in Boston[15] where four healthy canine clinical data (left ventricles) were provided. The validation was performed on local displacements. Our model has been calibrated by a quantitative sensitivity study as well as a personalized automatic calibration.
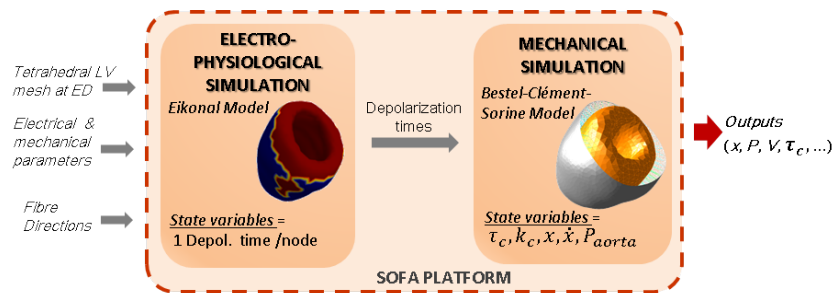

Figure 16. Complete electromechanical pipeline used for the simulations of four healthy canine hearts.

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

Participants: Chloé Audigier [correspondent], Hervé Delingette, Tommaso Mansi, Nicholas Ayache.

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

Radio Frequency Abation, Patient-Specific Simulation, Lattice Boltzmann Method, Computational Fluid Dynamics, Heat Transfer, Therapy Planning, Liver

Radiofrequency abation (RFA) is a minimally invasive therapy suited for liver tumor ablation. However a patient-specific predicitive tool to plan and guide the treatment is required. We developed a computational framework for patient-specific planning of RFA (see Fig.17) :

- a personalised forward model of RFA:
- A patient-specific detailed anatomical model of the liver is estimated from standard CT image and meshed to generate a tetrahedral volume mesh. The structures of interest include the parenchyma, lesion, hepatic vein and vena cava.
- A Computation Fluic Dynamic and porous media solver using the Lattice Boltzmann Method is used to compute the patient-specific blood flow in the hepatic circulatory system and the blood flow distribution inside the parenchyma.
- Bio-heat equation has been implemented with a Lattice Boltmann Method also to model efficiently the heat propagation in biological tissues accounting for the cooling effect of neighboring vessels. A cell death model have been combined to account for the cellular necrosis.

Then this forward model is used to estimate patient-specific model parameters as presented in the ABDI workshop at MICCAI 2014 [12]. This work presented obtained the best paper award of the workshop.
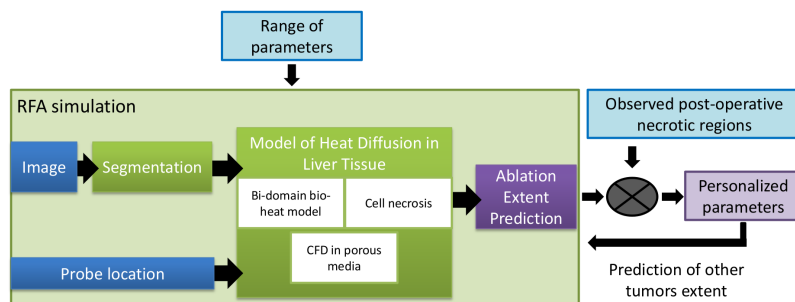
Figure 17. Steps of the proposed method for parameters personnalisation (blue: input, green: processes, purple: output).

5.4.6. *Multi-channel patch-based glioma segmentation*

Participants: Nicolas Cordier [correspondent], Hervé Delingette, Nicholas Ayache.

Part of 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, Glioma, Patch-based Segmentation, Tumor Simulation

The segmentation of glioblastoma, the most severe case of brain tumors, is a crucial step for diagnostic assessment and therapy planning. In order to perform the manual delineation of the tumor compartments, the clinicians have to concurrently screen multi-channel 3D MRI, which makes the process both time-consuming and subject to inter-expert delineation variability. We are building upon the patch-based segmentation framework, the state-of-the-art for the segmentation of healthy brain structures, to present automatic glioma segmentation algorithms. Our 2013 submission to the MICCAI Brain Tumor Segmentation Challenge has been improved by:
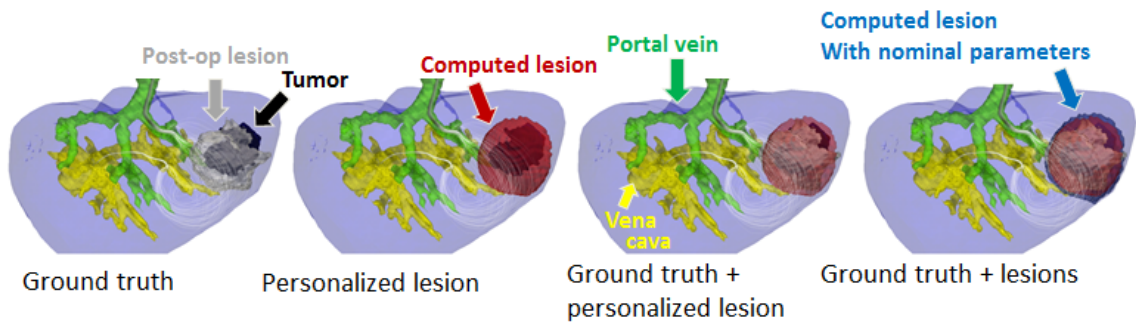
Figure 18. Predicted necrosis compared qualitatively well with ground truth (necrosis zone observed on a post-operative image).

- replacing the heuristic label fusion strategy with a more robust approach,
- integrating information such as statistics of appearance and position,
- generating configurations of synthetic training patches,
- filtering out the training patches for which the labels are less reliable (see Fig.19).
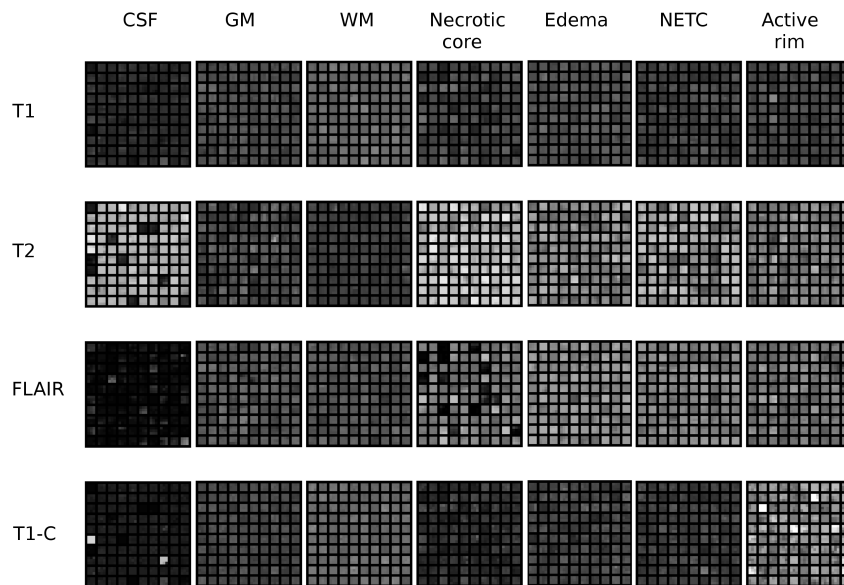
Figure 19. 2D slices of randomly sampled 3D multi-channel patches. From left to right: the different MR channels (T1, T2, T2-FLAIR, contrast enhanced T1). From top to bottom: cerbrospinal fluid (CSF), grey matter (GM), white matter (WM), necrotic tumor core, edema, non-enhancing tumor core (NETC), and active rim.

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

6. New Results

6.1. Highlights of the Year

Maureen Clerc was awarded the PIERRE FAURRE Prize by the French Academy of Sciences. This award recognizes her outstanding contributions to the modelling and interpretation of electrical signals in the brain. The ceremony took place at the Institut de France on October 14th, 2013.

Emmanuel Caruyer was awarded the AFRIF Best PhD thesis award 2013 for his work "Q-space diffusion MRI: Acquisition and Signal Processing" performed under the direction of Rachid Deriche. He received the award thesis AFRIF 2013 during RFIA Conference held from June 30 to July 4, 2014 in Rouen.

Rachid Deriche was awarded the title of Honorary Doctor (honoris causa) from the University of Sherbrooke, Canada. This award recognises his achievements and contributions to image processing, computer vision and computational brain imaging. The title was awarded at the academic conferment ceremony held on September 20th, 2014 at the University of Sherbrooke.

Théo Papadopoulo has been promoted to the position of Research Director Class 2, starting from October 1st, 2014.

6.2. Modeling in Diffusion MRI

6.2.1. *Non-Negative Spherical Deconvolution (NNSD) for estimation of fiber Orientation Distribution Function in single-/multi-shell diffusion MRI*

Participants: Jian Cheng [University of North Carolina at Chapel Hill,USA], Tianzi Jiang [LIAMA, China], Shen Dinggang [University of North Carolina at Chapel Hill,USA], Yap Pew-Thian [University of North Carolina at Chapel Hill,USA], Rachid Deriche.

Spherical Deconvolution (SD) is commonly used for estimating fiber Orientation Distribution Functions (fODFs) from diffusion-weighted signals. Existing SD methods can be classified into two categories: 1) Continuous Representation based SD (CR-SD), where typically Spherical Harmonic (SH) representation is used for convenient analytical solutions, and 2) Discrete Representation based SD (DR-SD), where the signal profile is represented by a discrete set of basis functions uniformly oriented on the unit sphere. A feasible fODF should be non-negative and should integrate to unity throughout the unit sphere SS2. However, to our knowledge, most existing SH-based SD methods enforce non-negativity only on discretized points and not the whole continuum of SS2. Maximum Entropy SD (MESD) and Cartesian Tensor Fiber Orientation Distributions (CT-FOD) are the only SD methods that ensure non-negativity throughout the unit sphere. They are however computational intensive and are susceptible to errors caused by numerical spherical integration. Existing SD methods are also known to overestimate the number of fiber directions, especially in regions with low anisotropy. DR-SD introduces additional error in peak detection owing to the angular discretization of the unit sphere. This work proposes a SD framework, called Non-Negative SD (NNSD), to overcome all the limitations above. NNSD is significantly less susceptible to the false-positive peaks, uses SH representation for efficient analytical spherical deconvolution, and allows accurate peak detection throughout the whole unit sphere. We further show that NNSD and most existing SD methods can be extended to work on multi-shell data by introducing a three-dimensional fiber response function. We evaluated NNSD in comparison with Constrained SD (CSD), a quadratic programming variant of CSD, MESD, and an L1-norm regularized non-negative least-squares DR-SD. Experiments on synthetic and real single-/multi-shell data indicate that NNSD improves estimation performance in terms of mean difference of angles, peak detection consistency, and anisotropy contrast between isotropic and anisotropic regions.

This work has been published in [11].

6.2.2. *Quantitative comparison of reconstruction methods for intra-voxel fiber recovery from diffusion MRI*

Participants: Alessandro Daducci [LTS5, Ecole Polytech. Fed. de Lausanne (EPFL)], Maxime Descoteaux [SCIL Lab., Sherbrooke University], Michael Paquette [SCIL Lab., Sherbrooke University], Sylvain Merlet, Emmanuel Caruyer, Rachid Deriche.

Validation is arguably the bottleneck in the diffusion magnetic resonance imaging (MRI) community. This work evaluates and compares 20 algorithms for recovering the local intra-voxel fiber structure from diffusion MRI data and is based on the results of the "HARDI reconstruction challenge" organized in the context of the "ISBI 2012" conference. Evaluated methods encompass a mixture of classical techniques well known in the literature such as diffusion tensor, Q-Ball and diffusion spectrum imaging, algorithms inspired by the recent theory of compressed sensing and also brand new approaches proposed for the first time at this contest. To quantitatively compare the methods under controlled conditions, two datasets with known ground-truth were synthetically generated and two main criteria were used to evaluate the quality of the reconstructions in every voxel: correct assessment of the number of fiber populations and angular accuracy in their orientation. This comparative study investigates the behavior of every algorithm with varying experimental conditions and highlights strengths and weaknesses of each approach. This information can be useful not only for enhancing current algorithms and develop the next generation of reconstruction methods, but also to assist physicians in the choice of the most adequate technique for their studies.

This work has been published in [12]

6.2.3. *Comparison of sampling strategies and sparsifying transforms to improve compressed sensing diffusion spectrum imaging*

Participants: Michael Paquette [SCIL Lab., Sherbrooke University], Sylvain Merlet, Guillaume Gilbert [SCIL Lab., Sherbrooke University], Maxime Descoteaux [SCIL Lab., Sherbrooke University], Rachid Deriche.

Diffusion Spectrum Imaging enables to reconstruct the ensemble average propagator (EAP) at the expense of having to acquire a large number of measurements. Compressive sensing offers an efficient way to decrease the required number of measurements. The purpose of this work is to perform a thorough experimental comparison of three sampling strategies and six sparsifying transforms to show their impact when applied to accelerate compressive sensing-diffusion spectrum imaging. In this work, we propose a novel sampling scheme that assures uniform angular and random radial q-space samples. We also compare and implement six discrete sparse representations of the EAP and thoroughly evaluate them on synthetic and real data using metrics from the full EAP, kurtosis, and orientation distribution function. The discrete wavelet transform with Cohen–Daubechies–Feauveau 9/7 wavelets and uniform angular sampling in combination with random radial sampling showed to be better than other tested techniques to accurately reconstruct the EAP and its features. In conclusion, it is important to jointly optimize the sampling scheme and the sparsifying transform to obtain accelerated compressive sensing-diffusion spectrum imaging. Experiments on synthetic and real human brain data show that one can robustly recover both radial and angular EAP features while undersampling the acquisition to 64 measurements (undersampling factor of 4).

This work has been published in [16]

6.2.4. *DSI 101: Better ODFs for Free*

Participants: Michael Paquette [SCIL Lab., Sherbrooke University], Sylvain Merlet, Maxime Descoteaux [SCIL Lab., Sherbrooke University], Rachid Deriche.

Diffusion Spectrum Imaging (DSI) is a well established method to recover the Ensemble Average Propagator (EAP). The orientation distribution function (ODF) is computed from this discretized EAP and used for tractography. However, there are several important implementation considerations that are tossed aside in the literature and the publicly available softwares. We investigate all the real steps necessary to go from the DSI signal to the ODF and provide applicable recommendations that greatly improve the accuracy of the local

orientation detected. These recommendations come "free-of-charge" as they are applicable to all existing DSI data and do not require a significant increase in computation time.

This work has been published in [26]

6.2.5. *Comparison between discrete and continuous propagator indices from Cartesian q-space DSI sampling*

Participants: Mauro Zucchelli [Dpt of Computer Science, University of Verona], Eleftherios Garyfallidis [SCIL Lab., Sherbrooke University], Michael Paquette [SCIL Lab., Sherbrooke University], Maxime Descoteaux [SCIL Lab., Sherbrooke University], Gloria Menegaz [Dpt of Computer Science, University of Verona], Sylvain Merlet.

DSI is often considered the state-of-the-art technique to analyze q-space measurements sampled from a Cartesian grid. The 3D fast Fourier transform is used to directly obtain a discrete version of the EAP (Ensemble Average Propagator). DSI was one of the first techniques used to infer complex fiber configurations as it allows resolving crossings. In principle, DSI also captures some radial information which,in theory, can be used to extract diffusion features of the EAP. However, a discrete propagator representation suffers from a limited frequency band, which makes infinite integration impossible. Hence, EAP derived indices 2,3 are problematic and quantitatively questionable, as one needs to artificially normalize and approximate the infinite integrals. Combined with the recent popularity of DSI in the Human Connectome Project, it is important to investigate the different angular and EAP indices that can be computed from these DSI datasets. In this work, we investigate alternatives to the discrete model-free approach of DSI and investigate the Simple Harmonic Oscillator based Reconstruction and Estimation 3 (SHORE) models based on the evaluation of (i) the orientation distri bution function (ODF) ; (ii) the return to the origin probability (RTOP) and (iii) the mean square displacement (MSD).

This work has been published in [33]

6.2.6. *Odf Maxima Computation Using Hill Climbing Algorithm*

Participants: Makhlouf Laouchedi [USTHB, Algeria], Thinhinane Megherbi [USTHB, Algeria], Linda Oulebsir-Boumghar [USTHB, Algeria], Rachid Deriche.

Methods like Diffusion Spectrum Imaging (DSI), High Angular Resolution Diffusion Imaging (HARDI) and the High Order Tensor techniques have been proposed to reconstruct specific functions like the Orientation Distribution Function (ODF) whose maxima correspond to the directions of the multiple fibers. In this work, we are interested to extract all the crossing fibers characterized as the maxima of the Orientation Distribution Function (ODF). A Hill Climbing algorithm based approach has been developed and implemented to efficiently and accurately extract all the fibers. Promising experimental results obtained with synthetic and real data illustrate the potential of the technique.

This work has been published in [24]

6.2.7. *Greedy NNLS: Fiber Orientation Distribution from Non-Negatively Constrained Sparse Recovery*

Participants: Aurobrata Ghosh, Rachid Deriche.

In this work, we validated experimentally the merits of the Non-Negative Least Squares (NNLS) for the constrained sparse recovery of the Fiber Orientation Distribution (FOD) and compared it with classical l1-minimization. The FOD is a robust model for mapping crossing white matter fibers. However, its angular resolution depends on the spherical harmonic basis order, which can imply a large number of acquisitions. Further, it is necessary to compute the maxima of the FOD to derive the fiber directions. It is possible to kill the two proverbial birds with a single stone by using a non-negatively constrained sparse recovery FOD estimation with NNLS.

From our experiments, we confirmed results from literature to show that NNLS converges to highly sparse solutions which are correctly constrained, while l1-minimization is less sparse, contains negative solutions and is unstable with noisy data. Finally, we discussed the NLS algorithm and attributed the sparsity to its design, which mirrors the design of Orthogonal Matching Pursuit (OMP)

This work has been published in [22]

6.2.8. *Crossing Fibers Detection with an Analytical High Order Tensor Decomposition*

Participants: Thinhinane Megherbi [USTHB, Algeria], Mouloud Kachouane [USTHB, Algeria], Linda Oulebsir-Boumghar [USTHB, Algeria], Rachid Deriche.

Diffusion magnetic resonance imaging (dMRI) is the only technique to probe in vivo and noninvasively the fiber structure of human brain white matter. Detecting the crossing of neuronal fibers remains an exciting challenge with an important impact in tractography. In this, we tackle this challenging problem and propose an original and efficient technique to extract all crossing fibers from diffusion signals. To this end, we start by estimating, from the dMRI signal, the so-called Cartesian tensor fiber orientation distribution (CT-FOD) function, whose maxima correspond exactly to the orientations of the fibers. The fourth order symmetric positive definite tensor that represents the CT-FOD is then analytically decomposed via the application of a new theoretical approach and this decomposition is used to accurately extract all the fibers orientations. Our proposed high order tensor decomposition based approach is minimal and allows recovering the whole crossing fibers without any a priori information on the total number of fibers. Various experiments performed on noisy synthetic data, on phantom diffusion, data and on human brain data validate our approach and clearly demonstrate that it is efficient, robust to noise and performs favorably in terms of angular resolution and accuracy when compared to some classical and state-of-the-art approaches.

This work has been published in [15] and [34].

6.2.9. *Complete set of Invariants of a 4th order tensor: the 12 tasks of HARDI from Ternary Quartics*

Participants: Théodore Papadopoulo, Auro Ghosh, Rachid Deriche.

In this work, we presented a simple and systematic method to compute a functionally complete set of invariants of a non-negative 3D 4th order tensor with respect to 3D rotations. Intuitively, this transforms the tensor's non-unique ternary quartic (TQ) decomposition (from Hilbert's theorem) to a unique canonical representation independent of orientation.

Invariants play a crucial role in diffusion MRI. In DTI (2nd order tensors), invariant scalars (FA, MD...)have been successfully used in clinical applications. But DTI has limiations and HARDI models (e.g. 4th order tensors) have been proposed instead. These, however, lack invariant features and computing them systematically is challenging.

The invariants we propose, can be computed from two simple reduction steps, which first reduce an orthogonal class and then a rotation transform class of equivalent representations from the TQ coefficients. The resulting invariants are, by construction, (1) functionally complete, (2) functionally irreducible (if desired), (3) computationally efficient and (4) reversible – or mappable to the TQ coefficients or shape. These were the novelties of our contribution in comparison to prior work.

This work has been published in [25]

6.2.10. *Fiber Orientation Distribution from Non-Negative Sparse Recovery*

Participants: Thinhinane Megherbi [USTHB, Algeria], Auro Ghosh, Linda Oulebsir-Boumghar [USTHB, Algeria], Rachid Deriche.

In this work, we tested our non-negatively constrained sparse recovery algorithm for estimating the FOD on single shell phantom data provided by the ISBI'2014 challenge. We used the NNLS algorithm to estimate high order FODs (24th order) from just 20, 30 and 60 gradient directions and for various b-values of 1000, 2000, and 3000.

From the results, which are yet to be published, but can be viewed in their preliminary form online, it is clear that amongst the single shell algorithms, ours was good at fitting the signal and estimating the number of compartments. It performed well even with as low as 20 gradient acquisitions. Its major shortcoming was in underestimating the crossing angle and this needs to be improved upon.

This work has been published in [35]

6.2.11. *How to get more out of a clinically feasable 64 gradient dMRI acquisition: Multi-Shell versus Single-Shell*

Participants: Rutger H.j Fick, Mario Zuccheli [Dpt of Computer Science, University of Verona], Gabriel Girard [SCIL Lab., Sherbrooke University], Maxime Descoteaux [SCIL Lab., Sherbrooke University], Gloria Menegaz [Dpt of Computer Science, University of Verona], Rachid Deriche.

For clinical applications the number of diffusion MRI (dMRI) samples that can be obtained is often limited by scanner time and patient comfort. For this reason one often uses short scanning protocols that acquire just 32 or 64 gradient directions using a single b-value to obtain diffusion measures such as the fractional anisotropy from Diffusion Tensor Imaging (DTI) or to estimate the white matter orientation using Constrained Spherical Deconvolution (CSD). Using 3D-SHORE and MAP-MRI, we show that by spreading the same number of dMRI samples over different b-shells (sampling angularly and radially) we can estimate not only the directionality of the white matter using the ODF, but also the radially dependent higher order diffusion measures that SHORE and MAP-MRI provide. This approach lends itself well for situations where acquisition time is limited, and is therefore particularly well suited for clinical applications.

This work has been submitted to ISMRM'2015.

6.3. From dMRI to Fiber Pathways

6.3.1. *Towards quantitative connectivity analysis: reducing tractography biases*

Participants: Gabriel Girard [SCIL Lab., Sherbrooke University], Kevin Whittingstall [SCIL Lab., Sherbrooke University], Maxime Descoteaux [SCIL Lab., Sherbrooke University], Rachid Deriche.

Diffusion MRI tractography is often used to estimate structural connections between brain areas and there is a fast-growing interest in quantifying these connections based on their position, shape, size and length. However, a portion of the connections reconstructed with tractography is biased by their position, shape, size and length. Thus, connections reconstructed are not equally distributed in all white matter bundles. Quantitative measures of connectivity based on the streamline distribution in the brain such as streamline count (density), average length and spatial extent (volume) are biased by erroneous streamlines produced by tractography algorithms. In this work, solutions are proposed to reduce biases in the streamline distribution. First, we propose to optimize tractography parameters in terms of connectivity. Then, we propose to relax the tractography stopping criterion with a novel probabilistic stopping criterion and a particle filtering method, both based on tissue partial volume estimation maps calculated from a T1-weighted image. We show that optimizing tractography parameters, stopping and seeding strategies can reduce the biases in position, shape, size and length of the streamline distribution. These tractography biases are quantitatively reported using in-vivo and synthetic data. This is a critical step towards producing tractography results for quantitative structural connectivity analysis.

This work has been published in [13]

6.3.2. *Choosing tractography parameters to improve connectivity mapping*

Participants: Gabriel Girard [SCIL Lab., Sherbrooke University], Kevin Whittingstall [SCIL Lab., Sherbrooke University], Maxime Descoteaux [SCIL Lab., Sherbrooke University], Rachid Deriche.

Diffusion-weighted imaging (DWI) is often used as a starting point for in vivo white matter (WM) connectivity to reconstruct potential WM pathways between brain areas. Tractography algorithms have many parameters which can influence reconstruction and connectivity. Various choices of parameters have been proposed. But how to choose the best set of parameters ? In this study, we varied three critical parameters while monitoring connectivity score using the Tractometer evaluation system on the International Symposium on Biomedical Imaging (ISBI) Challenge synthetic dataset. The three parameters were: The maximum deviation angle between two consecutive tractography steps (this addresses the hypothesis of smoothness of the WM pathways), the spherical function (SF) threshold (this aims at removing noisy propagation directions during the tractography process) and the initial SF threshold (this aims at removing initial noise at the seeds and to start tractography in a good tangent direction to the WM bundle).

This work has been published in [20]

6.3.3. *Connectivity directionally-encoded color map: a streamline-based color mapping*

Participants: Gabriel Girard [SCIL Lab., Sherbrooke University], Kevin Whittingstall [SCIL Lab., Sherbrooke University], Maxime Descoteaux [SCIL Lab., Sherbrooke University], Rachid Deriche.

In this work, we provide a novel method to map streamlines in a color image, which can be generated from any set of streamlines. We show that this novel orientation color-coded map based on streamline tractography can improve connectivity analysis.

This work has been published in [23]

6.4. From dMRI to Microstructures Recovery

6.4.1. *NMR characterization of cylinder radii distribution using a SHORE-based regularization method.*

Participants: Gonzalo Sanguinetti, Matt Hall [Centre for Medical Image Computing, Dept. Computer Science, UCL], Daniel Alexander [Centre for Medical Image Computing, Dept. Computer Science, UCL], Rachid Deriche.

In this work, we are interested in retrieving information about the axon diameter distributions in white matter fiber bundles using NMR, which are commonly modelled as ensembles of parallel cylinders. We add regularization to the 1D-SHORE basis which results in more stable characterization of diameter distributions. To validate the method, we simulate NMR signals using the open source toolkit CAMINO. The results illustrate the enhanced estimation accuracy given by the regularization and provide an alternative validation of the SHORE based method.

This work has been published in [30]

6.4.2. *Mapping Average axon diameters under long diffusion time*

Participants: Gonzalo Sanguinetti, Rachid Deriche.

This work proposes an original method to recover axon diameter distribution (ADD) parameters using nuclear magnetic resonance. White matter (WM) is modelled as a bi-compartmental medium composed of an intra axonal space where the diffusion is restricted and an extra axonal space where diffusion is hindered. Under the assumption of long diffusion time, we provide a novel and efficient model for the component of the signal due to the restricted part. This technique might be interpreted as an interesting simplification of the AxCaliber framework, which leads to a simpler model and an extremely faster acquisition protocol. To test and validate our method, we use the open-source toolkit Camino for computing Monte-Carlo simulations of NMR data and model the WM as 3D cubic environments, formed by parallel cylinders with gamma distributed radii. Promising experimental results illustrate the potential of the proposed method.

This work has been published in [29] and [28]

6.4.3. Magnitude and complex based diffusion signal reconstruction

Participants: Marco Pizzolato, Timothe Boutelier [Olea Medical, La Ciotat], Rachid Deriche.

In Diffusion Weighted Magnetic Resonance Imaging (DW-MRI) the modeling of the magnitude signal is complicated by the Rician distribution of the noise. It is well known that when dealing instead with the complex valued signal, the real and imaginary parts are affected by Gaussian distributed noise and their modeling can thus benefit from any estimation technique suitable for this noise distribution. We present a quantitative analysis of the difference between the modeling of the magnitude diffusion signal and the modeling in the complex domain. The noisy complex and magnitude diffusion signals are obtained for a physically realistic scenario in a region close to a restricting boundary. These signals are then fitted with the Simple Harmonic Oscillator based Reconstruction and Estimation (SHORE) bases and the reconstruction performances are quantitatively compared. The noisy magnitude signal is also fitted by taking into account the Rician distribution of the noise via the integration of a Maximum Likelihood Estimator (MLE) in the SHORE. We discuss the performance of the reconstructions as function of the Signal to Noise Ratio (SNR) and the sampling resolution of the diffusion signal. We show that fitting in the complex domain generally allows for quantitatively better signal reconstruction, also with a poor SNR, provided that the sampling resolution of the signal is adequate. This applies also when the reconstruction is compared to the one performed on the magnitude via the MLE.

This work has been published in [27]

6.4.4. Extracting a biomarker for the mean cross-sectional area from the ODF

Participants: Rutger H.j Fick, Gonzalo Sanguinetti, Rachid Deriche.

Finding new biomarkers related to the microstructure of white matter (WM) is an active area of research in the MRI community. As opposed to the usual MRI markers such as fractional anisotropy (FA), these biomarkers provide a closer insight on the tissue structure. We introduce a new microstructure based biomarker that is related to the axon diameter distribution (ADD) and can be obtained with a q-space imaging technique like DSI or MAP . This feature is related with the nature and purpose of WM paths in bothnormal and pathological conditions and is obtained from the Orientation Distribution Function (ODF) as twice its maximum value. We show that this value is related with the mean cross-sectional area (MCSA) of an ensemble of parallel axons. The same geometric feature was proposed as a scalar index of microstructure, but was not related to the ODF. In this work we give the formal relation between this microstructure feature and the ODF, and validate it using state-of-the-art numerical simulations.

This work has been published in [18].

6.4.5. An Analytical 3D Laplacian Regularized SHORE Basis and Its Impact on EAP reconstruction and Microstructure Recovery

Participants: Rutger H.j Fick, Demian Wassermann, Gonzalo Sanguinetti, Rachid Deriche.

In diffusion MRI, the reconstructed Ensemble Average Propagator (EAP) from the diffusion signal provides detailed insights on the diffusion process and the underlying tissue microstructure. Recently, the Simple Harmonic Oscillator based Reconstruction and Estimation (SHORE) basis was proposed as a promising method to reconstruct the EAP. However, the fitting of the basis is sensitive to noise. To solve this we propose to use the Laplacian of the SHORE basis as a natural regularization functional. We provide the derivation of the Laplacian functional and compare its effect on EAP reconstruction with that of separated regularization of the radial and angular parts of the SHORE basis. To find optimal regularization weighting we use generalized cross-validation and validate our method quantitatively on synthetic and qualitatively on human data from the Human Connectome Project. We show that Laplacian regularization provides more accurate estimation of the signal and EAP based microstructural measures.

This work has been published in [19]

6.4.6. *Using 3D-SHORE and MAP-MRI to obtain both Tractography and Microstructural Contrasts from a Clinical DMRI Acquisition*

Participants: Rutger H.j Fick, Mario Zuccheli [Dpt of Computer Science, University of Verona], Gabriel Girard [SCIL Lab., Sherbrooke University], Maxime Descoteaux [SCIL Lab., Sherbrooke University], Gloria Menegaz [Dpt of Computer Science, University of Verona], Rachid Deriche.

Diffusion MRI (dMRI) is used to characterize the directional- ity and microstructural properties of brain white matter (WM) by measuring the diffusivity of water molecules. In clinical practice the number of dMRI samples that can be obtained is limited, and one often uses short scanning protocols that ac- quire just 32 to 64 different gradient directions using a single gradient strength (b-value). Such 'single shell' scanning protocols restrict one to use methods that have assumptions on the radial decay of the dMRI signal over different b-values, which introduces estimation biases. In this work, we show, that by simply spreading the same number of samples over multiple b-values (i.e. multi-shell) we can accurately estimate both the WM directionality using 3D-SHORE and characterize the radially dependent diffusion microstructure measures using MAP-MRI. We validate our approach by undersampling both noisy synthetic and human brain data of the Human Connectome Project, proving this approach is well-suited for clinical applications.

This work has been submitted to ISBI'2015.

6.4.7. *Laplacian-Regularized MAP-MRI Improving Axonal Caliber Estimation*

Participants: Rutger H.j Fick, Demian Wassermann, Gonzalo Sanguinetti, Rachid Deriche.

In diffusion MRI, the accurate description of the entire diffusion signal from sparse measurements is essential to enable the recovery of microstructural information of the white matter. The recent Mean Apparent Propagator (MAP)-MRI basis is especially well suited for this task, but the basis fitting becomes unreliable in the presence of noise. As a solution we propose a fast and robust analytic Laplacian regularization for MAP-MRI. Using both synthetic diffusion data and human data from the Human Connectome Project we show that (1) MAP-MRI has more accurate microstructure recovery com- pared to classical techniques, (2) regularized MAP-MRI has lower signal fitting errors compared to the unregularized approach and a positivity constraint on the EAP and (3) that our regularization improves axon radius recovery on human data.

This work has been submitted to ISBI'2015.

6.4.8. *A Unifying Framework for Spatial and Temporal Diffusion in Diffusion MRI*

Participants: Rutger H.j Fick, Demian Wassermann, Marco Pizzolato, Rachid Deriche.

We propose a novel framework to simultaneously represent the diffusion-weighted MRI (dMRI) signal over diffusion times, gradient strengths and gradient directions. Current frameworks such as the 3D Simple Harmonic Oscillator Reconstruction and Estimation basis (3D- SHORE) only represent the signal over the spatial domain, leaving the temporal dependency as a fixed parameter. However, microstructure- focused techniques such as Axcaliber and ActiveAx provide evidence of the importance of sampling the dMRI space over diffusion time. Up to now there exists no generalized framework that simultaneously models the dependence of the dMRI signal in space and time. We use a functional basis to fit the 3D+t spatio-temporal dMRI signal, similarly to the 3D-SHORE basis in three dimensional 'q-space'. The lowest order term in this expansion contains an isotropic diffusion tensor that characterizes the Gaussian displacement distribution, multiplied by a negative exponential. We regularize the signal fitting by minimizing the norm of the analytic Laplacian of the basis. The continuous 3D+t signal representation can provide new insights on the anomalous nature of the dMRI signal in human tissues, i.e., when mean-squared molecular displacements varies slower than linearly with the diffusion time. From the fitting one can also estimate the axon radius distribution parameters along any direction using approaches similar to AxCaliber. We validate our technique on synthetic data generated using the theoretical model proposed by Callaghan et al. We show that our method is robust to noise and can accurately describe the restricted spatio-temporal signal decay originating from tissue models such as cylindrical pores. Moreover, we apply our method on real data from an ActiveAx acquisition. Overall our approach allows to represent the complete 3D+t dMRI signal which should prove helpful in understanding normal and pathologic nervous tissue.

This work has been submitted to IPMI'2015.

6.4.9. *Fast and Robust EAP reconstruction using a Laplacian Regularized SHORE basis and its Impact on Microstructure Recovery*

Participants: Rutger H.j Fick, Demian Wassermann, Emmanuel Caruyer [SBIA, University of Pennsylvania Medical School], Rachid Deriche.

In diffusion MRI, the reconstructed Ensemble Average Propagator (EAP) from the diffusion signal provides detailed insights on the diffusion process and the underlying tissue microstructure. Recently, the 3D Simple Harmonic Oscillator based Reconstruction and Estimation (3D-SHORE) basis was proposed as a promising method to reconstruct the EAP. However, the fitting of the basis is sensitive to noise. To solve this we propose to use the Laplacian of the SHORE basis as a natural regularization functional. We provide the derivation of the Laplacian functional and compare its effect on EAP reconstruction with that of separated regularization of the radial and angular parts of the SHORE basis and imposing positive-definiteness in the estimation of the EAP. We validate our method on phantom data with known ground truth and on human data from the Human Connectome Project. We show that Laplacian regularization of the 3D-SHORE basis provides faster and more accurate estimation of the signal and EAP.

This work has been submitted to NeuroImage.

6.5. Functional and structural models analysis

6.5.1. *Analyzing Brain Plasticity in Math Learning Using Automated Dissection and Analysis of White Matter Tracts Through dMRI*

Participants: Dietsje Jolles [Stanford Medical School], Demian Wassermann, Ritika Chokhani [Stanford Medical School], Jennifer Richardson [Stanford Medical School], Caitlin Tenison [Stanford Medical School], Roland Bammer [Stanford Medical School], Lynn Fuchs [Vanderbit University], Kaustubh Supekar [Stanford Medical School], Vinod Menon [Stanford Medical School].

In a collaboration with Stanford Medical School, we explored longitudinal changes in white matter connectivity triggered by intensive math learning. Plasticity of white matter tracts is thought to be essential for cognitive development and academic skill acquisition in children. However, a dearth of high-quality diffusion tensor imaging (DTI) data measuring longitudinal changes with learning, as well as methodological difficulties in multi-time point tract identification have limited our ability to investigate plasticity of specific white matter tracts. With this contribution, we examined learning-related changes of white matter tracts innervating inferior parietal, prefrontal and temporal regions following an intense two-month math tutoring program. DTI data were acquired from 18 third grade children, both before and after tutoring. A novel fiber tracking algorithm based on a White Matter Query Language (WMQL) was used to identify three sections of the superior longitudinal fasciculus (SLF) linking frontal and parietal (SLF-FP), parietal and temporal (SLF-PT) and frontal and temporal (SLF-FT) cortices, from which we created child-specific probabilistic maps. The SLF-FP, SLF-FT, and SLF-PT tracts identified with the WMQL method were highly reliable across the two time points and showed close correspondence to tracts previously described in adults. Notably, individual differences in behavioral gains after two months of tutoring were specifically correlated with plasticity in the left SLF-FT tract. Our results extend previous findings of individual differences in white matter integrity, and provide important new insights into white matter plasticity related to math learning in childhood. More generally, our quantitative approach will be useful for future studies examining longitudinal changes in white matter integrity associated with cognitive skill development.

This work has been published in [14].

6.5.2. *Quantifying Uncertainty in Diffeomorphic Medical Landmark Registration*

Participants: Demian Wassermann, Matt Toew [Harvard Medical School - Brigham and Women's Hospital], Marc Niethammer [University of North Carolina at Chapel Hill], William Wells Iii [Harvard Medical School - Brigham and Women's Hospital, MIT].

In a collaboration with Harvard Medical School, the Brigham and Women's Hospital, MIT and the University of North Carolina at Chapel Hill, we proposed a novel mathematical framework to represent uncertainty in diffeomorphic registration techniques. Particularly, we introduced a novel mathematical framework for representing uncertainty in large deformation diffeomorphic image registration. The Bayesian posterior distribution over the deformations aligning a moving and a fixed image is approximated via a variational formulation. A stochastic differential equation (SDE) modeling the deformations as the evolution of a time-varying velocity field leads to a prior density over deformations in the form of a Gaussian process. This permits estimating the full posterior distribution in order to represent uncertainty, in contrast to methods in which the posterior is approximated via Monte Carlo sampling or maximized in maximum a-posteriori (MAP) estimation. The framework was demonstrated in the case of landmark-based image registration, including simulated data and annotated pre and intra-operative 3D images. This type of registration can be extended to several anatomical objects such as white matter tracts represented as streamlines.

This work has been published in [32].

6.5.3. *Group Comparisons on White Matter Tracts in Native Space*
Participants: Eleftherios Garyfallidis [University of Sherbrooke], Demian Wassermann, Maxime Descoteaux [University of Sherbrooke].

Let us suppose that we want to study specific fiber bundles in different subjects. The common approach would be to use a voxel-wise analyses which will warp scalar volumes in a common space, e.g. MNI space, and show how every subject differentiates from an average template. However, we know that with averaging and warping much of the specific information about the individual subjects' differences is lost. In this work, we provide a solution to this problem by using local streamline registration of specific bundles from different subjects. We show that with this new method we can keep track of the differences from every subject to every other subject in our group study.

This study was performed in collaboration with the SCIL lab of Sherbrook University within the framework of the Brain Connectivities Associate Team and published in [21].

6.5.4. *Perfusion Deconvolution via SHORE and Laplacian Regularization*
Participants: Marco Pizzolato, Auro Ghosh, Timothé Boutelier [Olea Medical, La Ciotat], Rachid Deriche.

Perfusion imaging comprehensively refers to the recovery of parameters of interest which are related to the passage of blood in the parenchyma (i.e. the functional part) of a tissue. The amount of perfusion is related to both the functionality of the parenchyma and its level of activity. By means of imaging techniques such as Dynamic Susceptibility Contrast MRI it is possible, in each voxel, to measure the tissue concentration $Ct(t)$ of a tracer injected before the scanning in the vascular system. According to the indicator dilution theory1 this is related to the concentration measured in an arterial region $Ca(t)$ described by a convolution with $R(t)$ that is the unknown residue function expressing the remaining time-dependent tracer quantity in the voxel. Historically $R(t)$ is obtained exploiting the convolution theorem $R(t) = FT - 1FT[Ct(t)]/FT[Ca(t)]$. However deconvolution is an ill-posed problem making this method very sensitive to noise. Many regularization techniques have been proposed but among all the most adopted technique is truncated Singular Value Decomposition, tSVD. However tSVD is known to underestimate an important perfusion parameter that is the blood flow BF, which can be computed as the maximum peak of the recovered $R(t)$. In this work we propose to use the Simple Harmonic Reconstruction and Estimation framework (SHORE) to estimate $R(t)$ in order to obtain a better parameter estimation. We regularize SHORE using Laplacian regularization. We compare the results with tSVD.

This work has been submitted to ISMRM 2015.

6.5.5. *Perfusion MRI Deconvolution with Delay Estimation and Non-negativity Constraints*
Participants: Marco Pizzolato, Auro Ghosh, Timothé Boutelier [Olea Medical, La Ciotat], Rachid Deriche.

Perfusion MRI deconvolution aims to recover the time-dependent residual amount of indicator (residue function) from the measured arterial and tissue concentration time-curves. The deconvolution is complicated by the presence of a time lag between the measured concentrations. Moreover the residue function must be non-negative and its shape may become non-monotonic due to dispersion phenomena. We introduce Modified Exponential Bases (MEB) to perform deconvolution. The MEB generalizes the previously proposed exponential approximation (EA) by taking into account the time lag and introducing non-negativity constraints for the recovered residue function also in the case of non-monotonic dispersed shapes, thus overcoming the limitation due to the non-increasing assumtion of the EA. The deconvolution problem is solved linearly. Quantitative comparisons with the widespread block-circulant Singular Value Decomposition show favorable results in recovering the residue function.

This work has been submitted to ISBI 2015.

6.6. Forward and Inverse Problems in MEEG

6.6.1. FindSource3D - Source Localization Using Rational Approximation on Plane Sections

Participants: Todor Jordanov [BESA GmbH, Germany], Jean-Paul Marmorat [École des Mines ParisTech, Sophia Antipolis], Maureen Clerc, Juliette Leblond, Andre Waelkens [BESA GmbH, Germany], Théodore Papadopoulo.

A new method for EEG source localization based on rational approximation techniques in the complex plane was suggested. The method is used in the context of a nested sphere head model, in combination with a cortical mapping procedure [51]. This method was shown to perform perfectly for numerical simulations without noise but its performance with respect to different signal-to-noise ratios (SNRs), to different number of sources and to real EEG data was not investigated until now. The method, formally called FindSource3D (FS3D), is evaluated with data simulations and a real EEG data set.

This work has been published in [40].

6.6.2. Diffusion Magnetic Resonance information as a regularization term for MEG/EEG inverse problem

Participants: Brahim Belaoucha, Anne-Charlotte Philippe, Maureen Clerc, Théodore Papadopoulo.

Several regularization terms are used to constrain the Magnetoencephalography (MEG) and the Electroencephalography (EEG) inverse problem. It has been shown that the brain can be divided into several regions with functional homogeneity inside each one of them. To locate these regions, we use the structural information coming from the diffusion Magnetic Resonance (dMRI) and more specifically, the anatomical connectivity of the distributed sources computed from dMRI. To investigate the importance of the dMRI in the source reconstruction, this work compares the solutions based on dMRI-based parcellation to random parcellation.

This work has been published in [37].

6.6.3. Dictionary learning for multitrial datasets

Participants: Maureen Clerc, Sebastian Hitziger, Théodore Papadopoulo.

Following the path opened with the Consensus matching Pursuit method (CMP) [48], we continue our endeavour to avoid signal averaging using directly the raw signal with the assumption that events of interest are those that repeat in each trial. Towards such a goal, and to improve the simple dictionary used in CMP, we have adapted dictionary learning methods to multitrial bio-electric signals, by explicitly implementing jitter invariance [62]. This allows for a much more detailed data-driven description of events. For example, using local field potential signals of chemically induced spikes (in a rat model), we have been able to distinguish several spike shapes which show some coherence in time. The method has been recently extended to detect spike events in continuous signals (i.e. not organized in epochs). While it requires a good signal to noise ratio, the method is very general and has also been used for various other signal types (see section 6.7).

This work has been published in [39].

6.7. Coupling functional and structural models

6.7.1. *Propagation of epileptic spikes revealed by diffusion-based constrained MEG source reconstruction*

Participants: Anne-Charlotte Philippe, Théodore Papadopoulo, Christian Bénar [Hospital "La Timone", Marseille], Jean-Michel Badier [Hospital "La Timone", Marseille], Maureen Clerc, Rachid Deriche.

In this work, we study the propagation of an epileptic spike (from single event data). As in the two previous sections, a cortex parcellation is performed using structural information coming from diffusion MRI Then, a MEG inverse problem is defined on a parcellated source space which imposes constant activity on each parcel. This inverse problem is applied separately for measurements obtained in a given time range. The most active parcels over the time range are located and their time course are displayed. This allowed the study of the propagation of an epileptic spike via those active parcels. Results on real data shows varying spatial propagations of an epileptic spike for the same subject.

This work has been published in [41].

6.7.2. *Using diffusion MRI information in the Maximum Entropy on Mean framework to solve MEG/EEG inverse problem*

Participants: Brahim Belaoucha, Jean-Marc Lina [Centre de Recherches Mathématique, Montréal], Maureen Clerc, Anne-Charlotte Philippe, Christophe Grova [McGill University], Théodore Papadopoulo.

Magnetoencephalography (MEG) and Electroencephalography (EEG) inverse problem is well-known to require regularization in order to avoid ill-posedness. Usually, regularization is based on mathematical criteria (minimum norm, ...). Physiologically, the brain is organized in functional parcels and imposing a certain homogeneity of the activity within these parcels was proven to be an efficient way to analyze the MEG/EEG data. The parcels information can be computed from diffusion Magnetic Resonances Imaging (dMRI) by grouping together source positions shared the same connectivity profile (computed as tractograms from diffusion images). In this work, three parcel-based inverse problem approaches have been tested. The first two approaches are based on minimum norm with added regularization terms to account for the parcel information. They differ by the use of a hard/soft constraint in the way they impose that the activity is constant within each parcel [74]. The third approach is based on the Maximum Entropy on Mean (MEM) framework [42]. It models source activity with a random variable and parcels are also used as a regularization. Several tests have been conducted with synthetic and real data that encompass the MEG/EEG and the diffusion magnetic resonance signals to compare these three approaches in terms of active region-detection accuracy.

This work has been published in [36].

6.8. Brain Computer Interfaces

6.8.1. *CoAdapt P300 speller: optimized flashing sequences and online learning*

Participants: Maureen Clerc, Théodore Papadopoulo, Loïc Mahé.

Our work in Brain Computer Interfaces was centered around the visual P300 speller system: a virtual keyboard allowing to type words by detecting the P300 wave, an automatic deflection of the central component of the electric potential, occurring approximately 300 ms after the apparition of an intermittent and rare event, on which the user's attention is focussing. The idea behind the P300 speller is very simple: the system displays series of stimuli (flashes), over the keyboard elements, and detects whether or not the EEG recorded after each flash contains a P300. Its implementation is not so simple, because of the low amplitude of the P300 compared to the background EEG, and of the inter-subject variability of this signal.

The advantage of this system is not to require any training on the part of the user. However the BCI system has to be trained to detect the P300 component from the background EEG: this is done through a calibration phase.

We developed a new method to reduce the calibration phase, with a transfer learning method called "mixture of experts" (MOE). The MOE classifier makes its predictions by averaging the decisions from a pre-recorded database of classifiers coming from other recording sessions (with other subjects) [31]. The decisions were made by using an evidence accumulation scheme, which updated the prediction at every flash of the keyboard [17].

Part of this work has been implemented in the software: CoAdapt P300 Stimulator.

6.8.2. P300 speller: clinical feasibility study with Amyotrophic Lateral Sclerosis

Participants: Maureen Clerc, Théodore Papadopoulo, Loïc Mahé, Asya Metelkina, Violaine Guy [Nice University Hospital], Claude Desnuelle [Nice University Hospital].

From September 2013 to July 2014, we were very involved in running an experiment with the Centre de Référence Maladies Neuromusculaires et SLA (CRMN/SLA) of Nice University Hospital. This study, partly funded by "Assocation pour la Recherche sur la Sclérose Latérale Amyotrophique", was conducted on 20 patients, who routinely come to be examined at the hospital. Each patient came for 3 sessions where he/she was allowed to use the P300 speller, after being equipped with an electro-encephalography cap, and watching a video explaining the modus operandi of the P300 and on their role in the study.

The P300 speller system has been organized in a way to make it relatively easy to deploy in a clinical setting: it involves only one laptop, and requires limited intervention from the caregiver. The most intricate operation is to position the EEG headset and ensure a correct impedance (below 5 kΩ) for all electrodes.

Each session consisted of three blocks, after the initial calibration phase: a copy spelling task of two ten-letter words, a free spelling task of approximately twenty characters, and an optional block of free use of the system for writing. Finally, the patient was asked to answer a questionnaire. This study intends to investigate the feasibility of setting up and using the P300 speller, from an operational point of view at the hospital. Translational studies of this type are extremely important for the adaptation the BCI systems to the target patient populations, and a large-scale usability study for the P300 speller has never been done before in France.

6.8.3. BCI Challenge: A spell on you if you cannot detect errors!

Participants: Maureen Clerc, Théodore Papadopoulo, Jérémie Mattout [Centre de Recherche en Neurosciences de Lyon, INSERM], Emmanuel Maby [Centre de Recherche en Neurosciences de Lyon, INSERM].

We have proposed an international BCI Challenge on decoding Error Potential signals. The winners will be announced at the 7th International IEEE/EMBS Conference on Neural Engineering, Montpellier, in April 2015. The Challenge was open on the Kaggle platform on Nov 14, 2014 and will close on Feb 24, 2015, see: website. At this date the competition is still open, and it has so far attracted 212 participants forming 181 teams.

In the P300 speller paradigm (see above) and in other BCI where a discrete feedback can be presented to the user, the EEG evoked response to the feedback can be recorded and processed online in order to evaluate whether the item selection was correct or not. This decision, if reliable, could then be used to improve the BCI performance by implementing some error correction strategy. In this competition, participants are asked to submit an Error Potential detection algorithm, capable of detecting the erroneous feedbacks online and to generalize across subjects (transfer learning).

The data used in this competition was acquired in the scope of the CoAdapt ANR project.

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

6. New Results

6.1. Highlights of the Year

- We reanalyzed the so-called Marginal Value Theorem (MVT), first published in 1976, in a paper published in Ecology Letters [23]. This theorem, also used in human behavior and economics, establishes how individuals should behave to optimize resource exploitation. Despite the thousands of papers written on the subject, we obtained the first mathematical characterization of how habitat characteristics affect the optimal foraging strategy. Mathematical foundations for this work were given in [24].

- The analysis of metabolic networks is generally made under the assumption (so called "balanced growth") that there is no internal accumulation of metabolites. However, this hypothesis is clearly wrong for microalgae, which store lipids and carbohydrates during the day and consume it during the night. A new formalism, called DRUM (Dynamic Reduction of Unbalanced Metabolism) was developed [16], assuming that the balanced growth is valid only in subnetworks, but that there can be accumulation between these modules (which often represent spatial distribution in the cell). This approach was successfully used to represent the dynamics of carbon accumulation in the microalgae *Tisochrysis lutea* under light/dark cycles, or in response to a nitrogen starvation. It also well described the diauxic heterotrophic growth of *Chlorella pyrernoidosa* [11].

6.2. Mathematical methods and methodological approach to biology

6.2.1. Mathematical analysis of biological models

6.2.1.1. Mathematical study of semi-discrete models
Participants: Jean-Luc Gouzé, Frédéric Grognard, Ludovic Mailleret, Pierre Bernhard, Elsa Rousseau, Nicolas Bajeux, Bapan Ghosh.

Semi-discrete models have shown their relevance in the modeling of biological phenomena whose nature presents abrupt changes over the course of their evolution [99]. We used such models and analyzed their properties in several practical situations that are developed in Section 6.3.2 , some of them requiring such a modeling to describe external perturbations of natural systems, and others to take seasonality into account. External perturbations of interacting populations occur when some individuals are introduced or removed from a natural system, which occurs frequently in pest control applications, either through the direct removal of pests [62], or through the introduction of biological control agents [45], [60], [54]. Seasonality is an important property of most agricultural systems in temperate environments since the year is divided into a cropping season and a 'winter' season, where the crop is absent, as in our analysis of eco-evolutionary dynamics of plant pathogens [25], [59]

6.2.1.2. Model reduction and sensitivity analysis
Participant: Suzanne Touzeau.

Dynamic models representing complex biological systems with numerous interactions can reach high dimensions and include complex nonlinearities. Especially if data are scarce, identifying the model parameters is then a challenge. So we designed an ad-hoc method based on global sensitivity analysis to simplify the model and determine the most influential parameters. It was applied to a within-host immunological model [30], [61]. This application was part of Natacha Go's PhD thesis, supervised by S. Touzeau and C. Belloc (BioepAR, INRA & Oniris Nantes) [90].

6.2.2. Metabolic and genomic models

Participants: Jean-Luc Gouzé, Madalena Chaves, Alfonso Carta, Ismail Belgacem, Olivier Bernard, Caroline Baroukh, Jean-Philippe Steyer, Diego de Pereda Sebastian, Francis Mairet.

6.2.2.1. Continuous models analysis

Transcription and translation models in bacteria We study detailed models of transcription and translation for genes in a bacterium, in particular the model of gene expression of RNA polymerase. With techniques of monotone systems, and time scale hypotheses, we can show the stability of the fast part of these systems, and reduce them to much smaller models [49], [48], [47]. We also study other models of the global cellular machinery. This is part of the PhD theses of Ismael Belgacem and Alfonso Carta [12], and done in collaboration with Inria IBIS project-team.

A model of synthesis of a virulence factor In collaboration with J.-A. Sepulchre (INLN Nice), we model the production of a virulence factor by a bacterium in a continuous stirred tank reactor. The production of this enzyme is genetically regulated, and degrades a polymeric external substrate into monomers.

Analysis and reduction of biochemical models In collaboration with D. Ropers (Inria IBIS project team), we address the problem of reduction of large biochemical networks, to decompose the dynamic behavior of the whole system into simpler models. This is the subject of the thesis of S. Casagranda.

Design of a bistable switch to control cellular uptake In joint work with Diego Oyarzún (Imperial College), we explore the idea of constructing a synthetic bistable system using an unbranched metabolic chain with a global enzyme regulator. Bistability can be achieved by choosing an appropriate pattern of regulation and deriving conditions on the promoter dynamic ranges to guarantee a bistable uptake flux. This work started during the visit of Diego to Biocore in October 2014.

Analysis of signaling pathways leading to apoptosis In joint work with Jérémie Roux (Marie Curie Fellow, IRCAN Nice), a cascade of signaling modules leading to apoptosis (or programmed cell death) was implemented and studied through simulations. The goal of this work is to determine whether, and at which stage in the pathway, the system may exhibit bistability. This was the work of Xiao Han's internship.

6.2.2.2. Hybrid models analysis

Piecewise quadratic systems for studying growth rate in bacteria The class of piecewise affine systems was extended to deal with dynamics dependent on dilution due to cell growth rate, leading to switched-piecewise quadratic systems [85]. These new systems use an expression for growth rate that may depend on any number of variables and have several quadratic modes. The behavior of piecewise quadratic systems introduces new features, notably regarding solutions at the thresholds when the vector fields are opposing: not only sliding mode solutions but also oscillatory behavior may happen. Part of this work is in the PhD thesis of Alfonso Carta [12].

Attractor computation using interconnected Boolean networks The method developed in [10] has been extended towards a better characterization of the attractors of the interconnected system in terms of invariant sets [26]. The method was used to test growth rate models in E. Coli using Boolean networks.

Analysis of circadian rhythms in cyanobacteria The model describing the system responsible for the circadian rhythm of cyanobacteria previously proposed in [86] has been improved in [50]. Here, we have tested the robustness of the circadian rhythm with respect to the perturbations inherent to the noisy environment of the cell, including cell growth and division. The interconnection between two models was studied: circadian rhythm and a stochastic model for cell division.

Structure estimation for Boolean models of gene regulation networks The problem of estimating Boolean models of gene networks from few and noisy measurements is addressed in [84], joint work with C. Breindl and F. Allgöwer from the University of Stuttgart. The class of unate or canalizing Boolean functions has been further considered and represented by multi-affine polynomials, leading to a reformulation of the estimation problem as a mixed integer linear program.

Structural principles for the existence of limit cycles in two-dimensional piecewise affine models Using concavity and continuity properties of Poincaré maps, we have derived some structural principles which link the topology of the transition graph to the existence, number and stability of limit cycles in a class of two-dimensional piecewise affine biological models of genetic networks [14].

6.2.2.3. Estimation and control

Optimal allocation of resources in a bacterium We study by techniques of optimal control the optimal allocation between metabolism and gene expression during growth of bacteria [52], in collaboration with Inria IBIS project-team.

Estimation of biological models In a joint work with Diego de Pereda (visiting PhD student), we studied observers and interval observers for models of glucose concentration in diabetes.

6.3. Fields of application

6.3.1. Bioenergy

6.3.1.1. Modelling of microalgae production

Participants: Olivier Bernard, Antoine Sciandra, Frédéric Grognard, Philipp Hartmann, Ghjuvan Grimaud, Quentin Béchet, David Demory, 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 [28] and support model development. These experiments have been carried out in the Lagrangian simulator, 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 signal perceived by a cell in a raceway pond [89], derived from hydrodynamical studies [55]. An electronic platform was developed to reproduce this high frequency light signal. The experiments show that the microalgae adapt their pigments to the average light that they have received [28].

The effect in the cell cycle of both the light periodic signal, the temperature and a nitrogen limitation were studied. The strong interactions between the different phases of the cell cycle through checkpoints was highlighted [106]. Temperature turned out to play a key role in modulating metabolic fluxes and synchronization.

The effect of cement flue gas on microalgae growth has been tested. It was demonstrated that this CO_2 source can be used to feed microalgal industrial cultures [114].

Finally a new methodology to measure cell viability has been set up. This approach is very promising to distinguish between net and gross growth rate [22].

These works have been carried out in collaboration with A. Talec, S. Rabouille, E. Pruvost and C. Combe (CNRS/UPMC -Oceanographic Laboratory of Villefranche-sur-Mer).

In collaboration with the IFREMER-PBA team (Nantes) we contributed to a study of the possible associations between microalgae and bacteria to enhance overall productivity [98].

Metabolism of carbon storage and lipid production

A macroscopic model for lipid production by oleaginous microalgae [7] has been previously proposed. This model describes the accumulation of neutral lipids (which can be turned into biofuel), carbohydrates and structural carbon. A metabolic model has been set up and validated for the microalgae *Isochrysis luthea*. It predicts carbohydrate and lipid accumulation, under conditions of light/dark cycles and/or nitrogen deprivation [78], [88], [16].

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 [83]. 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 [55].

Modeling the photosynthesis response to fast fluctuating light

The impact of the hydrodynamics on the light perceived by a single cell was studied thanks to fluid dynamics simulations of a raceway pond [34]. 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 [96]. Single cell trajectories were simulated by this software, and the effect on photosynthesis efficiency was assessed using models of photosynthesis [95]. These results were compared to experimental measurements where the high frequency light was reproduced [89].

We also developed a model to reproduce the fluorescence of microalgae during a PAM protocol [51]

Modeling microalgae production processes

The integration of different models developed in the group [81], [101], [7] 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. 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 [82], [111] 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]. This model is used to predict lipid production in the perspective of large scale biofuel production. It was also used to assess the potential of France for microalgae, when taking into account the actual 2012 meteorology at the scale of France the use of lands, slope, proximity of nutrients and CO_2 [73].

In the framework of the ANR project Purple Sun, we develop an innovative system for microalgae production: a raceway pond under a greenhouse with semi-transparent photovoltaic panels. To this end, we include in the microalgae model the effect of light wavelength, and we develop a thermic model of the system in order to estimate the culture temperature.

Finally, we provide guidelines for the design of experiments with high informative content that allows an accurate estimation of the parameters concerning the effect of temperature and light on microalgae growth. The optimal experiment design problem was solved as an optimal control problem. E-optimal experiments were obtained by using two discretization approaches namely sequential and simultaneous. Simulation results showed the relevance of determining optimal experimental inputs for achieving an accurate parameter estimation [39].

Nitrogen fixation by nitrogenotrophs

The fixation of nitrogen by *Croccosphera watsonii* was represented with a macro metabolic model [92]. 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-Oceanographic Laboratory of Villefranche-sur-Mer).

Modeling thermal adaptation in microalgae

We have used the Adaptive Dynamics theory to understand how temperature drives evolution in microalgae. For a constant temperature, we have shown that the optimal temperature trait tends to equal the environment temperature. We then study the case where the temperature is periodically fluctuating [53]. We now use this method at the scale of the global ocean, validating our approach with experimental data sets from 194 species.

Including phytoplankton photoadaptation into biogeochemical models

The complexity of the marine ecosystem models and the representation of biological processes, such as photoadaptation, is very challenging to tackle so that their representation remains 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 [77], [15].

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

6.3.1.2. Control and Optimization of microalgae production

On-line monitoring

Interval observers give an interval estimation of the state variables, provided that intervals for the unknown quantities (initial conditions, parameters, inputs) are known [91]. Several developments were carried out in this direction to improve the design and performances of interval observers, and accounting for a specific structure (*i.e.* triangular) or property (*i.e.* Input to State Stable), [38]. Interval observers were designed for the estimation of the microalgae growth and lipid production within a production process [37] and validated experimentally [36].

Optimization of the bioenergy production systems

Based on simple microalgae models, analytical optimization strategies were proposed. We first focused on the optimal operating conditions for the biomass productivity under day/night cycles using Pontryagin's maximum principle (assuming a periodic working mode) [32].
On the other hand, we assessed strategies for optimal operation in continuous mode using the detailed model for raceways [108], [109]. Two strategies were developed. The first one consists 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 translated 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 attain biomass productivities very near to the maximal 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.
We also propose a nonlinear adaptive controller for light-limited microalgae culture, which regulates the light absorption factor (defined by the ratio between the incident light and the light at the bottom of the reactor). We show by numerical simulation that this adaptive controller can be used to obtain near optimal productivity under day-night cycles [103].

Interactions between species

Large scale culture of microalgae for bioenergy involves a large biodiversity (different mutants, invasion, growth-promoting bacteria [98]...). Control of such systems requires to consider the interactions between the different species. Such systems involve bacteria and microalgae, and the competition between these organisms can have several equilibrium points, which can be studied with Monod, Contois and Droop type models [33].

In the framework of the ANR Facteur 4 project, we propose to drive this competition exploring different strategies in order to select species of interest.

We have proposed an adaptive controller which regulates the light at the bottom of the reactor [104]. When applied for a culture with n species, the control law allows the selection of the strain with the maximum growth rate for a given range of light intensity. This is of particular interest for optimizing biomass production as species adapted to high light levels (with low photoinhibition) can be selected.

Strategies to improve the temperature response have been proposed. First we modeled the adaptive dynamics for a population submitted to a variable temperature [53]. This was then used to design experiments aiming at enlarging the thermal niche of a species. Experiments with periodic temperature stresses are currently carried out at the LOV.

Finally, in a more theoretical framework, we studied how to select as fast as possible a given species in a chemostat with two species at the initial instant. Using the Pontryagin maximum principle, we have shown that the optimal strategy is to maintain the substrate concentration to the value maximizing the difference between the growth rates of two species [17]. We now try to extend this result for n species with mutations.

6.3.2. *Design of ecologically friendly plant production systems*

6.3.2.1. *Controlling plant pests*
Participants: Frédéric Grognard, Ludovic Mailleret, Suzanne Touzeau, Nicolas Bajeux, Bapan Ghosh.

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 [6] [110], unveiling the crucial influence of within-predator density dependent processes. Since parasitoids may be more prone to exhibit positive density dependent dynamics rather than negative ones, which are prevalent among predatory biocontrol agents,the current modeling effort consists in studying the impact of positive predator-predator interactions on the optimal introduction strategies (PhD of Nicolas Bajeux, [45]). The influence of the spatial structure of the environment on biological control efficacy has also been investigated; first results indicate that spatial structure tends to influence it in quite a same way as intra-specific competition does [60]. An extension of that modeling framework was also studied, that considered state dependent impulsive feedback for the stabilization of a positive equilibrium [54].

Connected research on the influence of space on the establishment capacities of biological control agents is also being pursued both through computer simulations and laboratory experiments on parasitoids of the genus *Trichogramma*. This is the topic of the PhD thesis of Thibaut Morel Journel (UMR ISA); in particular, we show how landscape connectivity or spatial heterogeneity shape establishment dynamics in spatially structured environments [63], [64], [65].

Plant compensation, pest control and plant-pest dynamics

Introducing a plant compartment into our models, we first focused on plant-insect interactions and showed how the level and timing of the pest invasion and pests control interventions could have important effects on the plant's growth pattern and its final biomass. We then modeled plant compensation, which is the process by which some plants respond positively to recover from the effects of pest injury. We have shown that depending on plants and pests characteristics, as well as the level of pest attack, plant overcompensation may or may not happen [35]. Experiments have then been held at UMR ISA on tomato plants facing tutta absoluta invasion; tendencies to compensation have been evidenced, but need to be confirmed through larger scale experiments.

This work is part of the PhD thesis of Audrey Lebon (Cirad), supervised in collaboration with Yves Dumont (Cirad), which has been defended in December 2014.

6.3.2.2. *Controlling plant pathogens*
Participants: Frédéric Grognard, Ludovic Mailleret, Suzanne Touzeau, Elsa Rousseau.

Sustainable management of plant resistance

Because in addition to being eaten, plants can also get sick, we studied other forms of biological control dedicated to fight plant pathogens. One such method is the introduction of plant strains that are resistant to one pathogen. This often leads to the appearance of virulent pathogenic strains that are capable of infecting the resistant plants. It is therefore necessary to find ways to protect the durability of such resistances, which are a natural exhaustible resource. Experiments were conducted in INRA Avignon, followed by high-throughput sequencing (HTS) to identify the dynamics of several virus strains in competition within host plants. Different plant genotypes were chosen for their contrasted effects on genetic drift and selection they induce on virus populations. Those two evolutionary forces can play a substantial role on the durability of plant resistance. Therefore we fitted a mechanistic-statistical model to these HTS data in order to disentangle the relative role of genetic drift and selection during within-host virus evolution [68], [67]. This is the topic of Elsa Rousseau's PhD thesis, and is done in collaboration with Frédéric Fabre and Benoit Moury (INRA Avignon).

We also represented the pathogen spread in agricultural landscapes [40]. At this scale, we looked at how the landscape structure facilitates or impedes the disease spread among host patches. We showed that, when deploying a host with complete resistance to the pathogen along with a susceptible host, mixed landscapes were always more efficient to hamper the disease spread. However, when using a quantitatively resistant host, aggregating the hosts in different regions could result in a better control of the pathogen spread [41]. This work is part of Julien Papaïx's PhD thesis (MIA, INRA Jouy-en-Josas & BIOGER, INRA Grignon).

Eco-evolutionary dynamics of plant pathogens in seasonal environments

Understanding better pathogen evolution also requires to understand how closely related plant parasites may coexist. Indeed, such coexistence is widespread and is hardly explained through resource specialization. We showed that, in agricultural systems in temperate environments, the seasonal character of agrosystems is an important force promoting evolutionary diversification of plant pathogens [94]. Plant parasites reproduction mode may also 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 [25].

This work was part of the PhD thesis of Magda Castel (Agrocampus Ouest) and has been done in collaboration with Frédéric Hamelin (Agrocampus Ouest), Didier Andrivon (INRA Rennes) and Virginie Ravigné (CIRAD Montpellier).

6.3.3. Biological depollution

6.3.3.1. Control and optimization of bioprocesses for depollution

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

We have considered the problem of global stabilization of an unstable bioreactor model (e.g. for anaerobic digestion), when the measurements are discrete and in finite number ("quantized"). These measurements define regions in the state space, wherein a constant dilution rate is applied. We show that this quantized control may lead to global stabilization: trajectories have to follow some transitions between the regions, until the final region where they converge toward the reference equilibrium [71].

Although bioprocesses involve an important biodiversity, the design of bioprocess control laws are generally based on single-species models. In [56], we have proposed to define and study the multispecies robustness of bioprocess control laws: given a control law designed for one species, what happens when two or more species are present? We have illustrated our approach with a control law which regulates substrate concentration using measurement of growth activity. Depending on the properties of the additional species, the control law can lead to the correct objective, but also to an undesired monospecies equilibrium point, coexistence, or even a failure point. We now start to develop control laws more robust to the presence of additional species.

Moreno [107] has proposed an optimal strategy for fed-batch bioreactor with substrate inhibition. Thanks to the Pontryagin maximum principle and the Hamilton-Jacobi equation, we have shown that the same strategy is still optimal when mortality is included in the model [79]. We have also studied the problem when the singular arc is non-necessarily admissible everywhere (i.e. the singular control can take values outside the admissible control set). We have pointed out the existence of a frame point on the singular arc above which any singular

trajectory is not globally optimal. Moreover, we have provided an explicit way for computing numerically the switching curves and the frame point [46], [19].

6.3.3.2. Coupling microalgae to anaerobic digestion
Participants: Olivier Bernard, Antoine Sciandra, Jean-Philippe Steyer, Frédéric Grognard, Philipp Hartmann, Francis Mairet.

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 Phycover project is aiming at evaluating the potential of this process [113], [112].

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 [100] . On the other hand, we have tested the ability of ADM1 [115] (a reference model which considers 19 biochemical reactions) to represent the same dataset. This model, after modification of the hydrolysis step [102] has then been used to evaluate process performances (methane yield, productivity...) and stability though numerical simulations.

Finally, we have proposed and analysed a three dimensional model which represents the coupling of a culture of microalgae limited by light and an anaerobic digester. We first prove the existence and attraction of periodic solutions. Applying Pontryagin's Maximum Principle, we have characterized optimal controls, including the computation of singular controls, in order to maximize methane production. Finally, we determine numerically optimal trajectories by direct and indirect methods [18].

6.3.3.3. Life Cycle Assessment
Participants: Olivier Bernard, Jean-Philippe Steyer.

This work is the result of a collaboration with Arnaud Helias of INRA-LBE and Pierre Collet (IFPEN).

In the sequel of the pioneering life cycle assessment (LCA) work of [97], we continued to identify the obstacles and limitations which should receive specific research efforts to make microalgae production environmentally sustainable.
The improvements due to technological breakthrough (leading to higher productivities) have been compared to the source of electricity. It turns out that the overall environmental balance can much more easily be improved when renewable electricity is produced on the plant [27]. As a consequence, a new paradigm to transform solar energy (in the large) into transportation biofuel is proposed, including a simultaneous energy production stage. This motivated the design of the purple sun ANR-project where electricity is produced by semi transparent photovoltaic panels [74] under which microalgae are growing.

These studies have allowed to identify the obstacles and limitations which should receive specific research efforts to make this process environmentally sustainable [93].

Finally, some works are aiming at normalizing LCA for microalgae and proposing guidelines to make the LCA more easily comparable [87].

These works have been carried out in collaboration with E. Latrille and B. Sialve (INRA - Laboratory of Environmental Biotechnology, Narbonne).

6.3.4. Models of ecosystems

6.3.4.1. Optimality/games in population dynamics
Participants: Frédéric Grognard, Ludovic Mailleret, Pierre Bernhard.

Optimal foraging and residence times variations

In a pair of papers [23], [24], we reanalyzed the so-called Marginal Value Theorem (MVT), first published in 1976. This theorem, also used in human behavior and economics, establishes how individuals should behave to optimize resource exploitation. This result has been has been routinely applied in ecology to understand the foraging strategy of animals such as insect parasitoids used for biological control purposes. We obtained the first mathematical characterization of how habitat characteristics (e.g. patch quality, or the distance between resource patches) affect the optimal foraging strategy. This allowed to confirm or refine MVT predictions, and to provide new predictions in the more realistic case of heterogeneous habitats. Some counterintuitive predictions emerged: making resource patches richer can actually make individuals move more rapidly, contradicting generally admitted earlier predictions.

This work was conducted with Vincent Calcagno (UMR ISA) and Frédéric Hamelin (Agrocampus Ouest).

The handicap paradox

We have continued our investigation of the handicap paradox of sexual selection with the tools of signaling theory. Zahavi's handicap principle, and our game theoretic analysis, explain why an equilibrium displays the "handicap" feature [21]. However, the explanation seems somewhat contrived, so the next question is "how could evolution have reached such a state ?" We have investigated that question with the tools of adaptive dynamics, and reached the conclusion that, if one accepts adaptive dynamics as a model of evolution, and our model of sexual selection, the equilibrium described in our previous article is indeed the limit state of evolution [20].

This work was conducted with Frédéric Hamelin (Agrocampus Ouest).

CASTOR Project-Team

6. New Results

6.1. Physics

6.1.1. Physical studies

6.1.1.1. Parallel Kelvin-Helmholtz instability
Participants: Hervé Guillard, Marco Bilanceri, Céline Colin [CEA], Philippe Ghendrih [CEA], Giorgio Giorgiani, Boniface Nkonga, Frédéric Schwander [M2P2, AMU], Eric Serre [M2P2, AMU], Patrick Tamain [CEA].

In the scrape-off layer (SOL) of tokamaks, the flow acceleration due to the presence of limiter or divertor plates rises the plasma velocity in a sonic regime. These high velocities imply the presence of a strong shear between the SOL and the core of the plasma that can possibly trigger some parallel shear flow instability. The existence of these instabilities, denoted as parallel Kelvin-Helmholtz instability in some works have been investigated theoretically in [51] using a minimal model of electrostatic turbulence composed of a mass density and parallel velocity equations. This work showed that the edge plasma around limiters might indeed be unstable to this type of parallel shear flow instabilities. In this work, begun in 2013, we have performed large scale 3D simulations using the PlaTo platform of the same simple mathematical model to investigate this question. The numerical results confirm that in agreement with the theoretical expectations as well as with other numerical methods, the sheared flows in the SOL are subject to parallel Kelvin-Helmholtz instabilities. However, the growth rate of these instabilities is low and these computations require both a sufficient spatial resolution and a long simulation time. This makes the simulation of parallel Kelvin-Helmholtz instabilities a demanding benchmark but it also allows us to validate the parallel implementation of the PlaTo platform up to up to $\mathcal{O}(1000)$ CPU [14].

6.2. Numerical developments

6.2.1. Numerical developments

6.2.1.1. Conformal hexahedral mesh coarsening by agglomeration
Participants: Hervé Guillard, Pierre Cargemel, Youssef Mesri [IFPEN].

This work has been realized in the framework of a PhD contract with IFPEN that aims to produce a coarsening software for hex-dominant meshes. Reservoir simulation involves to compute dynamic flow of different phases in a porous medium. The initial state of the reservoir is usually pre-computed via geo-statistics methods extrapolating measures of the terrain. Therefore, the input of reservoir simulation is given as a very fine mesh containing heterogeneous data and numerical simulation on this fine mesh is usually non-practical. This work is therefore devoted to the study of an agglomeration strategy, to dynamically coarsen this fine hex-dominant mesh. The adaptivity may be driven by physics and/or geometric estimators. Ideally, the coarsening should be applied locally in low gradient regions, whereas high gradient regions keep the fine mesh. This work has been presented in the 23rd International meshing Roundtable [26]. The planned sequel of this work consists to use the notion of Central Voronoi Tesselation (CVT) to treat the regions where the mesh is not structured and to apply this strategy in different physical contexts from plasma physics to petroleum engineering.

6.2.1.2. Mapped Fourier Methods for stiff problems in toroidal geometry
Participant: Hervé Guillard.

Due to the particular geometry of tokamaks, a lot of numerical codes developed for their numerical simulations use Fourier methods. Fourier spectral or pseudo-spectral methods are extremely efficient for periodic problems. However this efficiency is lost if the solutions have zones of rapid variations or internal layers. For these cases, a large number of Fourier modes are required and this makes the Fourier method unpractical in many cases. This work investigates the use of mapped Fourier method as a way to circumvent this problem. Mapped Fourier method uses instead of the usual Fourier interpolant the composition of the Fourier interpolant with a mapping in such a way that in the computational space, the functions to represent are not stiff. This work gives some examples of the usefulness of this method and apply it to a simple model of pellet injection in tokamaks as an example of its potential interest for complex multi dimensional problem [34].

6.2.1.3. Multislope MUSCL method for general unstructured meshes
Participants: Hervé Guillard, Clément Le touze [ONERA], Angelo Murrone [ONERA].

To increase the accuracy in finite volume method, the concept of MUSCL reconstruction has been introduced in the pioneering work of van Leer in the 70'. This technique is still one of the most efficient to deal with the existence of discontinuous solutions in numerical simulations. In the MUSCL technique, a discontinuous linear approximation of the solution is reconstructed on each control volume. The main approximation problem of this method is therefore to reconstruct the slope of the solution.

The multislope concept has been recently introduced in the literature to deal with MUSCL reconstructions on triangular and tetrahedral unstructured meshes in the finite volume cell-centered context. Dedicated scalar slopes are used to compute the interpolations on each face of a given element, in opposition to the monoslope methods in which a unique limited gradient is used. The multislope approach reveals less expensive and potentially more accurate than the classical gradient techniques. Besides, it may also help the robustness when dealing with hyperbolic systems involving complex solutions, with large discontinuities and high density ratios. In this work, we have designed a generalized multislope MUSCL method for cell-centered finite volume discretizations. The method is freed from constraints on the mesh topology, thereby operating on completely general unstructured meshes. Moreover optimal second-order accuracy is reached at the faces centroids and the scheme is L^∞ stable. Special attention has also been paid to equip the reconstruction procedure with well-adapted dedicated limiters, potentially CFL-dependent. We have shown in [18] the ability of the method to deal with completely general meshes, while exhibiting second-order accuracy.

6.2.1.4. Development of a two temperature model
Participants: Hervé Guillard, Afeintou Sangam, Elise Estibals.

A two temperature (ions - electrons) model for non-magnetized plasma has been designed. The numerical scheme is a finite volume method with an approximate Riemann solver using the total energy equation and the electron entropy as main variables. This Riemann solver has been validated against standard shock tube problems and incorporated in the PlaTo platform. The solver has been implemented in toroidal geometry and tested successfully on realistic particular flows encountered in this context. The development of a reduced MHD model based on this two temperature scheme is currently studied.

6.2.1.5. Entropy Preserving Schemes for Conservation Laws
Participants: Christophe Berthon [University of Nantes], Bruno Dubroca [CEA/DAM/CESTA and University of Bordeaux 1], Afeintou Sangam.

A relaxation-type scheme has been proposed to approximate weak solutions of Ten-Moments equations with source terms [**2**]. These equations model compressible anisotropic flows. Following the technique introduced in [44], the proposed scheme is proved to be entropy preserving.

6.2.1.6. Eurofusion WPCD: Free boundary equilibrium code and control
Participants: Cédric Boulbe, Blaise Faugeras, Jean François Artaud [IRFM CEA Cadarache], Vincent Basiuk [IRFM CEA Cadarache], Emiliano Fable [Max-Planck-Institut für Plasmaphysik, Garching], Philippe Huyn [IRFM CEA Cadarache], Eric Nardon [IRFM CEA Cadarache], Jakub Urban [IPP, Academy of Sciences of the Czech Republic, Prague].

Our team is involved in the integrated modelling WPCD (Work Package Code Development) Eurofusion. This project is the continuation of the EFDA-ITM project. The goal of WPCD is to provide a european tool for tokamak simulations. Different physical codes can be coupled using Kepler environment. Machine description and physical data have been described using CPO (Consistent Physical Objet) which are used as standardized inputs and outputs for the codes.

In this project, we participate in the coupling of a free boundary equilibrium solver, the European Transport Solver (ETS) and a plasma shape and position controller. The workflow coupling TCV hybrid Simulink controller and Cedres++ using PF circuit connections has been finalized and tested on the TCV tokamak.
A new workflow coupling Cedres++ with ETS and the TCV controller has been developed and is being tested on a TCV test case. This workflow is an evolution of the coupling CEDRES++ - ETS described in [12].
A successful benchmark between the three free boundary equilibrium codes CEDRES++([15] [47]) , FREE-BIE ([43]), and SPIDER ([49]) has been done on static test cases. This activity will be continued to compare the time dependent versions of the three codes.

6.2.1.7. Optimal control for scenario optimization of discharges in tokamaks

Participants: Jacques Blum, Holger Heumann.

In this project we aim for an automatic determination of optimal voltage evolutions via an optimal control formulation based on a system of partial differential equations that describes the evolution of plasma equilibrium in a tokamak. Optimal voltage evolutions are the one that ensure that the evolution of the plasma runs through predescribed, user-defined states, defined e.g. as desired evolution of shape or position. The system of partial differential equations describing the evolution of the plasma is non-linear and we use a finite element formulation together with implicit time stepping for the discretization [15]. With this approach we end up with a large but finite dimensional optimization problem with non-linear constraints. We are using SQP (sequential quadratic programming), known to be one of the fastest methods of such problems, to solve the finite-dimensional optimization problem. The performance of SQP relies on accurate derivatives of the objective function and the constraints. The derivatives related to the free boundary, derived and implements during H. Heumann's PostDoc 2011/2012 for a static optimal control problem, appeared here again and are one of the important building blocks for treating the transient case. Both in CEDRES++ and FEEQ.M we have now the capability to solve first test cases to define optimal voltage evolutions. In contrast to the static case, where the linear algebraic systems in the SQP iteration remain reasonable small, the solution of the corresponding linear system in the transient case becomes very time-consuming, which somehow limits the applicability. We are testing variants of SQP, such as BFGS-like updates for the reduced Hessian, to see whether we could speed up and improve robustness of our calculations. Fast iterative solver for large sparse linear systems is another option that we started to investigate. Fast iterative solver for linear system in transient optimal control problems governed by partial differential equations is a very active area of research and we hope to benefit from the latest developments.

6.2.1.8. Boundary reconstruction for the WEST tokamak with VacTH

Participants: Jacques Blum, Sylvain Bremond [IRFM CEA Cadarache], Cédric Boulbe, Blaise Faugeras, Holger Heumann, Philippe Moreau [IRFM CEA Cadarache], Eric Nardon [IRFM CEA Cadarache], Remy Nouailletas [IRFM CEA Cadarache], François Saint Laurent [IRFM CEA Cadarache].

This work is under progress in collaboration with the CEA. The control of the plasma in the future WEST tokamak requires the identification of its boundary in real time during a pulse. The code VacTH under development in the team enables such an identification. Several numerical developments and experiments have been conducted in order to prepare the control of the plasma in the WEST tokamak. The equilibrium code CEDRES++, also developed in the team, is used to simulate a real plasma and to generate synthetic magnetic measurements from which the plasma boundary is reconstructed using the code VacTH. A control algorithm developed by the colleagues from the CEA then uses this knowledge of the plasma shape to adapt the currents flowing in the poloidal field coils in order to achieve a desired evolution of the plasma.

6.2.1.9. Equilibrium reconstruction for ASDEX UpGrade (AUG) with Vacth-Equinox
Participants: Blaise Faugeras, Rui Cohelo [IPFN, IST, Lisbonne], Patrick Mccarthy [National University of Ireland University College Cork].

Within the framework of the WPCD EUROFUSION the code VacTH-Equinox has been adapted to enable equilibrium reconstruction for AUG. The identification of the current density pedestal required the development of a specific regularization scheme allowing weaker regularization close to the plasma boundary and stronger close to the magnetic axis.

6.2.1.10. Taylor-Galerkin stabilized Finite Element
Participants: José Costa, Boniface Nkonga.

The theoretical part of Taylor-Galerkin/Variational multi-scales (TG/VMS) strategy applied to MHD and reduced MHD modeling has been achieved last year. The final method amounts to add in the finite element formulation, a self-adjoint operator associated to the most critical hyperbolic component of the system to be solved. The design of the critical contours and the identification of associated waves to be stabilized is problem dependent and related to the Jacobian matrix. We have focused this year on the validations of this strategy and the improvement of the linearization used for stabilization. For application to plasma configurations with X-point, we need to reconsider the consistency with equilibrium and the Bohm boundary conditions on open flux walls.

6.2.1.11. Toward full MHD numerical modeling with $\mathcal{C}^1$ FE
Participants: José Costa, Giorgio Giorgiani, Hervé Guillard, Boniface Nkonga.

In this context the single fluid full MHD model is considered and the divergence free constraint on the magnetic field is achieved by introduction of a potential vector. The use of the potential vector has the additional advantage that the toroidal component is the magnetic flux of the Grad-Shafranov equilibrium. However, using the potential vector as variable introduces higher order derivatives in the system and classical C0 finite elements cannot be directly applied. This is why our finite element strategies use shape/test functions whose derivatives have global continuity in space (smooth finite elements). The global approach uses cross product shape/test functions between poloidal(2D) and toroidal(1D). In the 2D poloidal plane, discretization uses either quadrangular or triangular elements.

This year we have focused on the numerical analysis associated to the full MHD discretization in configurations with open flux surfaces. In order to derive efficient strategies for the full MHD in the potential vector formulation, the Gauge condition on the potential vector and the boundary conditions have been enforced by penalizations. For the Gauge condition it gives rise to element contributions but also boundary integrals that should be computed on curved surfaces that sometime fitted the magnetic surfaces. Equations are formulated in semi-conservative form such as to apply integration by part. Therefore, boundary conditions can be viewed as evolution of fluxes or variables. Integral formulation on the boundary is very useful for higher order finite elements and also for easier treatment of corners. Indeed in this context the boundary conditions are edges/surfaces oriented and boundary corners are driven by the neighborhood edges penalizations. This strategy is the one that will be used for future developments.

2D Quadrangular Cubic Bezier Finite Elements: This finite element is used for a while for reduced MHD models in the software Jorek. Reduced MHD is used to project the momentum equation in a space orthogonal to the equilibrium. When full MHD models are used, the momentum equation needs to be projected in the equilibrium space and this projection should be consistent with the Grag-Shafranov equilibrium that is used to compute the initial state. This has been achieved by a proper computation of the JxB contribution in the momentum equation, taking into account the poloidal variation of the toroidal component of the magnetic field. Detailed analysis has been performed. The next year will be devoted to implementations and numerical validations.

2D Triangular Powel-Sabin Finite Elements: In order to avoid some mesh singularities when using quadrangular meshes for complex geometries and flux surfaces shapes, triangular elements are a possible option. It is not so easy to derive smooth finite element on triangle with reduced number of degree of freedom. The Bell reduced-quintic finite elements we have considered in the previous years have too much unknowns (6 per vertex). Powell-Sabin splines are piecewise quadratic polynomials with a global $\mathcal{C}^1$-continuity and 3 unknowns per vertex, they have a local support, they form a convex partition of unity, they are stable, and they have a geometrically intuitive interpretation involving control triangles. Construction of the Powel-Sabin splines needs some geometrical tools that have been developed: Minimum area enclosing triangle of a set of control points (sequential and parallel). This construction is applied to each vertex of the triangular mesh and used to derive the local shape/test functions. These Powel-Sabin splines have been used successfully in the area of computer aided geometric design for the modeling and fitting of surfaces. We have used the Powell-Sabin (PS) splines for the approximation of elliptic partial differential equations (including Grad-Shafranov) in a rectangular domain. In this context have recovered the optimal rate of convergence (order 3). Preliminary result has been obtained for hyperbolic isothermal 2D Euler equations with TG/VMS stabilization. Our aim in the coming years is to apply these PS splines to full MHD in a toroidal geometry.

6.2.1.12. *Genuinely multidimensional Riemann Solver*
Participants: Jeaniffer Vides, Boniface Nkonga.

Multidimensional Riemann solvers were pioneered by Abgrall. Abgrall, Maire, Nkonga, Després and Loubere have extensively developed them especially as node-solvers for Lagrangian hydrodynamics. Another strain of work comes from explorations by Wendroff and Balsara who took a space-time approach. In this work, the resolved state is obtained via space-time integration over a wave model, just as was done by Wendroff and Balsara. However, an algebraic approach is used for the development of the fluxes. It is, therefore, shown that the multidimensional fluxes can be obtained by application of jump conditions at the boundaries of the wave model. The problem is of course over determined with the result that the shock jump conditions are only satisfied approximately in a least squares sense. Even so, this work gives us new perspective on multidimensional Riemann solvers. The litteral satisfaction of the shock jump conditions (up to least squares approximation) makes it easier to understand multidimensional Riemann solvers as a natural extension of the one-dimensional Riemann solvers. Contributions have also been made on the development of a minimalist wave model, which might help in reducing dissipation. Further innovations are reported on the assembling of fluxes based on the structure of the wave model, and those innovations are potentially useful. For MHD the CT approach consists of constraining the transport of magnetic field so that the divergence is always kept zero. The method relies on exploiting the dualism between the flux components and the electric field. Since the electric field is needed at the edges of the mesh, the multidimensional Riemann solver can also provide the electric field. By running an extensive set of simulations, it is shown that the multidimensional Riemann solver is robust and can be used to obtain divergence-free formulations for MHD that perform well on several stringent calculations. Future work will improve this strategy by enriching the description of the strongly interaction of waves.

6.2.1.13. *Multi scales approximations of "Shallow water" flows*
Participants: Jeaniffer Vides, Boniface Nkonga.

The terminology "Shallow water" is used to characterize thin flows on curved surfaces. It is customary for this type of flows to use the incompressible Navier-Stokes equations to asymptotically derive reduced models for the evolution of the depth integrated speed and the thickness of the flow. Reduced model are mainly hyperbolic and finite volume method are often used for their numerical approximation. Approximations strategies are generally structured as follow:

- Construction of a global coordinate system associated with an assumed analytical surface.
- Reduction of the model relatively to the global coordinate system
- Approximation of the surface by a finite number of elements.
- Approximation of the reduced model using the discrete surface.

In the context of real applications, it is presumptuous to expect an analytical formulation of the surface. From the data provided by observation satellites, we can usually extract a discrete description of the surfaces that drives thin flow. Therefore, it is more practical to use the discrete description as the starting point of the resolution strategy. This is the angle of approach that we have considered. We locally define two mesh scales: the element scale and the cell scale. The discrete mapping and the reduced model are defined at the element scale and the average values that evolve in time are defined at the cell scale. First applications have been successfully performed. We will now continue your investigations and include relevant physics at each scale, including sheared flows. We will also examine the use of multi dimensional Riemann solver in this context.

6.2.1.14. Computational Magnetohydrodynamics with Discrete Differential Forms

Participants: Holger Heumann, Ralf Hiptmair [SAM, ETH Zürich, Switzerland], Cecilia Pagliantini [SAM, ETH Zürich, Switzerland].

Differential forms, or equivalently exterior calculus, are a natural framework for electromagnetics; not only for a better understanding of the theoretical foundation, but also for the development of numerical methods. Keywords are the Hodge decomposition or the de Rham complex that are at the bottom of recent development of efficient multigrid methods or stable mixed finite element methods. Thinking in terms of such co-ordinate free differential forms offers considerable benefits as regards to the construction of structure preserving spatial discretizations.

In the present project, we aim at developing a new approach for the numerical treatment of resistive magnetohydrodynamics where a Galerkin discretization of the electromagnetic part based on finite element exterior calculus (FEEC) will be coupled to advanced finite volume methods for the approximation of the balance laws for the fluid.

The latest developments involved the extension and analysis of the stablized Galerkin schemes for advection of differential forms introduced in [48] to the case of time-dependent and non-regular flow fields.

6.2.1.15. Hamilton-Jacobi Formulation for Vlasov-Poisson

Participants: Holger Heumann, Eric Sonnendrücker [IPP, Max-Planck-Institute Garching, Germany], Philip J. Morrison [Institute for Fusion Studies, Austin, USA].

The phase space mapping induced by the solution of the Vlasov-Poisson problem is a symplectic mapping (or canonical transformation in physics literature) solving Hamilton's equations. In this project we are developing numerical methods that are based on this formulation. We derived and implemented new finite difference schemes for the corresponding Hamilton-Jacobi equation, that circumvent the projection of the distribution function inherent in Lagrangian methods. First numerical results for standard test problems show the ability of increased resolution of fine-scale effects.

6.2.1.16. Entropy viscosity technique

Participants: Richard Pasquetti, Jean-Luc Guermond [Texas A & M University], Boyan Popov [Texas A & M University].

The entropy viscosity technique allows to address hyperbolic equations by introducing a strongly non linear viscous term where needed, especially at shocks. The basic idea is to set up a viscosity from the residual of the entropy inequality together with a $O(h)$ upper bound proportional to the local wave speed. In view of addressing situations where vaccum may appear in the tokamak, we have considered the shallow water equations with topography and in situations where dry-wet transitions occur. Using a RK scheme in time and a spectral element method (SEM) in space, we have proposed a variant of the entropy technique, that mainly consists of using the viscosity upper bound in the dry parts, to obtain satisfactory results. This work was presented in [30], [22] and a publication is under review.

6.2.1.17. Bohm boundary conditions

Participants: Sébastian Minjeaud, Richard Pasquetti.

In the frame of the ANR project ESPOIR, our partners have proposed a penalty method to enforce the Bohm criterion (Mach number greater than one at the tokamak plates). This approach has been justified by considering a "minimal transport model" that consists of a 1-dimensional non linear hyperbolic system of two equations, that govern the evolutions of the density and velocity. The approach and further developments are described in three recent papers published in the journal of computational physics. Considering the same hyperbolic system, we have proposed a direct way to enforce the Bohm criterion in the frame of an explicit time marching. Using a SVV stabilized SEM it is then possible to resolve the same problem with spectral accuracy. This paper is now in press and will be published as a JCP note.

6.2.1.18. A numerical scheme for fluid-particules flows

Participants: Florent Berthelin, Thierry Goudon, Sebastian Minjeaud.

We propose a numerical scheme for the simulation of fluid-particles flows with two incompressible phases. The numerical strategy is based on a finite volume discretization on staggered grids, with a flavor of kinetic schemes in the definition of the numerical fluxes. We particularly pay attention to the difficulties related to the volume conservation constraint and to the presence of a close-packing term which imposes a threshold on the volume fraction of the disperse phase. We are able to identify stability conditions on the time step to preserve this threshold and the energy dissipation of the original model. The numerical scheme is validated with the simulation of sedimentation flows.

6.2.1.19. Identification and forecast of ionospheric disturbances

Participants: Didier Auroux, Sebastian Minjeaud.

In the framework of ANR IODISSEE, in order to identify (and forecast) ionospheric disturbances leading to temporary losses of satellite-to-earth communications (GPS, Galileo), we used Striation software for data assimilation. We obtained the adjoint code thanks to automatic differentiation (Tapenade software from Inria). As the data from Demeter satellite were not available, we extracted synthetic data from a generic model run, and we tried to identified some physical parameters (electronic density, atomic mass, number of particles) of the initial condition from the observations. For a small physical time scale (approximately 1 hour), the identification works very well, and it is possible to retrieve the initial condition from a sparse and noisy observations, allowing us to forecast the evolution of the ionospheric plasma - and then to forecast the disturbances and plasma bubbles that trap GPS and Galileo signals. For longer physical time windows (5 to 10 hours), the identification does not work anymore. We plan to work with real data, if possible, and also with a more complex model (for instance Dynamo software).

COFFEE Project-Team (section vide)

<h2 style="text-align:center; color:red;">DEMAR Project-Team</h2>

5. New Results

5.1. Modelling and identification of the sensory-motor system

5.1.1. Whole Body Center of Mass Estimation with Portable Sensors: Using the Statically Equivalent Serial Chain and a Kinect

Participants: Alejandro Gonzalez de Alba, Mitsuhiro Hayashibe, Vincent Bonnet, Philippe Fraisse.

The trajectory of the whole body center of mass (CoM) is useful as a reliable metric of postural stability. If the evaluation of a subject-specific CoM were available outside of the laboratory environment, it would improve the assessment of the effects of physical rehabilitation. A method is developed tot enable tracking CoM position using low-cost sensors such that it can be moved around by a therapist or easily installed inside a patientâs home. We compare the accuracy of a personalized CoM estimation using the statically equivalent serial chain (SESC) method and measurements obtained with the Kinect to the case of a SESC obtained with high-end equipment (Vicon). We also compare these estimates to literature-based ones for both sensors. The method was validated with seven able-bodied volunteers for whom the SESC was identified using 40 static postures. The literature-based estimation with Vicon measurements had a average error 24.9 ± 3.7 mm; this error was reduced to 12.8 ± 9.1 mm with the SESC identification. When using Kinect measurements, the literature-based estimate had an error of 118.4 ± 50.0 mm, while the SESC error was 26.6 ± 6.0 mm. The subject-specific SESC estimate using low-cost sensors has an equivalent performance as the literature-based one with high-end sensors. The SESC method can improve CoM estimation of elderly and neurologically impaired subjects by considering variations in their mass distribution.
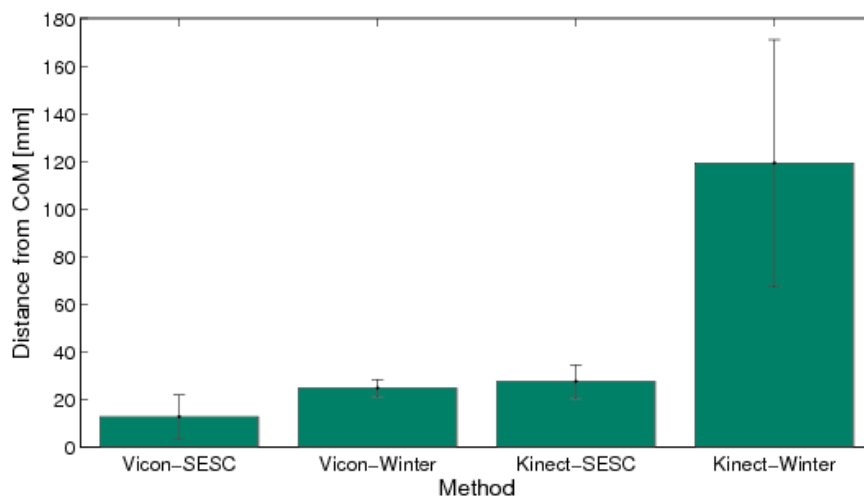

Figure 2. The bars correspond to the averaged results for all seven subjects. We observe an increase in the accuracy of the identified SESCs with respect to the literature estimates; we found that the performance of the Kinect-SESC is equivalent to that of the literature-based estimate using high-end sensors.

5.1.2. A Personalized Balance Measurement for Home-based Rehabilitation

Participants: Alejandro Gonzalez de Alba, Philippe Fraisse, Mitsuhiro Hayashibe.

We use the subject-specific center of mass (CoM) estimate offered by the statically equivalent serial chain (SESC) method and the zero rate of change of angular momentum (ZRAM) concept to evaluate balance for a series of dynamic motions. Two healthy subjects were asked to stand on a Wii balance board and their SESC parameters were identified. A set of dynamic motions was to evaluate the rate of change of centroidal angular momentum and the distance of the ZRAM point to the center line of the support polygon. We found a good match between both balance metrics. As an application example, the subjects performed a dynamic motion (such as walking and abruptly stopping) and the stability was determined in real-time using the ZRAM point from the personalized CoM trajectory. This was implemented with a real-time balance visualization tool based on Kinect measurements for home-based rehabilitation.
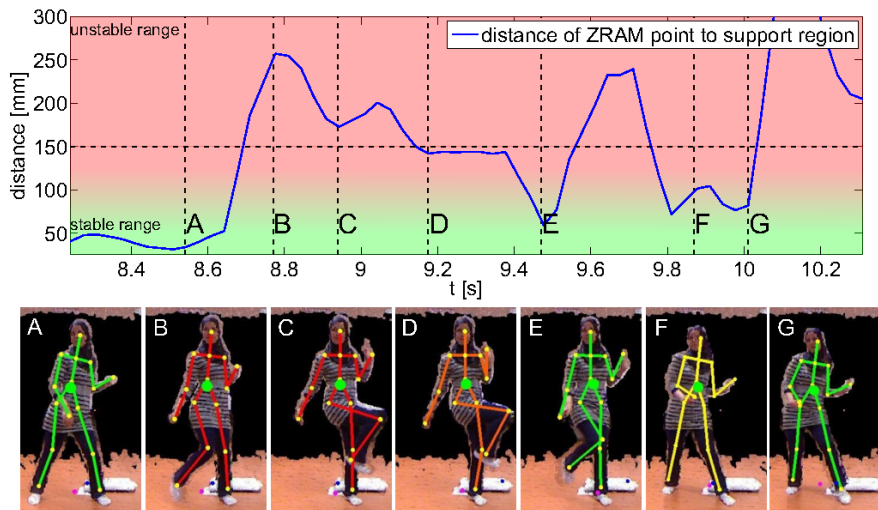

Figure 3. The distance between the ZRAM point and the center of the support polygon can be used to determine balance in real-time and shown as feedback during a physical rehabilitation program. In this example, the subject's skeleton is colored in red to indicate an unstable movement and in green for a stable one.

5.1.3. Methodology for automatic movement cycle extraction using Switching Linear Dynamic System

Participants: Roberto de Souza Baptista, Mitsuhiro Hayashibe, Antônio P. L. Bó [Univ. Brasilia].

Human motion assessment is key for motor-control rehabilitation. Using standardized definitions and spatiotemporal features - usually presented as a movement cycle diagram- specialists can associate kinematic measures to progress in rehabilitation therapy or motor impairment due to trauma or disease. Although devices for capturing human motion today are cheap and widespread, the automatic interpretation of kinematic data for rehabilitation is still poor in terms of quantitative performance evaluation. In this paper we present an automatic approach to extract spatiotemporal features from kinematic data and present it as a cycle diagram. This is done by translating standard definitions from human movement analisys into mathematical elements of a Switching Linear Dynamic System model. The result is a straight-forward procedure to learn a tracking model from a sample execution. This model is robust when used to automatically extract the movement cycle diagram of the same motion (the Sit-Stand-Sit, as an example) executed in different subject-specific manner such as his own motion speed.
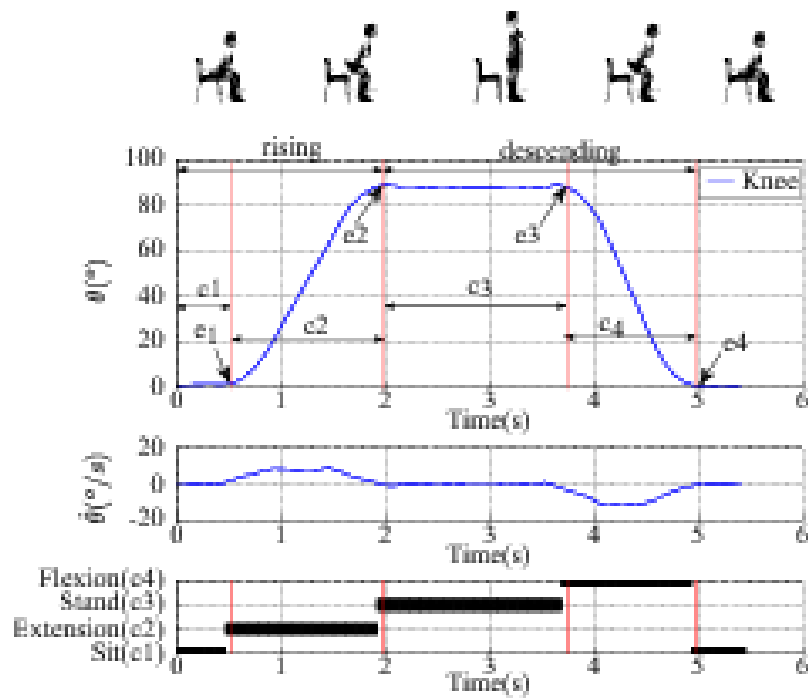
Figure 4. Training dataset consisting of one execution of the Sit-Stand-Sit movement cycle. Events (e_i), components (c_i) and the rising and descending phases are marked. θ and $\dot\theta$ indicates angle and angular velocity.

5.1.4. Real-time Muscle Deformation via Decoupled Modeling of Solid and Muscle Fiber Mechanics

Participants: Yacine Berranen, Mitsuhiro Hayashibe, David Guiraud, Benjamin Gilles.

This work presents a novel approach for simulating 3D muscle deformations with complex architectures. The approach consists in choosing the best model formulation in terms of computation cost and accuracy, that mixes a volumetric tissue model based on finite element method (3D FEM), a muscle fiber model (Hill contractile 1D element) and a membrane model accounting for aponeurosis tissue (2D FEM). The separate models are mechanically binded using barycentric embeddings. Our approach allows the computation of several fiber directions in one coarse finite element, and thus, strongly decreases the required finite element resolution to predict muscle deformation during contraction. Using surface registration, fibers tracks of specific architecture can be transferred from a template to subject morphology, and then simulated. As a case study, three different architectures are simulated and compared to their equivalent one dimensional Hill wire model simulations.
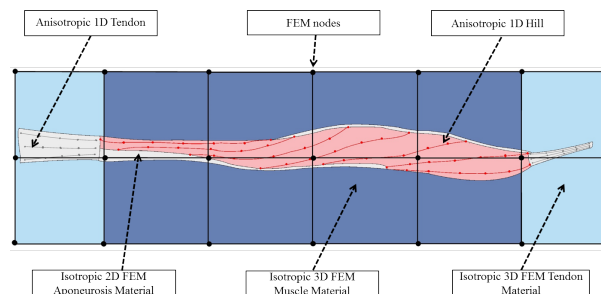

Figure 5. Muscle multi-model scheme: The different models are linked via barycentric embeddings. This approach strongly decreases the required finite element resolution to predict muscle deformation during contraction
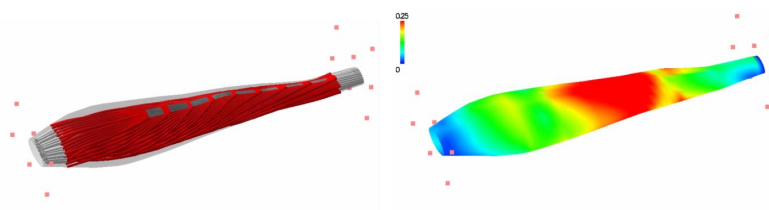

Figure 6. Left: Hybrid model of bipennate muscle, right: internal stress due to isometric contraction

5.1.5. Adaptive model for viscoelastic solids

Participants: Benjamin Gilles, Maxime Tournier, Matthieu Nesme, Francois Faure.

A new adaptive model for viscoelastic solids is presented in [50], [34]. Unlike previous approaches, it allows seamless transitions, and simplifications in deformed states. The deformation field is generated by a set of physically animated frames. Starting from a fine set of frames and mechanical energy integration points, the model can be coarsened by attaching frames to others, and merging integration points. Since frames can

be attached in arbitrary relative positions, simplifications can occur seamlessly in deformed states, without returning to the original shape, which can be recovered later after refinement. We propose a new class of velocity-based simplification criterion based on relative velocities. Integration points can be merged to reduce the computation time even more, and we show how to maintain constant elastic forces through the levels of detail. This meshless adaptivity allows significant improvements of computation time.

5.1.6. Functional Brain Stimulation: Filling the gap between micro- and direct electrical stimulation of the brain in order to better understand and innovate

Participants: Marion Vincent, Olivier Rossel, Mitsuhiro Hayashibe, Guillaume Herbet, Hugues Duffau [Neurosurgery Department, CHU, Montpellier], David Guiraud, François Bonnetblanc.

Micro-stimulation (MES) and Direct electrical stimulation (DES) of the brain are both used to perform in vivo functional mapping of the brain in fundamental neuroscience and neurosurgery respectively. The former is performed in animal experiments while the latter is performed on humans in the operative room. Very recently, a strong debate occurred to determine whether DES used during "wide-awake surgery" with success is a gold standard to study brain functions (Mandonnet et al., 2010; Borchers et al. 2012; Desmurget et al., 2013). In this debate, confusion is very often made between DES and MES, as these are considered to induce similar effects on the nervous tissues, with comparable behavioural consequences. However, electrical stimulation (ES) parameters used in both techniques are clearly different. More surprisingly, a strong biophysical rational of their choices is lacking. It may be due to historical, methodological and technical constraints that have guided empirically them. These constraints may have strongly shaped and limited experimental protocols in a standard way. By contrast, the gap between MES and DES may reveal a great potential for new experimental paradigms in ES of the brain in vivo. By considering this gap and new technical developments in the design of stimulators, it may be time to move on to alternative and innovative stimulations protocols, especially regarding and inspired from what is performed in functional electrical stimulation (FES) of peripheral nerves, for which more theoretical supports exist.

5.1.7. Modelling of structural and functional brain connectivity networks in Diffuse Low-Grade Glioma (DLGG)

Participants: Jija Syamala James, Anirban Dutta, François Bonnetblanc, David Guiraud, Nicolas Menjot de Champfleur [Department of Neuroradiology,CHU, Montpellier], Emmanuelle Le Bars [Institute of Human Functional Imaging 12 FH, CHU, Montpellier], Hugues Duffau [Neurosurgery Department, CHU, Montpellier].

The impairment of functional brain connectivity networks in Diffuse Low Grade Glioma (DLGG) subjects can lead to distinct functional deficits where the challenge remains in greater understanding of distribution of dynamic brain connectivity. Our multimodality (resting state functional MRI, Diffusion Tensor Imaging and tractography) neuroimaging study aims to evaluate the differences in spatial and temporal patterns of brain connectivity networks in DLGG patients, pre and postoperatively.

In order to identify the brain connectivity networks, we analysed resting state fMRI data of 22 mesio-frontal DLGG patients by using MELODIC (Multivariate Exploratory Linear Optimised Decomposition into Independent Components)-ICA module, implemented in Freesurfer Software Library [FSL] (www.fmrib.ox.ac.uk/fsl). First we evaluated the clinical efficacy of rsfMRI technique to non-invasively map the dynamic functional reorganizations of brain connectivity networks in glioma subjects. Further we observed the existence of altered inter-hemispheric functional connectivity in DLGG subjects postoperatively.

We are currently performing this analysis using larger sample size in order to find out

a) the invivo structure-function relationship of motor, language and visual brain networks by fusing the information from DTI and fMRI.
b) the correlation between brain connectivity networks and neurobehavioural performance.
c) the neuroplastic alterations in topographic organization and strength of connections before and after DLGG surgery.
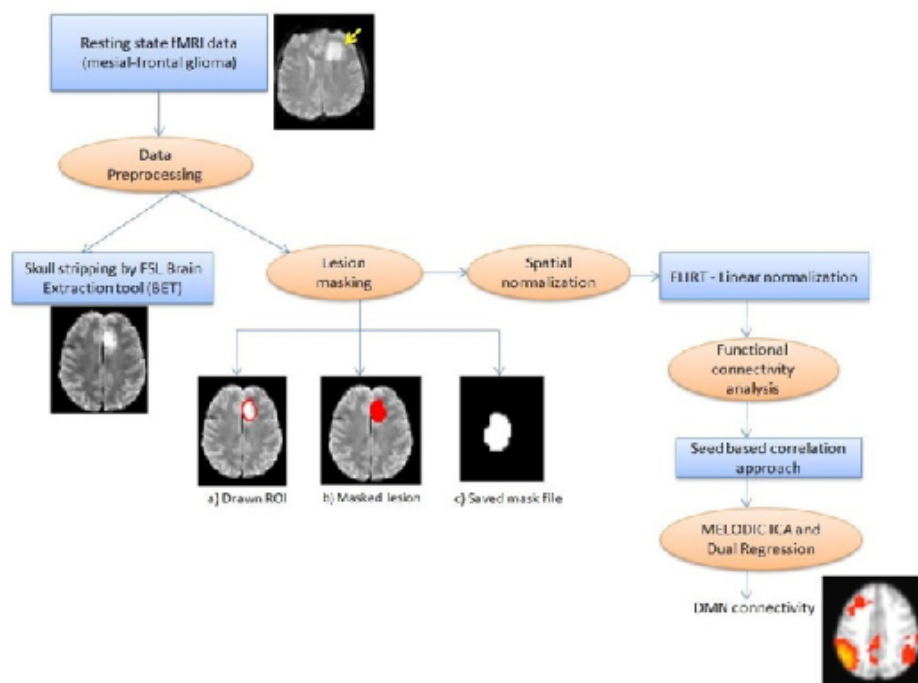
Figure 7. Steps involved in resting state functional connectivity analysis.

5.1.8. SPINSTIM: Direct spinal stimulation for rehabilitation of bladder, bowel and sexual functions in spinal cord injury

Participants: Christine Azevedo Coste, Luc Bauchet [Neurosurgery Department, CHU, Montpellier], Claire Delleci [CHU Bordeaux], Charles Fattal, Thomas Guiho, David Guiraud, Jean-Rodolphe Vignes [CHU Bordeaux].

For the general public, spinal cord injury (SCI) is often restricted to limb paralysis. In reality, by interrupting communication between encephalon and peripheral organs, medullary wounds lead to physiological deficiencies such as urinary (retention and/or incontinence), gastrointestinal (constipation) or sexual impairments; disorders which are the center of patient's expectations.

Spinal cord stimulation (SCS) is a general term which includes both epidural and intradural stimulation. Originally associated with the treatment of chronic neurogical pain (in the 1970ies), SCS led also to immediate and profound improvements of sensory and motor functions in recent studies both on SCI patients (few subjects involved) and rodents.

Despite these promising results some limitations have still to be overcome. Among them, the use of small animal models, the empirical aspect of the stimulation procedure and the impact of these protocols on intestinal and urinary functions are critical.

To counteract these limits, we want to explore intradural and epidural stimulations in an intermediate model- the house pig- and assess their impact on bladder, guts and genitals. In order to evaluate our approach, we will record EMG signals of lower limbs and sphincters (both urethral and anal), and simultaneously, we will monitor bladder and rectal pressure.

5.1.8.1. Development of point of care testing (POCT) device for neurovascular coupling from simultaneous recording of EEG and NIRS during non-invasive brain stimulation (NIBS) for closed-loop control of NIBS.

Participants: Mehak Sood, Utkarsh Jindal, Mitsuhiro Hayashibe, Stephane Perrey, Michael A. Nitsche, Abhijit Das, Shubhajit Roy Chowdhury, Anirban Dutta.

Our preliminary work showed that transcranial direct current stimulation (tDCS) can perturb local neuronal activity which can be used for assessing regional neurovascular coupling (NVC) functionality. It was postulated that tDCS leads to rapid dynamic variations of the brain cell microenvironment that perturbs the hemodynamic and electromagnetic responses. Based on these preliminary studies, we developed a POCT device for EEG-NIRS based screening and monitoring of neurovascular coupling functionality under perturbation with tDCS. The stroke case study showed detectable changes in the degree of NVC to a $0.526A/m^2$ square-pulse (0-30sec) of anodal tDCS where these alterations in the vascular system may result in secondary changes in the cortical excitability. The objective of this case study was to evaluate an empirical method to assess NVC using cross-correlation function (CCF) between mean (cortical) tissue oxy-haemoglobin concentration time-series and averaged PSD time-course from the EEG spectrogram. The CCF based assessment of the patient-specific status of NVC are currently being studied in a larger cohort with small vessel diseases. The overarching goal is closed-loop control of tDCS based on simultaneous recording of EEG and NIRS during non-invasive brain stimulation.

5.2. Synthesis and Control of Human Functions

5.2.1. Ergonomics of the control by a quadriplegic of hand functions

Participants: Christine Azevedo Coste, David Guiraud, Wafa Tigra, Charles Fattal.

In subjects with complete Spinal Cord Injury (SCI) above C7, the four limbs are paralyzed (quadriplegia). Recovery of grasping movements is then reported as a priority. Indeed, most activities of daily living are achieved through upper limbs. Thus, restoration of hand and forearm active mobility could significantly increase independence and quality of life of these people and decrease their need of human aid. Although most of the subjects plebiscite pharmacological or biological solutions, only orthotics and Functional Electrical Stimulation (FES) allow, so far, to restore hand movements but they are rarely used. Ergonomics and comfort
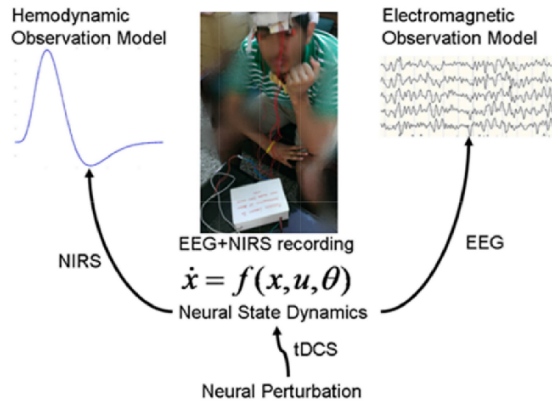
Figure 8. EEG-NIRS based simultaneous recording of the hemodynamic and electromagnetic responses to perturbations with transcranial direct current stimulation (tDCS).

of piloting mode could partly explain the low usage of these systems. In this context, our aim is to explore possible solutions for subjects to interact with such devices. In this article we propose to evaluate the capacity of active upper limb muscles contraction to be used to intuitively control FES in tetraplegic subjects. In this study, we assessed the ability to gradually contract different muscles: trapezius, deltoid, platysma and biceps. Three subjects with C6 to C7 neurological level of lesion were included. We show that over the active upper limb muscles tested, contraction of the trapezius muscle was considered by the subjects as the most comfortable and could be employed as an intuitive mode of control of functional assistive devices.

5.2.2. *Implementation of filtering, calibration and reconstruction algorithms dedicated to the use of inertial measurement units related to rehabilitation and movement analysis*

Participants: Christine Azevedo Coste, Benoît Sijobert, Roger Pissard-Gibollet.

This work has been done within SENSBIO ADT.

- Stride length estimation

    An algorithm has been implemented for the sep by step stride length calculation from only one shank located sensor algorithm (calibration, segmentation and reconstruction) Experimental validation was done on 10 healthy subjects (error <10%), 12 Parkinson disease subjects, and 7 hemiparetic subjects.

- Comparison of different reconstruction algorithms

    A Python programmed toolbox was developed for movement analysis (SensbioTK) in order to compare the performances of Mahony, Madgwick, Martin-Salaun sensor fusion algorithms.

- Real-time MoCap data processing Matlab based software

    The algorithm is based on Nexus SDK for rotation angle and translation calculation of a tracked object.

5.2.3. *Dominant Component in Muscle Fatigue Induced Hand Tremor during Laparoscopic Surgical Manipulation*

Participants: Sourav Chandra, Mitsuhiro Hayashibe, Asokan Thondiyath [IIT Madras].

Accuracy of laparoscopic surgery gets affected by the hand tremor of the surgeons. Though cognitive load is inevitable in such activity which promotes tremor, muscle fatigue induced tremor is significant among the most important source of tremor. Characteristic of fatigue induced hand tremor and its dominant directional properties are reported in this work. For a fixed laparoscopic tool grip with temporally synchronized predefined task protocols, characteristics of fatigue induced tremors have been studied. Dominant component of tremor is found to be in the sagittal plane in case of both static and dynamic tasks. The results shown in the figure, sagittal plane (z axis) component of hand tremor is higher than the other directions. In order to relate it with the muscle fatigue level, spectral properties of surface electromyography (SEMG) were also investigated simultaneously. A study of transient effect on tool positioning was also included, which conjointly advocates the other experimental results on fatigue induced hand tremor as well. Currently a better metric for muscle fatigue is being analyzed and studied with a purview of relation in between SEMG and hand tremor (fig.9).
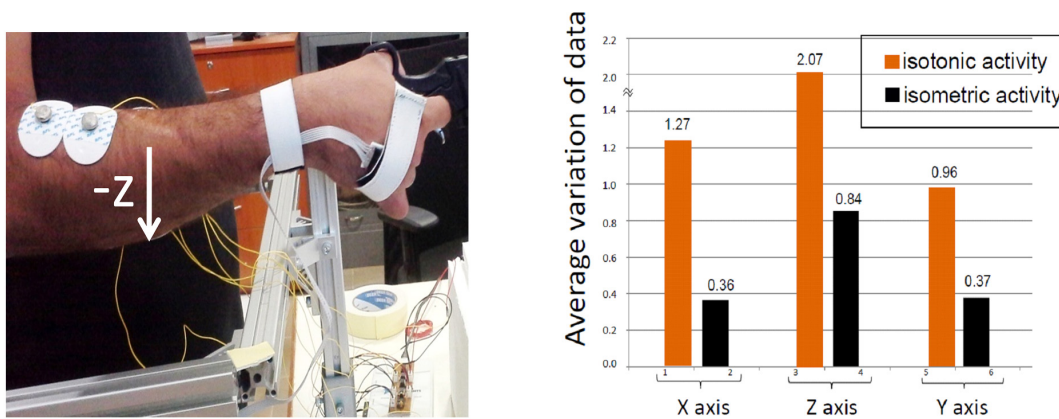

Figure 9. (left) Experimental hand tremor measurement setup (right) Dominant component of hand tremor in all three planes

5.2.4. Human arm optimal motion analysis in industrial task for improving ergonomics of painful workstations

Participants: Nahema Sylla, Vincent Bonnet [TUAT, Japan], Nahid Armande [PSA], Philippe Fraisse.

In PSA Peugeot Citroen factories, high precision requirements of workstations make them being manual. One of the main goal of the car manufacturer is to minimize the pain of workers while maintaining high efficiency of production lines. Consequently, assisting operators with an exoskeleton is a potential solution for improving ergonomics of painful workstations while respecting industrial constraints [33]. We have developed a new approach based on inverse optimization to better understand human arm motion in industrial screwing task. The process combines several criteria to minimize such as energy expenditure or trajectory smoothness leading to the optimal trajectory of a typical screwing task, often performed by workers. Estimated joint trajectories are similar with the measured ones, with a mean square error of 4 degrees. The resulting cost-function is mainly composed of energy expenditure and geodesic criteria. Results show the relevance of using composite cost function in human motion planning [52]. This method has been applied to evaluate a 7 DoF exoskeleton in terms of motion control. The results of our study show that the hybrid composition of the free arm movement was accurately determined. At contrary, when wearing the exoskeleton, which produces an arbitrary determined torque compensation, the motion is different from the naturally adopted one. This study is part of the evaluation and comprehension of the complex neuromuscular mechanism resulting in wearing an exoskeleton several hours per day for industrial tasks assistance [49].

5.2.5. A System for Real-time Online Estimation of Joint Torque with Evoked EMG under Electrical Stimulus

Participants: Zhan Li, Mitsuhiro Hayashibe, David Andreu, David Guiraud.

Functional electrical stimulation (FES) is a useful rehabilitation technique for restoring motor capability of spinal cord injured (SCI) patients by artificially driving muscle contraction through delivering electrical pulses. Real-time FES systems with online modulation ability are in great demand towards clinic applications. In this work, online estimation of joint torque with evoked electromyography (eEMG) in real-time environment is presented. The eEMG is acquired by National Instrument (NI) acquisition card and the stimulus is produced by wireless stimulator (manufactured by Vivaltis Inc., France). Kalman filter (KF) is adopted and embedded as the online estimator. Such real-time online torque estimation system produces promising results. Currently, the implementation to use EMG signal also from wireless Pod module is under investigation (fig.10).
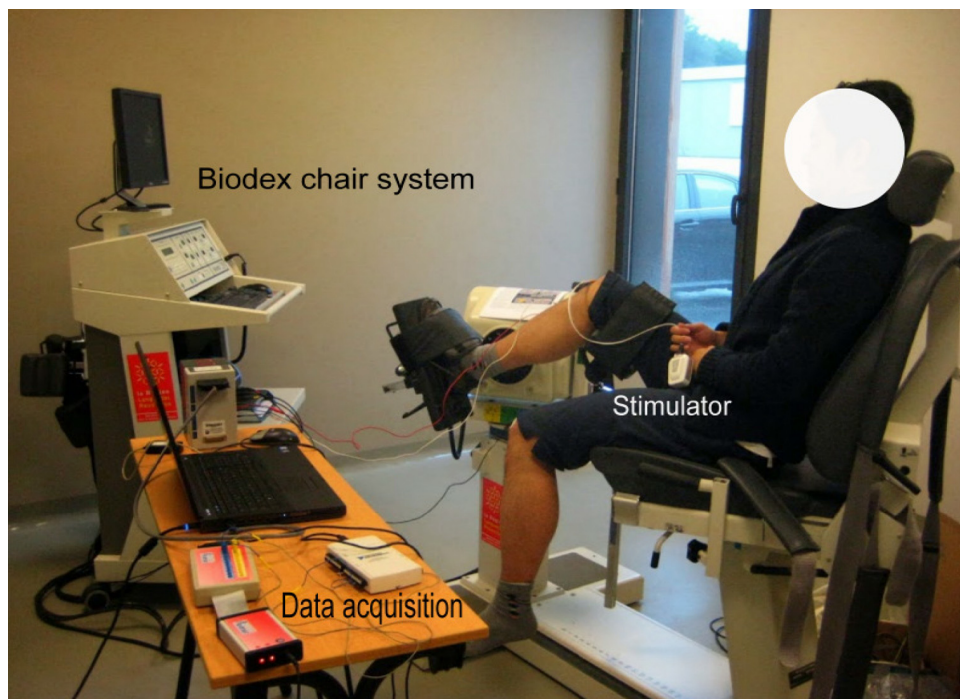

Figure 10. Real-time online FES-induced torque estimation system

5.2.6. Freezing of Gait Analysis and Detection

Participants: Christine Azevedo Coste, Benoît Sijobert, Roger Pissard-Gibollet, Christian Geny [CHU Montepllier, Neurologie].

We have extended and optimized the work on Freezing Of Gait (FOG) and destination detection (Maud Pasquier PhD thesis, 2013). A new software for FOG Criteria based on cadency and stride length calculation has been implemented. Results were compared to Freezing Index (Moore et al., 2008) based on frequencies analysis of legs vertical acceleration. The comparison between detection and reality is done on the basis of video analysis of the performed tasks. A software has been implemented for the video labelization Matlab/VLC based tool, with graphical user interface, for marking and synchronizing events from a video (MovieFOG). 14 Parkinson disease patients were included in the study [12][48] (fig.11).
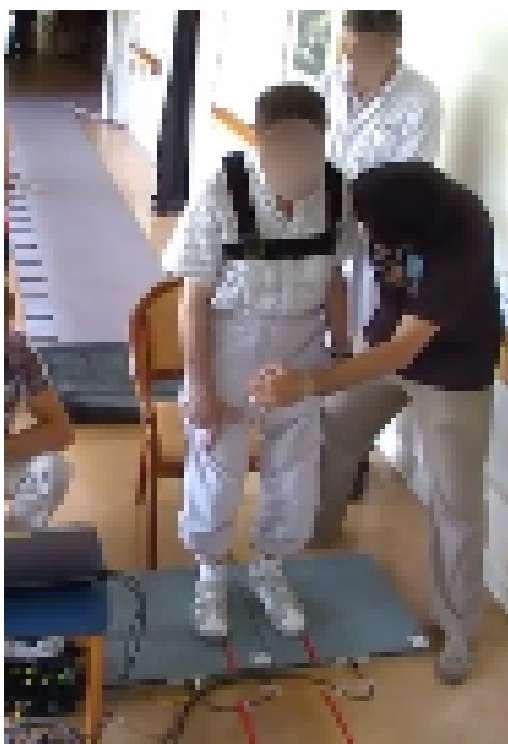
Figure 11. Freezing of Gait Analysis and Detection

5.2.7. *A novel brain-computer interfacing paradigm for control of multi-DOFs robot with adaptive EEG decoding and synergetic environment adaptation*

Participants: Saugat Bhattacharyya, Shingo Shimoda [RIKEN, Japan], Mitsuhiro Hayashibe.

The study proposes a novel brain-computer interfacing paradigm for control of a multi-joint redundant robot system. Here, the user would determine the direction of end-point movement of a 3-dof robot arm using motor imagery electroencephalography (EEG) signal with co-adaptive decoder while a synergetic motor learning algorithm manages a peripheral redundancy in multi DOF joints toward energy optimality through tacit learning. As in human motor control, torque control paradigm is employed for a robot to be sensitive to the given environment. The dynamic condition of the robot arm is taken into consideration by the learning algorithm. Thus, the user needs to only think about the end-point movement of the robot arm, which allows simultaneous multi-joints control by BCI. The k-Nearest Neighbor based decoder designed for this study is adaptive to the changing mental state of the user. Offline experiments on the decoder reveals that its classification accuracy gradually increases at each learning stage. Online experiments also reveals that the users successfully reach their targets with an average decoder accuracy of more than 65% in different end-point load conditions. The details of the BCI control paradigm, shown in Fig.12 , is as follows: Initially, the robot is trained to its dynamic environment using a tacit learning approach for a fixed period of time. In this study, the load carried by the robot is treated as the environmental changes along with link segment inertial configuration changes. As a result, the movement of the joints of the robot adapts to the changing load. After the training of the robot, the subject begins his/her task of visualizing the target and decides on the direction of motion of the robot. Here, we have used left and right movement imagery to move the robot in upward and downward direction, respectively. Subsequently, the EEG signals (corresponding to the movement imagined) are pre-processed and their corresponding features are extracted. These features are used as inputs to the decoder which determines the mental state of the user and sends command to the robot to move in the equivalent direction.

5.2.8. *Impact of the gaze direction on the skier trajectory*

Participants: Christine Azevedo Coste, Benoît Sijobert, Roger Pissard-Gibollet, Nicolas Coulmy [FFS annecy].

This work is done within a collaboration with the French Skiing Federation (FFS). Preliminary indoor tests for designing an experimental protocol using motion capture and inertial sensors was realized using a ski simulator and Motion Capture Tool from DEMAR (VICON BONITA). Indoor and outdoor experimentations were done with olympic skiers. In outdoor tests IMUs were combined with SMI eye tracking device. Preliminary analysis and data reconstruction has been done. A VPython script for 3D visualizing of skier movements was developed (fig.13).

5.2.9. *Development of a low-cost biofeedback system for electromyogram-triggered functional electrical stimulation therapy in conjunction with non-invasive brain stimulation*

Participants: Anirban Dutta, Christine Azevedo Coste, Mitsuhiro Hayashibe, Uttama Lahiri, Abhijit Das, Michael A. Nitsche, David Guiraud.

Functional electrical stimulation (FES) facilitates ambulatory function after paralysis by activating the muscles of the lower extremities. The FES-assisted stepping can either be triggered by a heel-switch, or by an electromyogram-(EMG-) based gait event detector. A group of six chronic (>6 months post-stroke) hemiplegic stroke survivors underwent transcutaneous FES-assisted training for 1 hour on stepping task with EMG biofeedback from paretic tibialis anterior (TA) and medial gastrocnemius (GM) muscles, where the stimulation of the paretic TA or GM was triggered with surface EMG from the same muscle. During the baseline, post-intervention, and 2-day-postintervention assessments, a total of 5 minutes of surface EMG was recorded from paretic GM and TA muscles during volitional treadmill walking. Two-way ANOVA showed significant effects in terms of P-values for the 6 stroke subjects, 0.002, the 3 assessments, 0, and the interaction between subjects and assessments, 6.21E-19. The study showed a significant improvement from baseline in paretic GM and TA muscles coordination during volitional treadmill walking. Moreover, it was found that the EMG-triggered
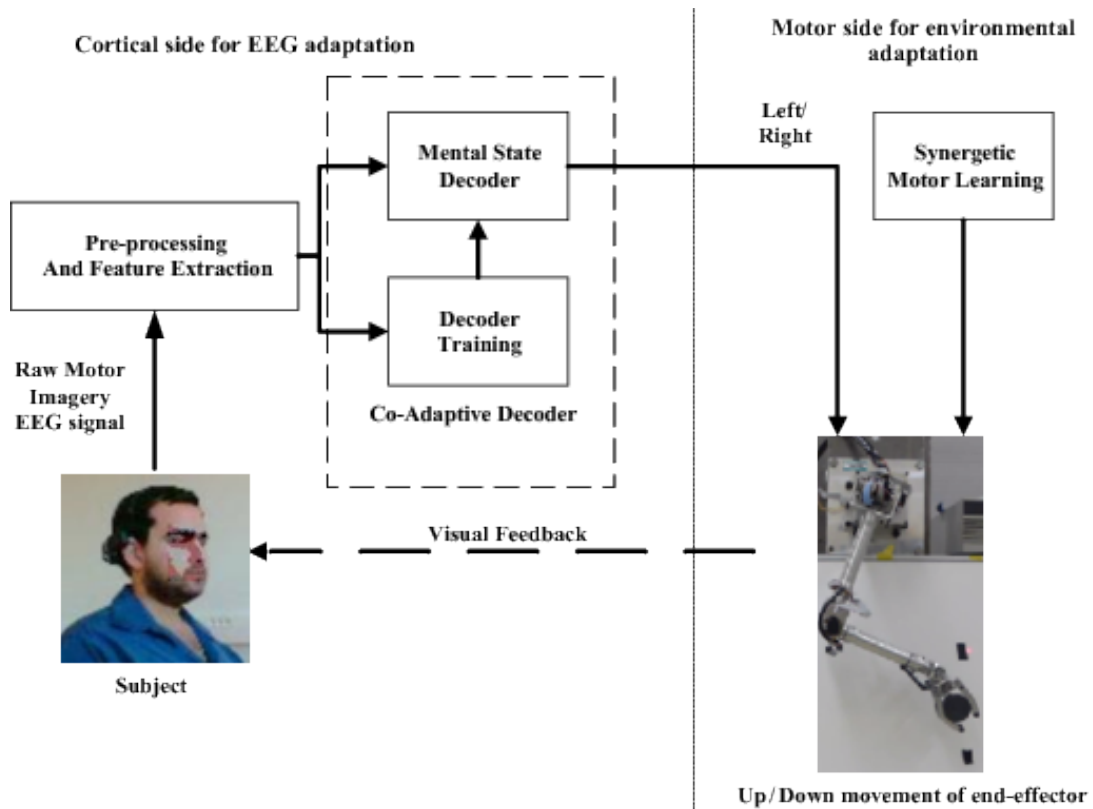
Figure 12. BCI paradigm employed in this study for control of multi-DOFs robot using adaptive left-right motor imagery decoder and synergetic motor learning for peripheric redundancy management (via tacit learning).
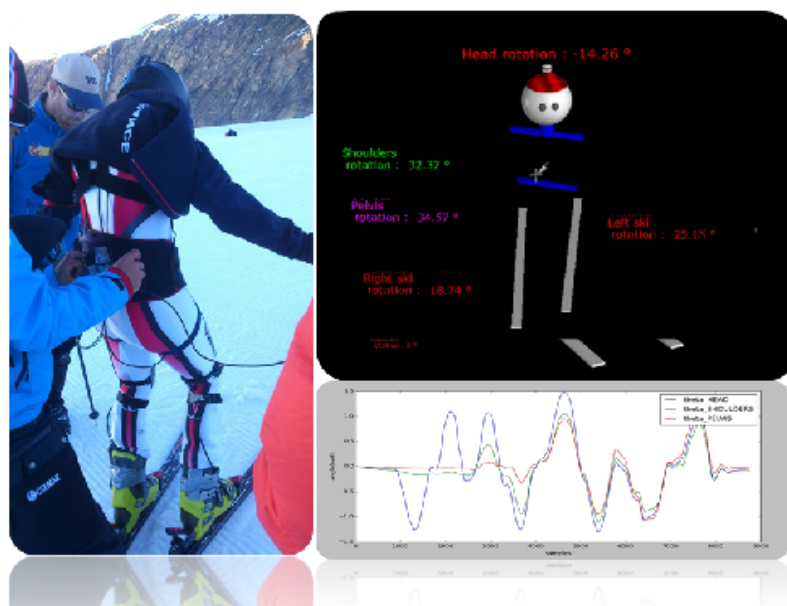
Figure 13. Impact of the gaze direction on the skier trajectory

FES-assisted therapy for stand-to-walk transition helped in convergence of the deviation in centroidal angular momentum from the normative value to a quasi-steady state during the double-support phase of the nonparetic. Also, the observational gait analysis showed improvement in ankle plantarflexion during late stance, knee flexion, and ground clearance of the foot during swing phase of the gait. Currently, we are conducting preliminary stroke studies to evaluate non-invasive brain stimulation as an adjunct to EMG-triggered FES therapy for movement rehabilitation [19][22].

5.2.10. Development and bench-testing of a low-cost eye tracking system (ETS) to measure gaze abnormality in stroke towards virtual reality based visuomotor therapy task.

Participants: Deepesh Kumar, Abhijit Das, David Guiraud, Michael A. Nitsche, Anirban Dutta, Uttama Lahiri.

We conducted a preliminary usability study while incorporating our novel low-cost ETS to measure one's eye gaze indices in response to presented visual task. The ETS provided gaze-related biomarkers which has the potential to be mapped to the probable abnormalities in one's eye movement pattern in stroke. Our preliminary findings with stroke-survivors and age-matched healthy participants indicate the potential of our low-cost ETS to provide quantitative measures of the difference in gaze-related biomarkers between the two groups of participants. Based on these preliminary results, we are conducting a clinical stroke study on ETS based screening and monitoring of performance during a virtual reality based visuomotor balance therapy task.

5.3. Neuroprostheses and technology

5.3.1. Abstraction and composition for formal design of neuroprotheses

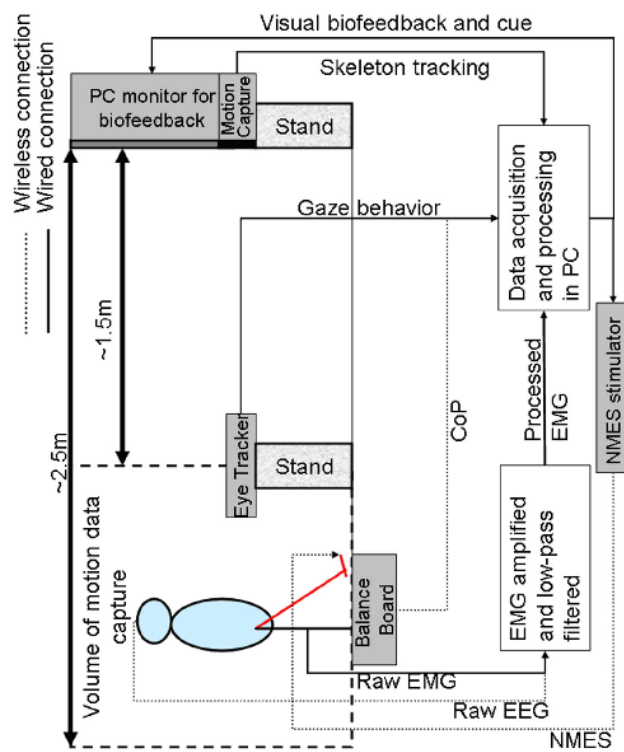**Participants:** David Andreu, Hélène Leroux, Karen Godary-Dejean.

Figure 14. Human-machine-interface integrating low-cost sensors for post-stroke balance rehabilitation
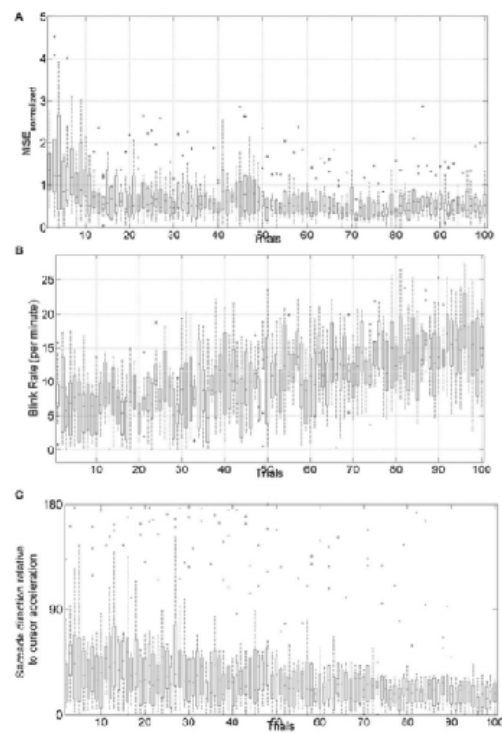
Figure 15. A. Box-plot of normalized mean squared error (MSE) across 10 healthy subjects, B. Box-plot of the blink rate during the visuomotor task, C. Box-plot of saccadic direction relative to the cursor acceleration during the visuomotor task.

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 (HILECOP). One of the inherent difficulties for designer is, on the behavioral part, to account for exceptions. This often leads 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 te on fig. 13), whatever is the current state of the sub-net.

We also solved state evolution conflicts introducing (automatically) priorities between transitions, to avoid reaching inconsistent global state while synchronously executing the model.

The new formalism (including all improvements) has been defined [45], as well as the model transformation based equivalent PNML generation for using existing analysis tools. The VHDL code generation has also been defined [59].

All this work has been applied to an industrial example, that of a neural stimulator developed in collaboration with MXM industrial partner. Results have shown the significant contribution of the theoretical approach to the stimulation device reliability, while preserving both surface and power consumption of the given digital part of the device.

Ongoing work, developed through I. Merzoug PhD thesis, concerns the improvement of the analysis of synchronously implemented Petri nets.

5.3.2. *New FES dedicated digital processor for neurostimulator*
**Participants:** David Andreu, David Guiraud.

We designed (patent pending) and prototyped a new neural FES dedicated processor and its associated (more compact and efficient) set of instructions, as well as an embedded sequencer for accurate timing in sequencing stimulations to be performed (by the stimulator). The new neural stimulator is based on a dedicated ASIC (Application Specific Integrated Circuit), that is able to drive up to 24 channels of stimulation in absolute synchronization, and with a programmable and controlled current level distribution (patent pending). This ASIC also allows for impedance measurement. The functions of the stimulator are currently implemented in two separate chips: an analog stimulation front-end (ASIC) and a field-programmable gate array (FPGA) embedding the logic control. The FPGA embeds the new FES dedicated processor setting the output stage configuration (poles configuration and current ratio between them) and running potentially complex stimulation profiles (with a $1\mu s$ time step and 5 μA current step); example of generated stimulations are shown in Fig.16 . It also embeds the protocol stack allowing for remote programming and online control. Online control relies on advanced and efficient modulation mechanisms, e.g. coefficient based modulation preserving balanced stimulation.

And last but not least, it also embeds a monitoring module ensuring the respect of safety constraints stemming both from target tissue protection and electrode integrity preservation; this reference model based monitoring module ensures (configurable) current and quantity of injected charges limits and thus safe stimulation whatever are electrodes to be used (particularly for thin-film micro-electrodes). Safety limits must be defined by users (partners) according to the target and electrodes to be used.

5.3.3. *Fast simulation of hybrid dynamical systems*
Participants: Abir Ben Khaled [IFPEN], Daniel Simon, Mongi Ben Gaid [IFPEN].

When dealing with the design of complex systems, simulation is an indisputable step between concept design and prototype validation. Realistic simulations allow for the preliminary evaluation, tuning and possibly redesign of proposed solutions ahead of implementation, thus lowering the risks. To be confident in the result, building such simulations needs high fidelity models both for the components and for their interaction. Models
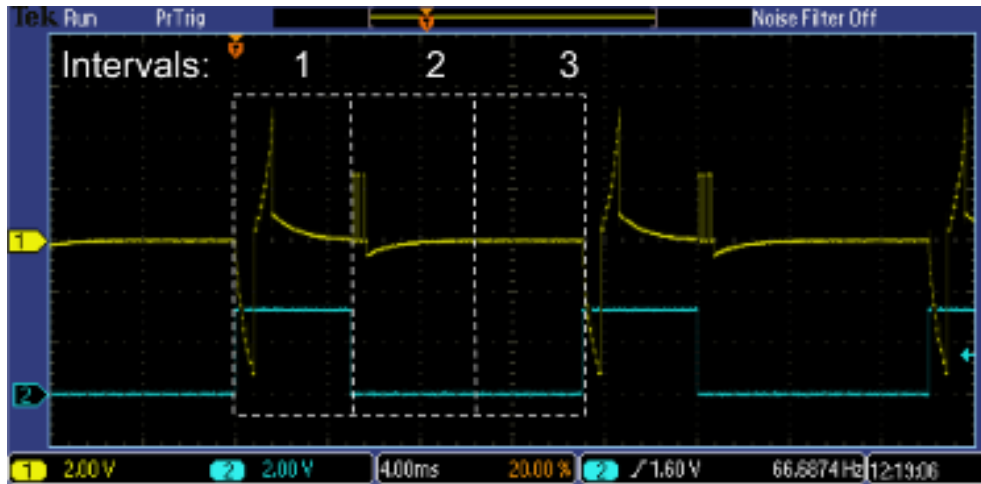
Sequence of stimulation with 3 intervals
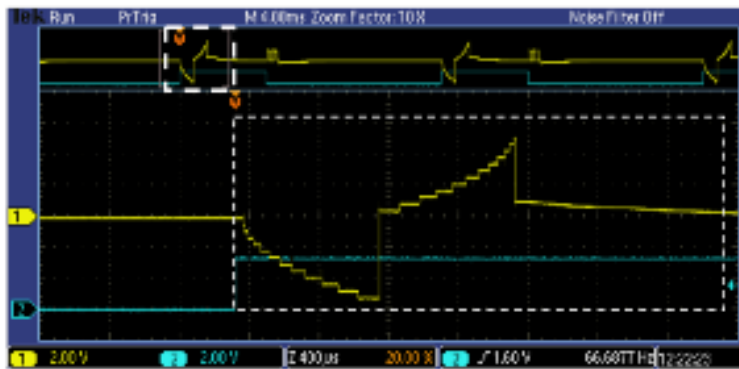
zoom on the stimulation profile generated at interval 1

zoom on the stimulation profile generated at interval 2

Figure 16. Sequence of stimulation generated by the new neurostimulator

of dynamical systems (as, for example, muscular fibers) are often given as a set of Ordinary Differential Equations (ODEs). However, the simulation of high-fidelity models is time consuming, and reaching real-time constraints is out of the capabilities of monotithic simulations running on single cores.

The aim of the on-going work is to speed up the numerical integration of hybrid dynamical systems, eventually until reaching a real-time execution, while keeping the integration errors inside controlled bounds. The basic approach consists in splitting the system into sub-models, which are integrated in parallel. I has been shown that an efficient partition must minimize the interactions between sub-models, in particular by confining discontinuities processing inside each component. Automatic partitioning, based on some particular incidence matrices of the original system, has been investigated to introduce a finely-grained co-simulation method enabling numerical integration speed-ups [14]. It is obtained using a partition across the model into loosely coupled sub-systems with sparse communication between modules. The proposed scheme leads to schedule a large number of operations with a wide range of execution times. A suitable off-line scheduling algorithm, based on the input/output dynamics of the models, is proposed to minimize the simulation errors induced by the parallel execution.The method was tested with an automotive engine model, but it is generic and can be applied to other systems of hybrid ODEs/DAEs, as are large sets of muscular fibers.

However, slack synchronization intervals may generate integration errors in the final result. Rather than using costly small integration and communication steps, an enhanced method uses context-based extrapolation is investigated to improve the trade-off between integration speed-ups, needing large communication steps, and simulation precision, needing frequent updates for the models inputs. the method uses extrapolations of the behavior of the models over the synchronization intervals. Test results on a hybrid dynamical engine model, based on FMI for model exchange, show that well chosen context-based extrapolation allows for significant speed-up of the simulation with negligible computing overheads [37].

5.3.4. *Control and scheduling co-design for stimulation systems*

Participants: Daniel Simon, David Andreu, Samy Lafnoune [Master2 Robotique].

In the FES distributed system developed by the team and marketed by Vivaltis, external electrodes are sticked on the body to either stimulate a muscle or to measure sensitive information (e.g., EMG). Each electrode is connected to a pod, which can be either a stimulation pod or a measurement pod. These Pods are controlled and coordinated by a controler through wireless radio-frequency (RF) links.

The communication frames between the controler and the pods use frames, each frame contains the receiver address and a code which correspond to the deired action. Simple communication stacks, based on the reduced OSI model, are implemented in each node of the wireless network, and the current version only provides static scheduling of the messages.

In fact static scheduling is not at all optimal when the generation of complex motions needs a tight coordination between several sensors and actuators over closed-loop controllers. The high sensitivity of the RF link with respect to varying networking loads and environment conditions calls for an adaptive scheduling of the messages via the regulation of a QoS criterion. The design of such feedback controllers wil rely on previous work [53], [54].

As any modification of the existing devices can be costly and as experiments involving livings cannot be done easily, a real-time simulation system has been design and implemented. The system includes continuous models of the muscular and skeleton systems, models of the wireless network, simplified communicaion stacks and control code running inside real-time threads. The system is open, running C code inside posix threads under Linux, so that the models can be progressively detailed and enriched when necessary [56].

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

6. New Results

6.1. Highlights of the Year

Antoine ROUSSEAU and 5 co-authors released in 2014 the book *Brèves de Maths* [16]. This work (in french) selected more than 100 posts from the blog breves-de-maths.fr, in the framework of the international initiative "Mathematics of the Planet Earth". In this book (see cover 5), no complicated numbers, no weird equation, but short and clear sentences together with nice drawings to illustrate everyday life topics on our planet with the beauty of mathematics.

Figure 5. Brèves de Maths. Ed. Nouveau Monde, 2014

6.2. A Schwarz coupling method for dimensionally heterogeneous problem

Participant: Antoine Rousseau.

We study and analyze in [10] an efficient iterative coupling method for a dimensionally heterogeneous problem. We consider the case of 2-D Laplace equation with non symmetric boundary conditions with a corresponding 1-D Laplace equation. We first show how to obtain the 1-D model from the 2-D one by integration along one direction, by analogy with the link between shallow water equations and the Navier-Stokes system. Then, we focus on the design of an Schwarz-like iterative coupling method. We discuss the choice of boundary conditions at coupling interfaces. We prove the convergence of such algorithms and give some theoretical results related to the choice of the location of the coupling interface, and the control of the difference between a global 2-D reference solution and the 2-D coupled one. These theoretical results are illustrated numerically.

6.3. Bioremediation of water ressources

Participants: Antoine Rousseau, Alexis Pacholik.

Together with fellows from the MODEMIC team, we proposed a strategy for the bioremediation of water ressources such as lakes or lagoons. The originality is that the water treatment has to be done outside of the resource, in order not to dislocate its fragile ecological equilibrium.

The objective is to reach a targeted acceptable state for the resource within the minimal time. The patent [19] has been filed in connection with this work.

6.4. A well-balanced and positive preserving DG scheme for the SW equations

Participants: Arnaud Duran, Fabien Marche.

We consider in [5] the discontinuous Galerkin discretization of the nonlinear Shallow Water equations on unstructured triangulations. We propose an efficient combination of ingredients that leads to a simple high-order robust and well-balanced scheme, based on the alternative formulation of the equations known as the pre-balanced shallow water equations. We show that the preservation of the motionless steady states can be achieved, for an arbitrary order of polynomial expansion. Additionally, the preservation of the positivity of the water height is ensured using the recent method introduced in [59]. Some comparisons with a recent finite-volume MUSCL approach are also performed. The well-known tsunami test case shown in figures 6 and 7 has been computed here with high order DG scheme on unstructured triangulation.

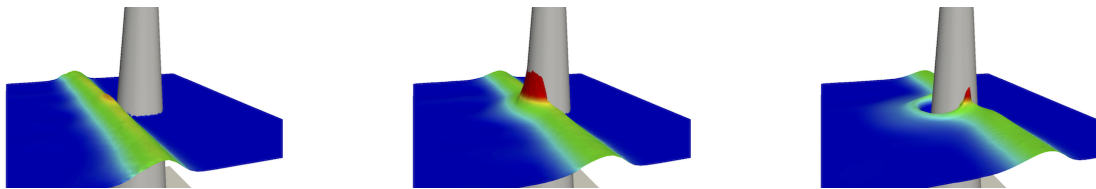*Figure 6. Tsunami wave over a conical island - Lateral view of the free surface at times t=5,6 and 7s.*
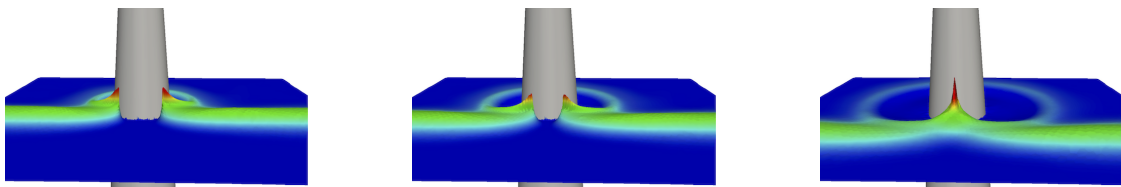

Figure 7. Tsunami wave over a conical island - Rear view of the free surface at times t=8,9 and 10s.

6.5. A well-balanced and positive preserving DG scheme for the GN equations

Participants: Arnaud Duran, Fabien Marche.

We introduce in [4] a discontinuous-Galerkin Finite-Element method to approximate the solutions of a new family of 1d Green-Naghdi models. These new models are shown to be more computationally efficient, while being asymptotically equivalent to the initial formulation with regard to the shallowness parameter. Using the free surface instead of the water height as a conservative variable, the models are recast under a pre-balanced formulation and discretized using a nodal expansion basis. Independently from the polynomial degree in the approximation space, the preservation of the motionless steady-states is automatically ensured, and the water height positivity is enforced. A simple numerical procedure devoted to stabilize the computations in the vicinity of broken waves is also described. The validity of the resulting model is assessed through extensive numerical validations.

6.6. A new class of fully nonlinear and weakly dispersive Green-Naghdi models

Participant: Fabien Marche.

We introduce in [8] a new class of two-dimensional fully nonlinear and weakly dispersive Green-Naghdi equations over varying topography. These new Green-Naghdi systems share the same order of precision as the standard one but have a mathematical structure which makes them much more suitable for the numerical resolution, in particular in the demanding case of two dimensional surfaces. For these new models, we develop a high order, well balanced, and robust numerical code relying on a hybrid finite volume and finite difference splitting approach. The hyperbolic part of the equations is handled with a high-order finite volume scheme allowing for breaking waves and dry areas. The dispersive part is treated with a finite difference approach. Higher order accuracy in space and time is achieved through WENO reconstruction methods and through an SSP-RK time stepping. Particular effort is made to ensure positivity of the water depth.

6.7. Upscaling transfer properties in heterogeneous porous media

Participant: Vincent Guinot.

In [9] the passive solute transport was studied in a periodic, artificial porous medium. A Laplace analysis of the breakthrough curves indicates that the widely used, classical Advection-Dispersion (AD) model cannot reproduce the contaminant transport features accurately. Neither can fractional dynamics-based, anomalous dispersion models. The models failing to reproduce the features of contaminant transport is shown to be due to the Fick-like, gradient-based operator used to represent dispersion, that induces infinite signal propagation speed, even when fractional models are used. The Laplace analysis shows that advection processes are predominant at all time and space scales. The size of the Representative Elementary Volume is shown to be 20 to 30 periods.

<div style="text-align:center">**MODEMIC Project-Team**</div>

6. New Results

6.1. Highlights of the Year

Yeasts play a central role in the wine making process. To study the yeasts in a stable environment and physiological state, a Multi-Stage Continuous Fermentor (MSCF) has been designed by the research Unit SPO (Sciences For Oenology). This device mimics the steps of the batch fermentation process. In this paper, the problem of the control of the sugar concentrations in each of the four reactors of the MSCF is considered. The cascade structure of the device leads to a constraint on the input flow rates (the control variables). A control strategy based on a linearizing control law coupled with a state observer and an anti windup component is proposed and finally implemented on the experimental process (see also 6.3.2).
BEST PAPER AWARD :
[41] **19th IFAC World Congress 2014**. C. CASENAVE, D. DOCHAIN, J. HARMAND, M. PEREZ, A. RAPAPORT, J.-M. SABLAYROLLES.

6.2. Mathematical models for microbial ecology

6.2.1. *Differential equations models*

Participants: Céline Casenave, Jérôme Harmand, Claude Lobry, Alain Rapaport, Alejandro Maximiliano Rojas.

Anaerobic digestion refers to the transformation of biodegradable material by micro-organisms in absence of oxygen (it can be found in waste-water treatments or industrial fermentation, and occurs naturally in soils). It receives an increasing consideration due to recent technological advances, but also because it is a source of renewable energy (bio-gas, fuel...). The anaerobic digestion is a complex set of bio-processes, for which there is a strong expectation of tractable models. We have proposed and studied new mathematical models that takes into account the following features:

- The available anaerobic digestion models used for control purposes do usually only consider soluble matter. In fact, part of the pollutants are not soluble but are under a particulate form. In order to establish whether adding the dynamics of such matter into the models is important for the system behavior or not, we have studied new models and established that depending on the kinetics of this additional reaction step, the qualitative behavior of the process may be significantly modified [26].

- Microbial food chains are present in anaerobic digestion where the different reaction steps can be seen as such: the waste products of the organisms at one trophic level (i.e. one reaction step) are consumed by organisms at the next trophic level (i.e. the next reaction step). In [55] we study a model of a two-tiered microbial 'food chain' with feedback inhibition, which was recently presented [63] as a reduced and simplified version of the anaerobic digestion model ADM1 of the International Water Association (IWA). It is known that in the absence of maintenance (or decay) the microbial 'food chain' is stable. In [63], using a purely numerical approach and ADM1 consensus parameter values, it was shown that the model remains stable when decay terms are added. In [55] we prove that introducing decay in the model preserves stability whatever its parameters values are and for a wide range of kinetics.

For the study of spatial heterogeneity in the models, we have carried on mathematical analyses of the properties of interconnected chemostats, in particular when growth rates present a substrate inhibition. In addition to the stabilizability properties discovered last year on "buffered" interconnections [33], we have studied this year yielding performances at steady state for the stabilizing configurations and characterized the set of the most efficient ones. For such configurations, we have shown that under certain circumstances, a "by-pass" of the main tank could be the best solution to ensure a global stabilization.

We have also analyzed two kinds of models, suited to specific characteristics of the microbial activity in soils:

- In [30], we have studied analytically and numerically a piece-wise linear model of carbon mineralization by two functional groups of micro-organisms in view of predicting the "priming effect" in soil ecosystems. The conclusion is that under a climate change, the augmentation of $C0^2$ will not affect to primary production and carbon storage when the plants are limited by nitrogen, but surprisingly a higher carbon input in soil should lead to a deplete of sequestered carbon and the increase of nitrogen release.

- In collaboration with Géosciences Rennes (Jean-Raynald de Dreuzy, Tristan Babey) and in the scope of the co-supervision of the PhD of Alejandro Rojas (also in the collaboration within the associated team with Chile), we investigate the equivalence between networks that represent interconnections of mobile/immobile zones in mass transfer models for soil ecosystems. For Structured INteracting Continua (SINC) models, that are described as the combination of a finite number of diffusion-dominated interconnected immobile zones exchanging with an advection-dominated mobile domain, we have proved an equivalence with Multi-Rate Mass Transfer (MRMT) and proposed a method for the identification of the equivalent MRMT model [14]. Moreover, we have shown the role of the controllability properties of a sub-system, in addition to the irreducibility of the network graph, for the input/output equivalence between several representations (work in preparation).

6.2.2. *Stochastic and hybrid discrete-continuous dynamical models*

Participants: Fabien Campillo, Bertrand Cloez, Coralie Fritsch.

Hybrid mass-structured chemostat models
Within the context of Coralie Fritsch thesis [12], we adopt a new modeling approach where instead on focusing on one type of model we propose different models and their interconnections, on the numerical viewpoint as well as the analytical one. Namely we propose an hybrid model of the chemostat where the population of bacteria is individually-based, each individual being described by its mass, and the subtract concentration is represented as a classic differential equation. We proved the convergence of this model in high population size toward an integro-differential system [20]. We proposed specific numerical schemes for the two approaches (see 5.3) [27].

Evolutionary invasion analysis and simulation for the chemostat
Still in the context of Coralie Fritsch thesis [12] and following her last year stay at the University of Helsinki in Otso Ovaskainen's Research Group, we consider an hybrid mass-structured mass-structured chemostat models with trait. The trait could for example be the factor of mass dissymmetry in the binary fission of a bacteria. In this context we proved an equivalence between invasion fitness for the hybrid-IBM model and the integro-differential system. We also numerically exhibit an evolutionarily singular strategy: with this given trait a monomorphic resident population cannot be invaded by a mutated population; the result is true for the two models.

PDE and stochastic models
In collaboration with M. Joannides and I. Larramendy-Valverde (I3M, University of Montpellier) we consider a stochastic growth model for which extinction eventually occurs almost surely. The associated complete Fokker–Planck equation describing the law of the process is established and studied. In dimension one, e.g. for the stochastic logistic model this equation combines a PDE and an ODE (paper under revision); in dimension two, e.g. for the stochastic chemostat model this equation combines a 2D PDE and a 1D PDE [22]. We then design a finite differences numerical scheme under a probabilistic viewpoint.

6.2.3. *Other modeling approachs*

Participants: Anne Bisson, Jérôme Harmand, Alain Rapaport.

A collaboration with the UMR Eco & Sols has led to the development of a (static) probabilistic model for inferring nature and number of interactions in communities assembly [29]. This model has brought new insights on a data set from reconstituted soil ecosystems. Because of the curse of dimensionality, we have begun this year to extend this approach to "assembling motifs" instead of considering all the possible assemblages (paper in preparation).

In [28], ecological trade-offs between species are studied to explain species coexistence in ecological communities. In our model, plant species compete for sites where each site has a fixed stress condition. Species differ both in stress tolerance and competitive ability. We derive the deterministic discrete-time dynamical system for the species abundances. We prove the conditions under which plant species can coexist in an stable equilibrium. We compare our model with a recently proposed, continuous-time dynamical system for a tolerance-fecundity trade-off in plant communities, and we show that this model is a special case of the continuous-time version of our model.

6.3. Analysis and supervision of bioprocesses

6.3.1. *Models development and identification*

Participants: Fabien Campillo, Amine Charfi, Yessmine Daoud, Jérôme Harmand, Sonia Hassam, Guilherme Pimentel, Alain Rapaport.

Membrane bioreactors combine a filtration process (with a membrane) and a suspended growth rate bioreactor. This recent technology present many advantages compared to conventional ones, but is more sophisticated and requires refined control because of the fouling process. We have proposed new modeling approaches of such bioreactors, where a fouling mechanism is explicitly described:

- Membrane bioreactors allow a perfect separation of biomass and treated waters. However, membranes are subject to clogging - also called fouling - by large organic molecules and solids (biomass, suspended solids, etc...). This phenomenon represents the main problem that limits the development of membranes bioreactors. It is well documented but very studies tried to formalize it under the form of dynamical model, even more under anaerobic conditions. This is what we did in the framework of Amine Charfi's PhD [11]: he contributed to a better understanding of membrane fouling dynamics and proposed a number of strategies to avoid irreversible clogging [24], [25].

- In the scope of the PhD supervision of Guilherme Pimentel, we have proposed a simple three time scales model in view of the control of the cake formation [13], [37]. This model has been validated on real data from a pilot plant at Univ. Mons (Belgium).

As already mentioned in 6.2.1 , anaerobic bioreactors are able to produce valuable energy. However, they are subject to destabilization in case of organic overload. It is thus necessary to develop appropriate models dedicated to the synthesis of stabilizing control feedbacks. Two strategies are followed to obtain such models. Either simple models are proposed from the knowledge we have about the process, either reduced models are obtained from more complex ones. Yessmine Daoud just began her PhD thesis following the first strategy: more precisely, we study simple inhibition models of the anaerobic digestion and tries to establish simple equivalence between these models and the well known ADM1 model. Sonia Hassam, in her PhD, works within the framework of the second strategy to propose simple models obtained in reducing complex ones like the ADM1.

As already proposed last year, we have set a methodology to identify from data observed on a chemostat plant a multi-specific model that suits better than a mono-specific one, when data from molecular biology are available. In [35], we combine molecular fingerprints obtained at some discrete times (such as the ones provides by the DGGE or SSCP techniques) with on-line macroscopic measurements. In a similar spirit, within the framework of a collaboration with the LOMIC at Banyuls, we have analysed molecular data obtained via SSCP technique to monitor the structure of microbial communities. It was shown that aerial transport of bacteria from desert into the sea influenced its bacterial diversity [32], [31].

As far as stochastic approaches are concerned, the thesis of Mohsen Chebbi aims at developing stochastic models of membrane bioreactors following the approach proposed in [57]. A mathematical and simulation framework has been established, as well as the bases of vectorial simulation techniques in `Matlab`. Developments of Monte Carlo techniques for the identification of bioprocesses are investigated in the thesis of Oussama Hadj-Abdelkader, improving classical particle filtering approaches (sequential Monte Carlo) by integrating MCMC (Monte Carlo Markov Chain) procedures. A software approach has been adopted in `C++`.

6.3.2. Synthesis of control laws

Participants: Térence Bayen, Walid Bouhafs, Céline Casenave, Amel Ghouali, Jérôme Harmand, Zeyneb Khedim, Claude Lobry, Alain Rapaport, Victor Riqueleme, Matthieu Sebbah.

We investigate two kinds of bioprocesses to be controlled, arising in industrial biotechnology (digesters, wastewater purification...) or in the bioremediation of natural environments (lakes, landfill...).

6.3.2.1. Industrial biotechnology

As it is often the case in industry, we distinguish two kinds of process operating: continuous processes, for which the volumes of the bio-reactors are constant, and fed-batch processes, for which the filling rate is the control.

We tackle several optimal control problems related to the maximization of productivity of continuous bioprocesses:

- As far as anaerobic digestion is concerned, we have considered an optimal control problem for the chemostat model with substrate inhibition. The originality of this problem relies on the fact that the quantity to be maximized is not simply a state of the model (the substrate or the biomass) but the volume of biogas produced (which is a nonlinear function of the state) within a given time interval. Optimal strategies have been proposed for a class of initial conditions of the system (PhD thesis of Amel Ghouali). Other optimal control problems are studies by Walid Bouhafs to establish the optimality of controls initially proposed by Djalel Mazouni, a former PhD candidate supervised by the team. Using a tricky projection of the problem into another state space, he has shown that the problem of optimizing the degradation of two different substrates by two antagonist bacterial communities could be solved in a very general sense as the minimization of a specific functional for a very large class of nonlinear systems [19].

- The work [15] arises in the context of selection of species (widely used in agriculture and biotechnology in order to improve productivity). For microorganisms, the selection process can be based on genetic tools. Our methodology in [15] is to drive the competition between species in a chemostat. We consider a two species chemostat model with one limiting substrate, and our aim is to optimize the selection of the species of interest. Thanks to the Pontryagin Maximum Principle, we introduce a singular feeding strategy which allows to reach the target, and we prove that the feedback control provided by this strategy is optimal. The optimal synthesis of the problem in presence of more than two species will be investigated in a future work.

- In [38], we study the problem of minimal time for a chemostat system with one limiting substrate and one species. Given a target point, the problem consists in finding an optimal feeding strategy steering any initial condition of the system to this target. This is typically of interest whenever the input substrate concentration changes yielding in a new steady state. We consider the case where the growth rate function is of Haldane type implying the existence of a singular arc that is non-necessary admissible everywhere. We provide an optimal synthesis of the problem using tools from optimal control theory.

- The work [16] studies the coupling of a culture of micro-algae limited by light and an anaerobic digester in a two-tank bioreactor (the model combines a periodic day-night light for the culture of micro-algae and a classical chemostat model for the digester). We first prove the existence and attraction of periodic solutions of this problem for a one day period. Then, we study the optimal control problem of optimizing the production of methane in the digester during a certain time frame, the control on the system being the dilution rate (the input flow of micro-algae in the digester). We also investigate the dependence of the optimal cost with respect to the volume ratio of the two tanks.

In collaboration with researchers of the unit SPO (Sciences For Oenology), we have proposed a control law of a multi-stage continuous fermentor (MSCF) designed for the study of the wine fermentation, that has been implemented on the experimental process [41]. We have also finalized the controllability analysis and minimal time feedback synthesis of models of cascade of continuous bioreactors under input constraints [18], [40] (that is also related to the control of MSCF).

As far as fed-batch processes are concerned, an extension of former results of the team about the minimal time control of fed-batch processes with impulse controls is presented in [34],

The paper [17] is devoted to the study of the minimal time problem of a fed-batch reactor, under the presence of a saturation point on the singular locus (this typically occurs whenever the growth rate function is of Haldane type and when typically the maximum input flow rate is not high enough to maintain the substrate concentration constant). This brings non-intuitive issues for the optimal synthesis (existence of switching curve and point of prior saturation).

6.3.2.2. Bioremediation of natural environments

In the scope of the associated team with Chile and the supervision of the postdoctoral stay of Matthieu Sebbah in Chile, we have addressed a new model of landfill remediation when controlling the leachate recirculation [53]. We have applied the same methodological approach than the one for the work [18], [40] mentioned in the previous section, which consists in characterizing first the sub-domains for which the target can be optimally reached with a constant extreme control (no recirculation or maximal speed of recirculation), and further the nature of optimal commutations outside these sets. This analysis provides information for the practitioners on the benefit to implement sensors and real-time controllers.

Also in the scope of the associated team with Chile (see 7.3.2.1) and the co-supervision of the PhD of Victor Riquelme, we have carried on the study of optimal syntheses for the minimal time treatment of natural water reservoirs (such as lakes) [52]. We have proved that the minimal time strategy consists in a most-rapid approach to homogeneous concentrations, even though the optimal control problem is non convex. Moreover, we have shown that a large diffusion increases the treatment time when the resource is everywhere highly polluted , while it can at the opposite be beneficial when only part of the resource is polluted. This feature should serve the practitioners in the choice of pumps positioning in a originally clean water resource that is suddenly affected by a local pollution. This work is in connection with the INRA/Inria patent [47] that has been deposited jointly with LEMON Team.

6.4. Other application domains

Participants: Fabien Campillo, Céline Casenave.

Semi-Markov land use dynamic
With IRD (GRED Montpellier) and the Univ. of Fianarantsoa (Madagascar) we pursued our study on land use dynamics models corresponding to parcels located on the edge of the forest corridor, Madagascar. We use semi-Markov chain to infer the land-use dynamics. In addition to the empirical and maximum likelihood methods, we estimate the semi-Markov kernel by a Bayesian approach [21].

Ice cream crystallization
We study the problem of the control of an ice cream crystallization process, part of the European CAFE project, in collaboration with CESAME (Univ. Catholique de Louvain-la-neuve), Irstea Antony and AgroParisTech. The goal is to control the viscosity of the ice cream at the outlet of the continuous crystallizer. On the basis of a population balance equation describing the evolution of the crystal size distribution of the ice cream, and an energy balance equation, we have proposed an input-output reduced order model of the process, that has been identified and validated on experimental data [23]. A nonlinear control strategy based on an adaptive linearizing control law coupled with a Smith predictor to account for the measurement delay has been proposed, and validated on the experimental pilot plant (paper in preparation).

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

5. New Results

5.1. Highlights of the Year

- Laure Blanc-Féraud was General Program chair of the conference IEEE ISBI 2014 in Beijing.

5.2. Sparse 3D reconstruction in fluorescence imaging

Participants: Emmanuel Soubies, Laure Blanc-Féraud, Sébastien Schaub, Gilles Aubert.

Sparse reconstruction Super-resolution microscopy techniques allow to overstep the diffraction limit of conventional optics. Theses techniques are very promising since they give access to the visualisation of finer structures which is of fundamental importance in biology. In this work we deal with Multiple-Angle Total Internal Reflection Microscopy (MA-TIRFM) which allows reconstructing 3D sub-cellular structures of a single layer of $\sim 300\ nm$ behind the glass coverslip with a high axial resolution. The 3D volume reconstruction from a set of 2D measurements is an ill-posed inverse problem and requires some regularization. Our aim in this work is to propose a new reconstruction method for sparse structures that is robust to Poisson noise and background fluorescence. The sparse property of the solution can be seen as a regularizer using the ℓ_0-norm. Let us denote $f \in R^N$ the unknown fluorophore density, then the problem states as

$$\widehat{f} = \arg \min_{f \in R^N} \left(J_d(f) + \lambda \|f\|_0 \right) \tag{1}$$

where J_d is defined from the likelihood function of the observation given f, $\lambda > 0$ is a weight parameter and $\| \cdot \|_0$ denotes the ℓ_0-norm (which counts the number of nonzero components of f). In order to solve this combinatorial problem, we propose a new algorithm based on a smoothed ℓ_0-norm allowing minimizing the non-convex energy (1). Following [20], the idea is to approach the ℓ_0-norm by a suitable continuous function depending on a positive parameter and tending to the ℓ_0-norm when the parameter tends to zero. Then the algorithm solves a sequence of functionals which starts with a convex one (on a large convex set) and introduce progressively the non-convexity of the ℓ_0-norm (Graduated Non Convexity approach). Figure 1 shows the accuracy of the method on a simulated membrane.
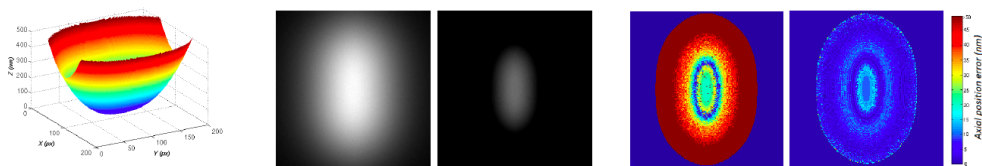
Figure 1. From left to right: Simulated membrane, Microscope acquisition (numerical simulations) with two different incident angles. The two images on the right represent position errors (nm) in the axial direction of the reconstructed membrane obtained with different algorithms: Richardson-Lucy algorithm without regularization (left) and our algorithm with $\lambda = 0.001$ (right).

Axial profile calibration In order to turn on real sample reconstructions we need to perform a calibration of the TIRF microscope. Its principle is based on an evanescent wave with an exponential theoretical decay. However this decay is generally not a pure exponential in practice and we need to have a good knowledge about it. Then based on a phantom specimen of known geometry (bead) we are working on a method to estimate experimentally/numerically this decay profile and calibrate all parameters of the system.

5.3. Penalty analysis for sparse solutions of underdeterminated linear systems of equations

Participants: Emmanuel Soubies, Laure Blanc-Féraud, Gilles Aubert.

In many applications such as compression to reduce data storage, compressed sensing to recover a signal from fewer measurements, source separation, image decomposition and many others, one aims to compute a sparse solution of an underdetermined linear systems of equations. Thus finding such sparse solutions is currently an active research topic. This problem can be formulated as a least squares problem regularized with the ℓ_0-norm. We consider the penalized form

$$\widehat{\mathrm{x}} = \arg \min_{\mathrm{x} \in \mathrm{R}^N} \left(\frac{1}{2} \|A\mathrm{x} - d\|^2 + \lambda \|\mathrm{x}\|_0 \right) \tag{2}$$

where $A \in \mathrm{R}^{M \times N}$, $d \in \mathrm{R}^M$ represents the data and $\lambda > 0$ is an hyperparameter characterizing the trade-off between data fidelity and sparsity.

It is well known that reaching a global solution of this $\ell_2 - \ell_0$ functional is a NP-hard combinatorial problem. Besides the non-convexity of this 'norm', its discontinuity at zero makes the minimization of the overall functional a hard task. In this work we focus on non-convex continuous penalties widely used to approximate the ℓ_0-norm which usually lead to better results than the classical ℓ_1 convex relaxation since they are more 'ℓ_0-like'. Based on some results in one dimension, we propose the Exact ℓ_0 penalty (El0). In one dimension and when the matrix A is orthogonal, replacing the ℓ_0-norm in (2) by this penalty gives the convex hull of the overall function. Then we have proved, for any matrix $A \in \mathrm{R}^{M \times N}$, that the global minimizers of the $\ell_2 -$ El0 objective function are the same as for the $\ell_2 - \ell_0$ functional. We also demonstrate that all the local minimizers of this approximated functional are local minimizers for $\ell_2 - \ell_0$ while numerical experiments show that the reciprocal is in general false and that the objective function penalized with El0 admits less local minimizers than the $\ell_2 - \ell_0$ functional. Then, this work provides in some way an equivalence between the initial $\ell_2 - \ell_0$ problem and its approximation using the El0 penalty. One can address problem (2) by replacing the ℓ_0-norm with the El0 penalty which provides better properties for the objective function although the problem remains non-convex. Recently, some authors have proposed algorithms and proved their convergence to critical points of non-smooth non-convex functionals like $\ell_0 -$El0. Based on such algorithms, we propose a macro algorithm and prove its convergence to a (local) minimizer of the initial $\ell_2 - \ell_0$ functional.

5.4. Motion compensation in two-photon microscopy temporal series

Participants: Caroline Medioni, Grégoire Malandain, Florence Besse, Xavier Descombes.

Acquisitions of 3D image sequences over long period of time, in particular, have enabled neurobiologists to follow complex processes such as the development of neuronal populations or degenerative events occurring in pathological contexts, improving our understanding of the mechanisms involved in brain development and function. In most cases, live samples are moving/growing during long-term imaging. Therefore it is required to compensate for this global 3D motion before measuring the dynamics of the structure of interest. We have proposed a method to compute a coherent 3D motion over a whole temporal sequence of 3D volumes (Figure 2), which is able to capture subtle sub-voxelic displacements.

5.5. Axon Growth Imaging and Modeling

Participants: Agustina Razetti, Caroline Medioni, Florence Besse, Xavier Descombes.

The modeling part of this work has been made in collaboration with S. Komech, E. Pechersky and E. Zhizhina from IITP (Russian Academy of Science)
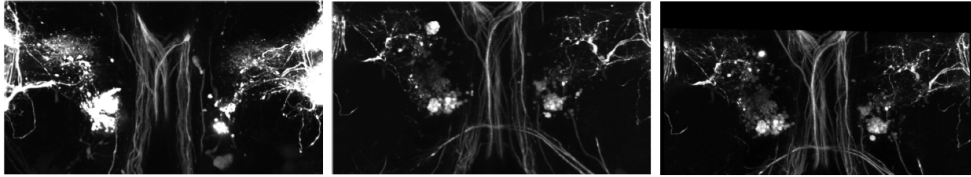
Figure 2. Left and middle: mip views of both the first and the last volumes of a temporal series. Right: mip view of the last volume after motion compensation.

In Drosophila brain, at metamorphosis, Mushroom Body gamma neurons undergo axonal remodeling characterized by a pruning of larval branches followed by regrowth and branching/arborization of adult processes. Axonal regrowth at this stage is essential to consolidate the adult brain and its success is determined by the trajectories followed by the axons and their branches. These trajectories depend on both extracellular guidance signals, and on a complex internal molecular machinery capable to read these signals and act in consequence. F. Besse's team at the IBV Institute has identified genes involved in this regrowth and branching processes [19]. A better understanding of the role of these genes will help to unravel the molecular mechanisms behind these fundamental processes, and lead to a better understanding of the neuronal morphology in both healthy and pathological conditions.

During this PhD project, mathematical and computational tools will be developed to characterize and compare the axonal regrowth and branching dynamics. Different populations of regrowing gamma axons will be considered (i.e. wild type and presenting mutations in relevant genes). The study will be based on both static 3D confocal images of axonal trees, and two-photon in vivo 4D image sequences showing either a single GFP positive regrowing axon or the entire population of regrowing gamma axons marked with GFP. In a first part we have begun to collect data. This includes three parts: i) fly stock maintenance, crossing and selection; ii) sample preparation consisting in pupal brain dissection, medium preparation and sample final assembly; iii) imaging: using different microscopy techniques, eg.confocal/two-photon, microscopy and light sheet microscopy, and acquisition of 4D image sequences. Thanks to an imaging technique developed by C. Medioni in F. Besse's laboratory, we have been able to monitor axonal regrowth and branching at early steps in individual neurons, and to follow them for about 15 hours using the two-photon microscope. These movies will be used in the future to develop the mathematical modeling of axonal regrowth/branching process (see figure 3). Our early works concerning modeling have consisted in investigating some models based on continuous time random walks and characterizing the main axon branch through topological entropy [8] [22].

5.6. Markov Chain for Axon Growth Modeling

Participants: Alejandro Mottini, Xavier Descombes, Florence Besse.

In this work we have defined a 2D discrete stochastic model for the simulation of axonal biogenesis [8]. The model is defined by a third order Markov Chain. The model considers two main processes: the growth process that models the elongation and shape of the neurites and the bifurcation process that models the generation of branches. The growth process depends, among other variables, on the external attraction field generated by a chemoattractant molecule secreted by the target area.

For the validation, we have fluorescently labeled single neurons within intact adult Drosophila fly brains, and have acquired 3D fluorescent confocal microscopy images of their axonal trees. Both normal neurons and neurons in which the function of the imp (mutant type 1) or profilin (mutant type 2) genes was inactivated were imaged. imp encodes a conserved RNA binding protein controlling subcellular mRNA transport and local protein synthesis, and is essential for axonal remodeling. profilin encodes a regulator of the actin cytoskeleton involved in axonal pathfinding. Mutations in these two conserved genes have been linked to neurological pathologies.
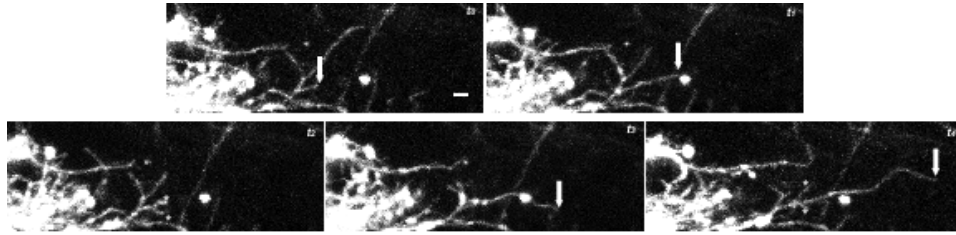
Figure 3. Images extracted from a movie recorded on a two-photon microscope (maximum intensity projection): single axon at the regrowing stage within Drosophila brain. Axons are marked with GFP. t0-t1: elongation step, t2: retraction and branching event, t3: elongation and t4: elongation and branching step. Arrows shows axonal tips and asterisks, the formation of branches. Scale bars for each image: 10μm.

Each image stack has a resolution of $0.093967 \times 0.093967 \times 0.814067 \mu m$ and two channels. The morphology of single axonal trees is visible in the first channel and was manually segmented by an experienced biologist. The morphology of the overall neuronal structure in which axons are developing is visible in the second channel. In total, 53 images (18 normal, 21 type 1 mutant and 14 type 2 mutant) were used. In order to study the attraction field of the populations, all stacks were registered against the first image of the normal population. This was performed using the second channel of each image.

We then have estimated the model parameters to generate two fields for each population, a scalar field that represents the axon flexibility and a vector field that represents the attraction field. Since we obtain some estimates on a sparse set of points in the x, y plane, we extrapolate the fields using a Gaussian Markov Random Field. By qualitatively analyzing the resulting images we have determined that there is no relevant difference on the attraction field between the three populations. We have observed that the field points towards the target area and that its norm is stronger at the starting point of the axons and weaker near the target area, which is consistent with biological expectations. The same procedure was used to analyze the difference on the scalar fields for each population (see Figure 4). In this case we can detect a significant difference between the populations.
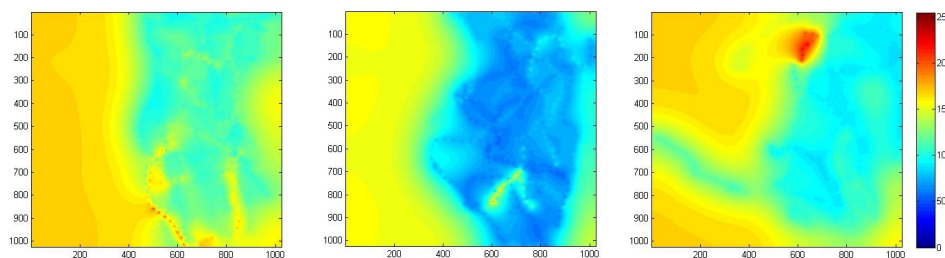

Figure 4. Markov Chain parameter scalar field for the normal (left), mutant type 1 (middle) and mutant type 2 (right) populations.

5.7. Cells detection using segmentation competition

Participants: Emmanuelle Poulain, Emmanuel Soubies, Sylvain Prigent, Xavier Descombes.

Image segmentation has been widely investigated in particular in the context of bioimaging for cells detection. In some cases, the background is clearly identifiable so that a binary mask of the objects can be computed using simple techniques such as thresholding. Therefore, isolated objects are easily recognizable while splitting clusters of objects, which are connected components in the binary mask, remains a challenging task. In fluorescent microscopy devices used for live imaging – *e.g.* confocal, biphoton, Selective Plane Illumination Microscope (SPIM) – an additional difficulty comes from the multiple degradations of the acquired images such as strong noise, spatially varying blur and light attenuation which makes the segmentation a hard task even for selecting a suitable threshold for the background. Since many years, researchers have developed several methods to perform such segmentation. An efficient approach consists in generating seeds that define regions using geometric information through a distance, as in the markers controlled watershed algorithm [21], or image gradient for the active contour approach. These approaches give accurate results providing the seeds are well chosen that is still an open issue. Bayesian approaches, such as marked point process, avoid this bottleneck by selecting randomly generated shapes through the minimization of an energy function. However, they are restricted to low dimensional parametric shapes, such as disks or ellipses, due to computational issues. Tuning the parameters of the segmentation algorithms mentioned above in order to obtain accurate results on the whole image can also be extremely tricky whereas it is much easier to obtain accurate results on different parts of the image using different sets of parameters. To overcome these limits we propose to combine both approaches by generating shapes from state of the art segmentation algorithms using random seeds and/or different sets of parameters. These shapes define a dictionary of candidates from which a competition process, using the Multiple Birth and Cut algorithm, extracts the most relevant shapes. We have validated this selection approach on synthetic data and on a multicellular tumor spheroid slice by comparing the obtained results with two different state of the art segmentation methods to build the dictionary of shapes and compare the performance of our competition approach with the ImageJ particle analyser (see figure 5).
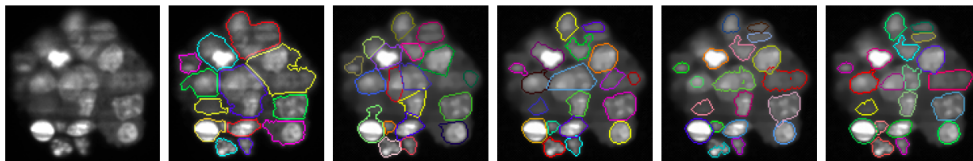
Figure 5. Spheroids of tumor cells stained with a fluorescent nuclear marker (top line left). Fiji Particle Analyzer result (top line right). Results of our approach with dictionary generated by RS-SKIZ with 800 repetitions and a threshold value fixed to 70 (middle line left) MTRS-SKIZ with 80 repetitions and 10 different thresholds (middle line right) MPRS-FM with 100 repetitions and 8 different values for the Fast Marching parameter used to stop the expansion (bottom line left), concatenation of MTRS-SKIZ and MPRS-FM dictionaries (bottom line right).

5.8. Graph cut and attractive interactions

Participants: Tarun Yellamraju, Emmanuel Soubies, Sylvain Prigent, Xavier Descombes.

Marked point processes have proved to be very efficient for segmenting a collection of objects from digital images. The multiple birth and death algorithm provides an optimization framework that allows reasonable time computation. This algorithm has to be embedded in a simluated annealing framework which involves parameters tuning (initial temperature and cooling scheme). This tedious task can be overcame considering a graph cut algorithm instead of the death step. The algorithm then consists in successively adding new random objects in the configuration and selecting the most relevant using the graph cut algorithm. In the graph construction a node is associated to each object. Unfortunately, the regularity condition imposed by the graph cut prevents to consider attractive interactions such as clustering or alignment constraints, which restricts the model to repulsive properties such as non overlap between objects. To overcome this restriction

we have investigated new graph constructions by considering nodes defined by clusters of interacting objects. Different strategies have been compared to avoid being tracked in local minima defined by clusters while minimizing the number of required iterations. First results have been obtained on a seeds detection problem (see figure 6).
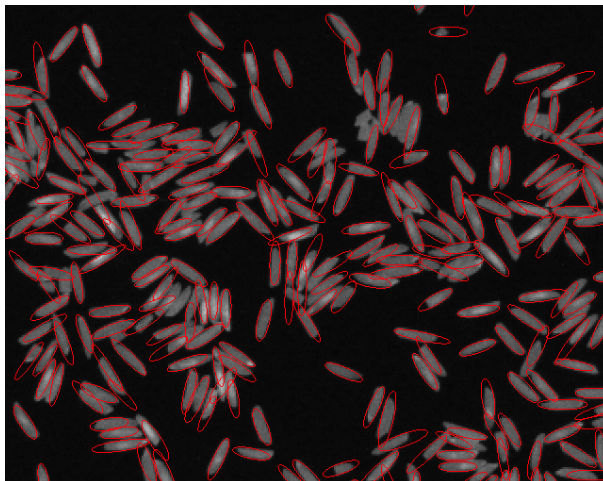
Figure 6. Rice seeds detection using a marked point processes and a birth and cut algorithm (Rice image has been given by Alpha MOS and LAAS-CNRS).

5.9. Cell-to-cell ascidian embryo registration

Participants: Gaël Michelin, Grégoire Malandain.

This work is made in collaboration with Léo Guignard and Christophe Godin (Virtual Plants) and Patrick Lemaire (CRBM), within the Morphogenetics Inria Project Lab.

Recent microscopy techniques allow imaging temporal 3D stacks of developing organs or embryos with a cellular level of resolution and with a sufficient acquisition frequency to accurately track cell lineages. Imaging multiple organs or embryos in different experimental conditions may help to decipher the impact of genetic backgrounds and environmental inputs on the developmental program. For this, we need to precisely compare distinct individuals and to compute population statistics. The first step of this procedure is to develop methods to register individuals.

From a previous work of cell segmentation from microscopy images [6], we propose an approach to extract the Left-Right symmetry plane of embryos at early stages (Figure 7). Then we use the symmetry information to both register these embryos at a similar developmental stage and obtain a cell-to-cell mapping. We assessed the symmetry plane extraction on more than 100 images from 10 individuals between 32-cells and late-neurula development stage. The cell-to-cell registration was performed on 5 distinct individuals at 64-cells and 112-cells stage (Figure 8).

5.10. Quantitative comparison of micro-vasculatures

Participants: Manon Linder, Grégoire Malandain.

This work is made in collaboration with Cécile Duplaa and Thierry Couffinhal (INSERM).
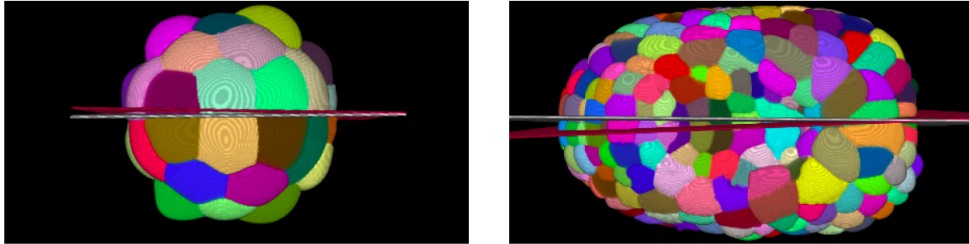
Figure 7. Left-Right symmetry plane initialization (red) and final estimation (white) on (left) a 32-cells stage embryo and (right) a neurula stage embryo.
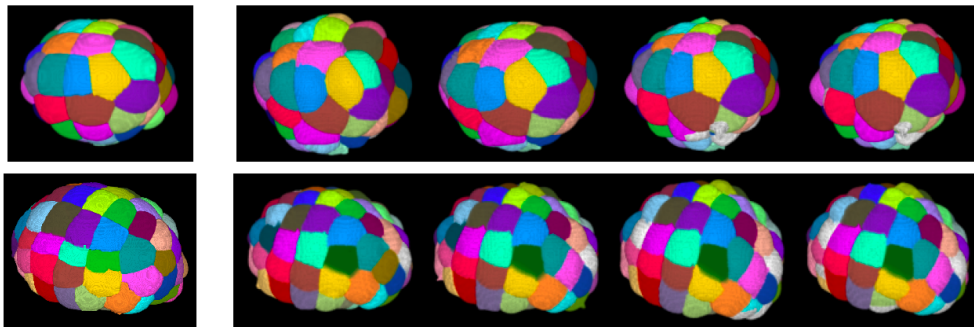
Figure 8. Cell-to-cell mapping between reference image (left) and test images (right) at 64-cells stage (first line) and at 112-cells stage (last line). The reference images are taken from the same individual, the test images are taken at different time-points of a second individual. On the test images, white cells are those that have not been matched to a reference cell.

Angiogenesis is a key component of ontogenesis, but also of tumor development or in some pathology repair (i.e. ischemia). Deciphering the underlying mechanisms of vessel formation is of importance. We aimed at identifying and characterizing the genetic components that are involved in this development. This requires to compare the effect of each gene with respect to the others, hence appeals for quantitative comparisons. We developed a methodology that first transforms a vascular image into a tree and second quantitatively analyze 3D vascular trees (see Figure 9) We conduct real experiments with images of the renal arterial network of different mutant mice, through the development of quantitative measurements that allow for group study.
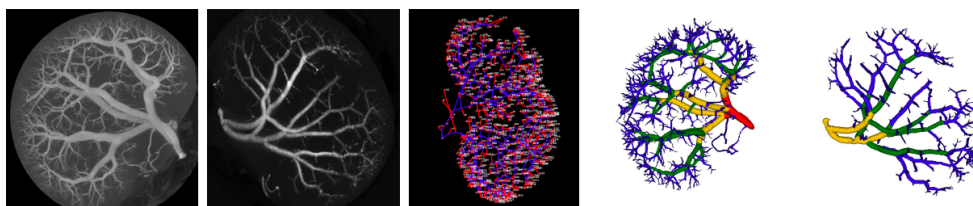
Figure 9. From left to right: two mip views of mouse kidneys acquired with a micro-CT (control and Fzd4 & Fzd6 KO mouse); a labeled tree built from an image; Diameter-defined Strahler classification of arterial trees.

5.11. Pre-clinical molecular dynamic image analysis: 99mTc- pertechnetate biodistribution model of murine stomach with micro-SPECT

Participants: Marine Breuilly, Grégoire Malandain.

This work is jointly conducted with Thierry Pourcher, Jacques Darcourt, Philippe Franken, Kaouthar Chatti, and Philippe Pognonec from the Transporter in Imagery and Oncologic Radiotherapy team (TIRO, CEA-CAL-UNSA).

This project investigates the potential retention of iodide in the stomach, for a better understanding of the iodide biodistribution in the body and more precisely of its potential antiseptic role. To that end, we study the uptake of the ^{99m}Tc-pertechnetate (an iodide analog) within the murine stomach observed thanks to a SPECT camera. Using the coupled SPECT and CT device dedicated to small animals, functional information targeted by a specific radiotracer (^{99m}Tc-pertechnetate) can be imaged dynamically.

The temporal evolution of the uptake is analysed thanks to a dedicated multi-compartment model. The addressed challenges consist in 1) estimating the time-activity curves for the different compartments, and 2) identifying the model parameters.

- ^{99m}-pertechnetate is an iodide analog regarding to the NIS gene. Thus iodide uptake kinetics can be studied through the study of ^{99m}Tc- pertechnetate biodistribution.
- Dynamic SPECT images exhibit a progressive accumulation of ^{99m}Tc-pertechnetate in the stomach wall and diffusion in the stomach cavity.
- The workflow that has been previously proposed in [18] was tested on a larger dataset of five subjects, yielding promising results: The computed model parameters are coherent, and the computed parameter values suggested that there is some iodide retention in the stomach wall.
- A comparison of the dedicated method for extraction of time activity curves with the ones extracted with Pixies software is on-going.
- A comparison of the dedicated method for solving the inverse problem of the compartmental analysis with methods developed by the Turku PET Centre is on-going.

5.12. Massal motility measures to automatically predict fertility scores

Participants: Ana Rita Lopes Simoes, Eric Debreuve.

This work has been done in the scope of the ANR project MOTIMO. We developed a method for automatic scoring of sperm samples in order to predict fertility for the farming industry. The method was applied to samples from rams and goats. A given sample is a video composed of a hundred frames (see Fig. 10).

We analyzed video samples acquired according to four modalities or protocols: drop, chamber, fluorescent beads with a 4x zoom, and fluorescent beads with 10x zoom. Two options have been considered. (1) An optical flow method has been applied to the videos in order to estimate the apparent motion of the seminal fluid (see Fig. 10). Some statistical features of interest (such as entropy) were extracted from the obtained motion fields in order to characterize the sperm massal motility. (2) The second option consisted of tracking the beads on the 4x or the 10x fluorescence videos (see Fig. 10). Some features of interest were also extracted from the resulting bead trajectories. Then using either of these feature sets (obtained with options (1) and (2)), a regression analysis (linear and kernel SVM) was conducted on a subset of the available videos (the learning set) so as to define a prediction function taking features as input and outputting a fertility score. The scores computed by this function were compared to scores assigned by experts. We used the coefficient of determination (commonly denoted by R^2) as a performance measure of the learned prediction function. The best results were obtained using tracking on the ram videos showing fluorescent beads with a 4x zoom ($R^2 = 0.9$). The results on the goat videos were not as satisfying ($R^2 \simeq 0.65$), but the partner providing the videos expressed some concerns about the quality of the acquisition campaign for these data.
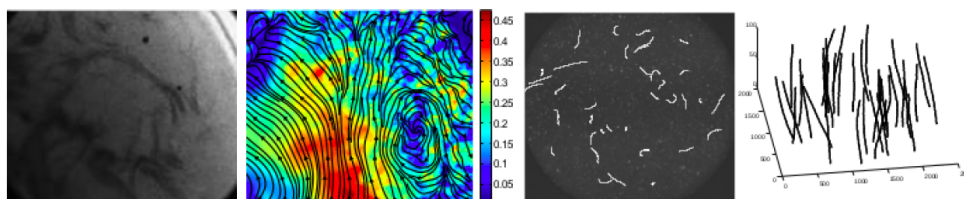
Figure 10. Predicting fertility based on massal motility measures. (Left) One frame of a video for the "drop" modality; (Middle) An example of computed optic flow; (Right: image+plot) An example of bead trajectories.

5.13. Sample selection for SVM learning on large data sets

Participants: Sonia Chaibi, Xavier Descombes, Eric Debreuve.

Support Vector Machines (SVM) represent a popular framework of supervised learning. However, it is not well adapted to large data sets since learning is performed by an optimization procedure involving the whole data set. Yet, in the end, only a small subset of the samples (the so-called support vectors) is retained for prediction. Of course, efficient algorithms exist. Still, it can be interesting to filter out as many samples as possible (the ones that will surely not be part of the support vectors) before initiating the learning procedure.

Sonia Chaibi, a PhD student from UBMA, Algeria, visited the team for a month to collaborate on this subject. The method relies on successive unsupervised sample clustering steps. After each clustering, the homogeneity of the clusters in terms of sample class assignment is used to decide which samples are unlikely to be close to the separation hyperplane (and hence unlikely to be selected as support vectors), and which samples are apparently close to this hyperplane. The former ones can be discarded, thus reducing greatly the number of samples to be processed by the SVM algorithm, while the latter ones are kept, preserving the precision of the separation hyperplane as much as possible.

5.14. Morphological Analysis and Feature Extraction of Neurons from Mouse Cortices Multiscale 3D Microscopic Images

Participants: Alexis Zubiolo, Xavier Descombes, Eric Debreuve.

This work is jointly conducted with Kawssar Harb and Michèle Studer (iBV).

We propose a framework to analyze the morphology of mouse neurons in the layer V of the cortex from 3D microscopic images. We are given 8 sets of images, each of which is composed of a 10x image showing the whole neurons, and a few (2 to 5) 40x images focusing on the somas. The framework consists in segmenting the neurons on both types of images to compute a set of specific morphological features, and in establishing the correspondence between the neurons to combine the features we obtained, in a fully automatic fashion. On the 10x images, we use a multiple birth and cut algorithm to segment the sections of the apical dendrites. Merging these intersections provides the localization of the first branching of the apical dendrite (see Fig. 11 (left)). On the 40x images, we compute an hysteresis threshold to obtain a first segmentation (somas and dendrites starts) and apply iterative morphological operators to reconstruct the full dendrites (see Fig. 11 (middle)). The correspondence map between the two types of images is done using a bipartite graph matching model that associates each neuron configuration of a 40x image – a constellation – to a subset of neurons in the 10 image – the galaxy – (see Fig. 11 (right)).
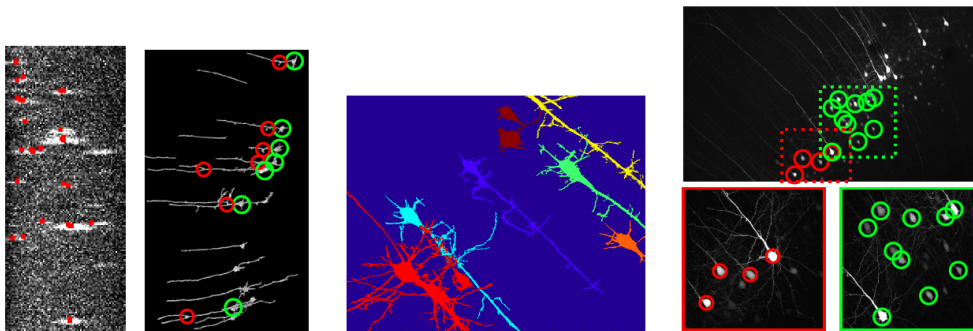
Figure 11. Left: neuron segmentation on the 10x image. Middle: full neuron reconstruction from the 40x image. Right: 10x-40x image maching.

5.15. Whole-Slide Image Analysis of Renal Cell Carcinoma

Participants: Ana Rita Lopes Simoes, Eric Debreuve, Alexis Zubiolo, Xavier Descombes.

This work is jointly conducted with Thierry Pourcher, and Philippe Pognonec (TIRO, CEA-CAL-UNSA), and Damien Ambrosetti (CHU, Nice).

We study histology images of kidney cancer that present different subpopulations of cells (tumor, healthy tissue, stroma, fat, blood, ...). The goal is to analyze the images to help determine the cancer type and stage. Given the resolution of the images ($0.25\mu m$) that leads to very large images (around 100k×100k pixels), a multiscale approach has been considered. At a larger scale, we focus on the cellular architecture and the vascular networks. Regions of interest (ROIs) have been detected with a pixelwise clustering based on neighborhood features (see Fig. 12 (left)). At a smaller scale, we extract more precise information from the cells (nucleus and cytoplasm sizes, shapes and colors, ...). The nuclei of the cells have been segmented using an Hessian determinant-based method (see Fig. 12 (middle)) which enables us to establish statistics about their size. Information on the vascular arborization has been extracted with a Frangi vesselness followed by a cleaning and gap filling post-processing (see Fig. 12 (right)).
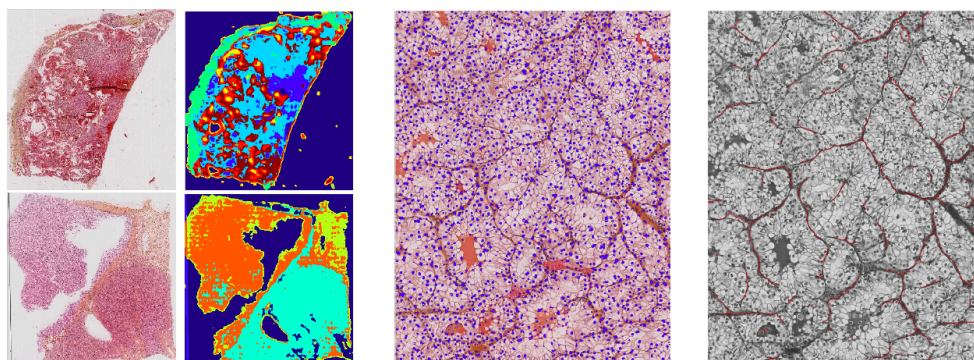
Figure 12. ROI detection (left), Nuclei segmentation in blue (middle) and Vascular arborization extraction in red (right)

NEUROMATHCOMP Project-Team

5. New Results

5.1. Highlights of the Year

Olivier Faugeras received the Okawa prize for his pioneering contributions for computer vision and for computational neuroscience. The ceremony will be held in Tokyo in March 2015.

5.2. Neural Networks as dynamical systems

5.2.1. *Heteroclinic cycles in Hopfield networks*

Participants: Pascal Chossat, Martin Krupa.

Learning or memory formation are associated with the strengthening of the synaptic connections between neurons according to a pattern reflected by the input. According to this theory a retained memory sequence is associated to a dynamic pattern of the associated neural circuit. In this work we consider a class of network neuron models, known as Hopfield networks, with a learning rule which consists of transforming an information string to a coupling pattern. Within this class of models we study dynamic patterns, known as robust heteroclinic cycles, and establish a tight connection between their existence and the structure of the coupling.

This work is available as and has been submitted to a Journal.

5.2.2. *Periodic forcing of stabilized E-I networks: Nonlinear resonance curves and dynamics*

Participants: Romain Veltz, Terry Sejnowski [Salk Institute].

Inhibition stabilized networks (ISNs) are neural architectures with strong positive feedback among pyramidal neurons balanced by strong negative feedback from in-hibitory interneurons, a circuit element found in the hippocampus and the primary visual cortex. In their working regime, ISNs produce damped oscillations in the γ-range in response to inputs to the inhibitory population. In order to understand the proper-ties of interconnected ISNs, we investigated periodic forcing of ISNs. We show that ISNs can be excited over a range of frequencies and derive properties of the resonance peaks. In particular, we studied the phase-locked solutions, the torus solutions and the resonance peaks. More particular, periodically forced ISNs respond with (possibly multi-stable) phase-locked activity whereas networks with sustained intrinsic oscilla-tions respond more dynamically to periodic inputs with tori. Hence, the dynamics are surprisingly rich and phase effects alone do not adequately describe the network re-sponse. This strengthens the importance of phase-amplitude coupling as opposed to phase-phase coupling in providing multiple frequencies for multiplexing and routing information.

This work has been submitted to a Journal and is available as [38].

5.3. Mean field approaches

5.3.1. *A Large Deviation Principle and an Expression of the Rate Function for a Discrete Stationary Gaussian Process*

Participants: Olivier Faugeras, James Maclaurin.

We prove a Large Deviation Principle for a stationary Gaussian process over $\mathbb{R}^b$, indexed by $\mathbb{Z}^d$ (for some positive integers d and b), with positive definite spectral density and provide an expression of the corresponding rate function in terms of the mean of the process and its spectral density. This result is useful in applications where such an expression is needed.

This work has been accepted for publication in Entropy [20].

5.3.2. A representation of the relative entropy with respect to a diffusion process in terms of its infinitesimal-generator

Participants: Olivier Faugeras, James Maclaurin.

In this paper we derive an integral (with respect to time) representation of the relative entropy (or Kullback-Leibler Divergence) $R(\mu|P)$, where μ and P are measures on $C([0,T];\mathbb{R}^d)$. The underlying measure P is a weak solution to a Martingale Problem with continuous coefficients. Our representation is in the form of an integral with respect to its infinitesimal generator. This representation is of use in statistical inference (particularly involving medical imaging). Since $R(\mu||P)$ governs the exponential rate of convergence of the empirical measure (according to Sanov's Theorem), this representation is also of use in the numerical and analytical investigation of finite-size effects in systems of interacting diffusions.

This work has been accepted for publication in the Journal Entropy [21].

5.3.3. Asymptotic description of stochastic networks of rate neurons with correlated synaptic weights

Participants: Olivier Faugeras, James Maclaurin.

We study the asymptotic law of a network of interacting neurons when the number of neurons becomes infinite. Given a completely connected network of neurons in which the synaptic weights are Gaussian correlated random variables, we describe the asymptotic law of the network when the number of neurons goes to infinity. Unlike previous works which made the biologically unplausible assumption that the weights were i.i.d. random variables, we assume that they are correlated. We introduce the process-level empirical measure of the trajectories of the solutions to the equations of the finite network of neurons and the averaged law (with respect to the synaptic weights) of the trajectories of the solutions to the equations of the network of neurons. The result is that the image law through the empirical measure satisfies a large deviation principle with a good rate function. We provide an analytical expression of this rate function. This work has appeared in the Comptes Rendus de l'Academie des Sciences. Serie 1, Mathematique [22].

We have continued the development, started in [22], of the asymptotic description of certain stochastic neural networks. We use the Large Deviation Principle (LDP) and the good rate function H announced there to prove that H has a unique minimum, a stationary measure on the set of trajectories $\mathcal{T}^{\mathbb{Z}}$. We characterize this measure by its two marginals, at time 0, and from time 1 to T. The second marginal is a stationary Gaussian measure. With an eye on applications, we show that its mean and covariance operator can be inductively computed. Finally we use the LDP to establish various convergence results, averaged and quenched. This work has also appeared in the Comptes Rendus de l'Academie des Sciences. Serie 1, Mathematique [23].

5.3.4. Asymptotic description of stochastic networks of integrate-and-fire neurons

Participants: François Delarue [UNS, LJAD], James Inglis [EPIs TOSCA and NeuroMathComp], S Rubenthaler [UNS, LJAD], Etienne Tanré [EPI TOSCA].

J. Inglis, together with F. Delarue (Univ. Nice – Sophia Antipolis), E. Tanré (Inria TOSCA) and S. Rubenthaler (Univ. Nice – Sophia Antipolis) completed their study of the mean-field convergence of a highly discontinuous particle system modeling the behavior of a spiking network of neurons, based on the integrate-and-fire model. Due to the highly singular nature of the system, it was convenient to work with a relatively unknown Skorohod topology. The resulting article [46] has been accepted for publication in *Stochastic Processes and Related Fields*.

5.3.5. Asymptotic description of stochastic networks of spiking neurons with dendrites

Participants: James Inglis [EPIs TOSCA and NeuroMathComp], Denis Talay [EPI TOSCA].

J. Inglis and D. Talay introduced in [49] a new model for a network of spiking neurons that attempted to address several criticisms of previously considered models. In particular the new model takes into account the role of the dendrites, and moreover includes non-homogeneous synaptic weights to describe the fact that not all neurons have the same effect on the others in the network. They were able to obtain mean-field convergence results, using new probabilistic arguments.

5.3.6. *Asymptotic description of stochastic networks of realistic neurons and synapses*
Participants: Mireille Bossy [EPI TOSCA], Olivier Faugeras, Denis Talay [EPI TOSCA].

In this note, we clarify the well-posedness of the limit equations to the mean-field N -neuron models proposed in [1] and we prove the associated propagation of chaos property. We also complete the modeling issue in [1] by discussing the well-posedness of the stochastic differential equations which govern the behaviour of the ion channels and the amount of available neurotransmitters.

This work has been submitted for publication to a Journal and is available as [40].

5.3.7. *On the Hamiltonian structure of large deviations in stochastic hybrid systems*
Participants: Paul Bressloff [Prof. University of Utah, Inria International Chair], Olivier Faugeras.

We develop the connection between large deviation theory and more applied approaches to stochastic hybrid systems by highlighting a common underlying Hamiltonian structure. A stochastic hybrid system involves the coupling between a piecewise deterministic dynamical system in $\mathbb{R}^d$ and a time-homogeneous Markov chain on some discrete space Γ. We assume that the Markov chain on Γ is ergodic, and that the discrete dynamics is much faster than the piecewise deterministic dynamics (separation of time-scales). Using the Perron-Frobenius theorem and the calculus-of-variations, we evaluate the rate function of a large deviation principle in terms of a classical action, whose Hamiltonian is given by the Perron eigenvalue of a $|\Gamma|$-dimensional linear equation. The corresponding linear operator depends on the transition rates of the Markov chain and the nonlinear functions of the piecewise deterministic system. The resulting Hamiltonian is identical to one derived using path-integrals and WKB methods. We illustrate the theory by considering the example of stochastic ion channels. This work has been submitted for publication to a Journal and is available as [41].

5.4. Neural fields theory

5.4.1. *Neural fields with noise*
Participants: Olivier Faugeras, James Inglis.

We extend the theory of neural fields which has been developed in a deterministic framework by considering the influence spatio-temporal noise. The outstanding problem that we here address is the development of a theory that gives rigorous meaning to stochastic neural field equations, and conditions ensuring that they are well-posed. Previous investigations in the field of computational and mathematical neuroscience have been numerical for the most part. Such questions have been considered for a long time in the theory of stochastic partial differential equations, where at least two different approaches have been developed, each having its advantages and disadvantages. It turns out that both approaches have also been used in computational and mathematical neuroscience, but with much less emphasis on the underlying theory. We present a review of two existing theories and show how they can be used to put the theory of stochastic neural fields on a rigorous footing. We also provide general conditions on the parameters of the stochastic neural field equations under which we guarantee that these equations are well-posed. In so doing we relate each approach to previous work in computational and mathematical neuroscience. We hope this will provide a reference that will pave the way for future studies (both theoretical and applied) of these equations, where basic questions of existence and uniqueness will no longer be a cause for concern. This work has appeared in the Journal of Mathematical Biology [19].

5.4.2. *A center manifold result for delayed neural fields equations*
Participants: Romain Veltz, Olivier Faugeras.

Lemma C.1 in [83] is wrong. This lemma is used in the proof of the existence of a smooth center manifold, Theorem 4.4. An additional assumption is required to prove this existence. We spell out this assumption, correct the proofs and show that the assumption is satisfied for a large class of delay functions τ. We also weaken the general assumptions on τ.

This work has been accepted for publication in the SIAM Journal on Mathematical Analysis and is available as [37].

5.5. Spike trains statistics

5.5.1. *Exact computation of the Maximum Entropy Potential of spiking neural networks models*

Participants: Bruno Cessac [correspondent], Rodrigo Cofre.

Understanding how stimuli and synaptic connectivity in uence the statistics of spike patterns in neural networks is a central question in computational neuroscience. Maximum Entropy approach has been successfully used to characterize the statistical response of simultaneously recorded spiking neurons responding to stimuli. But, in spite of good performance in terms of prediction, the fitting parameters do not explain the underlying mechanistic causes of the observed correlations. On the other hand, mathematical models of spiking neurons (neuro-mimetic models) provide a probabilistic mapping between stimulus, network architecture and spike patterns in terms of conditional probabilities. In this paper we build an exact analytical mapping between neuro-mimetic and Maximum Entropy models.

This work has been published in [18] and presented in [44], [29], [30], [27].

5.5.2. *Parameter Estimation for Spatio-Temporal Maximum Entropy Distributions: Application to Neural Spike Trains*

Participants: Bruno Cessac [correspondent], Hassan Nasser.

We propose a numerical method to learn maximum entropy (MaxEnt) distributions with spatio-temporal constraints from experimental spike trains. This is an extension of two papers, [10] and [4], which proposed the estimation of parameters where only spatial constraints were taken into account. The extension we propose allows one to properly handle memory effects in spike statistics, for large-sized neural networks.

This work has been published in [25] and presented in [44], [29], [30], [27].

5.6. Synaptic plasticity

5.6.1. *Large Deviations of an Ergodic Synchoronous Neural Network with Learning*

Participants: Olivier Faugeras, James Maclaurin.

In this work we determine a Large Deviation Principle (LDP) for a model of neurons interacting on a lattice $\mathbb{Z}^d$. The neurons are subject to correlated external noise, which is modelled as an infinite-dimensional stochastic integral. The probability law governing the noise is strictly stationary, and we are therefore able to find a LDP for the probability laws Π^n governing the ergodic empirical measure μ^n generated by the neurons in a cube of length $(2n+1)$ as n asymptotes to infinity. We use this LDP to determine an LDP for the neural network model. The connection weights between the neurons evolve according to a learning rule / neuronal plasticity, and these results are adaptable to a large variety of specific types of neural network. This LDP is of great use in the mathematical modelling of neural networks, because it allows a quantification of the likelihood of the system deviating from its limit, and also a determination of which direction the system is likely to deviate. The work is also of interest because there are nontrivial correlations between the neurons even in the asymptotic limit, thereby presenting itself as a generalisation of traditional mean-field models.

This work is available [47] and is under review in a Journal.

5.7. Visual Neuroscience

5.7.1. *On the effects on cortical spontaneous activity of the symmetries of the network of pinwheels in visual area V1*

Participants: Romain Veltz, Pascal Chossat, Olivier Faugeras.

We consider the problem of describing mathematically the spontaneous activity of V1 by combining several important experimental observations including 1) the organization of the visual cortex into a spatially periodic network of hypercolumns structured around pinwheels, 2) the difference between short-range and long-range intra-cortical connections, the first ones being rather isotropic and producing naturally doubly-periodic patterns by Turing mechanisms, the second one being patchy and 3) the fact that the Turing patterns spontaneously produced by the short-range connections and the network of pinwheels have similar periods. By analyzing the Preferred Orientation (PO) map, we are able to classify all possible singular points of the PO maps (the pinwheels) as having symmetries described by a small subset of the wallpaper groups. We then propose a description of the spontaneous activity of V1 using a classical voltage-based neural field model that features isotropic short-range connectivities modulated by non-isotropic long-range connectivities. A key observation is that, with only short-range connections and because the problem has full translational invariance in this case, a spontaneous doubly-periodic pattern generates a 2-torus in a suitable functional space which persists as a flow-invariant manifold under small perturbations, hence when turning on the long-range connections. Through a complete analysis of the symmetries of the resulting neural field equation and motivated by a numerical investigation of the bifurcations of their solutions, we conclude that the branches of solutions which are stable over an extended set of parameters are those corresponding to patterns with an hexagonal (or nearly hexagonal) symmetry. The question of which patterns persist when turning on the long-range connections is answered by 1) analyzing the remaining symmetries on the perturbed torus and 2) combining this information with the Poincaré-Hopf theorem. We have developed a numerical implementation of the theory that has allowed us to produce the patterns of activities predicted by the theory, the planforms. In particular we generalize the contoured and non-contoured planforms predicted by previous authors and predict the existence of mixed contoured/non-contoured planforms. We also found that these planforms are most likely to be time evolving. This work is available as a preprint [36] and has been submitted to a Journal.

5.7.2. *Decoding the retina with the first wave of spikes*

Participants: Geoffrey Portelli, John Barrett [Institute of Neuroscience, Medical School, Newcastle University, Newcastle UK], Evelyne Sernagor [Institute of Neuroscience, Medical School, Newcastle University, Newcastle UK], Timothée Masquelier [Institut de la Vision, UPMC Université Paris 06, Paris, France], Pierre Kornprobst [correspondent].

How a population of retinal ganglion cells (RGCs) encode the visual scene remains an open question. Several coding strategies have been investigated out of which two main views have emerged: considering RGCs as independent encoders or as synergistic encoders, i.e., when the concerted spiking in a RGC population carries more information than the sum of the information contained in the spiking of individual RGCs. Although the RGCs assumed as independent encode the main information, there is currently a growing body of evidence that considering RGCs as synergistic encoders provides complementary and more precise information. Based on salamander retina recordings, it has been suggested [66] that a code based on differential spike latencies between RGC pairs could be a powerful mechanism. Here, we have tested this hypothesis in the mammalian retina. We recorded responses to stationary gratings from 469 RGCs in 5 mouse retinas. Interestingly, we did not find any RGC pairs exhibiting clear latency correlations (presumably due to the presence of spontaneous activity), showing that individual RGC pairs do not provide sufficient information in our conditions. However considering the whole RGC population, we show that the shape of the wave of first spikes (WFS) successfully encodes for spatial cues. To quantify its coding capabilities, we performed a discrimination task and we showed that the WFS was more robust to the spontaneous firing than the absolute latencies are. We also investigated the impact of a post-processing neural layer. The recorded spikes were fed into an artificial lateral geniculate nucleus (LGN) layer. We found that the WFS is not only preserved but even refined through the LGN-like

layer, while classical independent coding strategies become impaired. These findings suggest that even at the level of the retina, the WFS provides a reliable strategy to encode spatial cues.

This work is ongoing and was presented as a poster at CNS 2014. See [26] for more details.

5.7.3. Microsaccades enable efficient synchrony-based visual feature detection by retinal ganglion cells

Participants: Timothée Masquelier [Institut de la Vision, UPMC Université Paris 06, Paris, France], Geoffrey Portelli, Pierre Kornprobst [correspondent].

Fixational eye movements are common across vertebrates, yet their functional roles, if any, are debated [70]. To investigate this issue, we exposed the Virtual Retina simulator [84] to natural images, generated realistic drifts and microsaccades [59], and analyzed the output spike trains of the parvocellular retinal ganglion cells (RGC). We first computed cross-correlograms between pairs of RGC that are strongly excited by the image corresponding to the mean eye position. Not surprisingly, in the absence of eye movements, that is when analyzing the tonic (sustained) response to a static image, these cross-correlograms are flat. Adding some slow drift ($\approx$20 min/s, self- avoiding random walk) creates long timescale (>1s) correlations because both cells tend to have high firing rates for central positions. Adding microsaccades ($\approx$0.5° in 25ms, that is $\approx$20°/s) creates short timescale (tens of ms) correlations: cells that are strongly excited at a particular landing location tend to spike synchronously shortly after the landing. What do the patterns of synchronous spikes represent? To investigate this issue, we fed the RGC spike trains to neurons equipped with spike timing-dependent plasticity (STDP) and lateral inhibitory connections [73]. Neurons self-organized, and each one selected a set of afferents that consistently fired synchronously. We then reconstructed the corresponding visual stimuli by convolving the synaptic weight matrices with the RGC receptive fields. In most cases, we could easily recognize what was learned (e.g. a face), and the neuron was selective (e.g. only responded for microsaccades that landed on a face). Without eye movements, or with only the drift, the STDP- based learning failed, because it needs correlations at a timescale roughly matching the STDP time constants [65]. Microsaccades are thus necessary to generate a synchrony-based coding scheme. More specifically, after each microsaccade landing, cells that are strongly excited by the image corresponding to the landing location tend to fire their first spikes synchronously. Patterns of synchronous spikes can be decoded rapidly – as soon as the first spikes are received – by downstream "coincidence detector" neurons, which do not need to know the landing times. Finally, the required connectivity to do so can spontaneously emerge with STDP. As a whole, these results suggest a new role for microsaccades – to enable efficient visual feature learning and detection thanks to synchronization – that differs from other proposals such as time-to-first spike coding with respect to microsaccade landing times.

This work is ongoing and was presented as a poster at CNS 2014.

5.7.4. A new retina-inspired descriptor for image classification

Participants: Cristina Hilario [Pattern Analysis and Computer Vision, PAVIS, Istituto Italiano di Tecnologia, Genova, Italy], Diego Sona [Pattern Analysis and Computer Vision, PAVIS, Istituto Italiano di Tecnologia, Genova, Italy], Kartheek Medathati, Pierre Kornprobst [correspondent].

Recent studies on the visual system reveal that retina is smarter than scientists believed. One low level processing occurring at the retina is feature extraction, becoming an inspiration to build novel image descriptors for image categorization. However only few methods have taken advantage of this idea, such as FREAK descriptor [51], which consists of a circular grid of a concentric distribution of overlapping receptive fields (RFs) in which average image intensities are compared pairwise. In this work we extended such a descriptor but sticking much more to biological data and models of retina. Each RF in our model is described with a linear-nonlinear model (LN) taking into account inhibitory surrounds with parameters based on biological findings. Based on the activity of retinal ganglion cells, we investigated several methods to define a set of descriptors. The performance of each descriptor was tested on computer vision datasets for texture and scene categorization.

This work is ongoing and was presented as a poster at the 1st Workshop of Visual Image Interpretation in Humans and Machine (VIIHM, EPSRC Network for Biological and Computer Vision in the UK).

5.7.5. *Shifting stimulus for faster receptive estimation of ensembles of neurons*

Participants: Daniela Pamplona, Bruno Cessac, Pierre Kornprobst [correspondent].

The spike triggered averaged (STA) technique has been widely used to estimate the receptive fields (RF) of sensory neurons [58]. Theoretically, it has been shown that when the neurons are stimulated with a white noise stimulus the STA is an unbiased estimator of the neuron RF (up to a multiplicative constant). The error decreases with the number of spikes at a rate proportional to the stimulus variance [75]. Experimentally, for visual neurons, the standard stimuli are checkerboards where block size is heuristically tuned. This raises difficulties when dealing with large neurons assemblies: When the block size is too small, neuron's response might be too weak, and when it is too large, one may miss RFs. Previously online updating the stimulus in the direction of larger stimulus-neural response correlation [62] or mutual information [69], [68] has been proposed. However, these approaches can not be applied for an ensemble of cells recorded simultaneously since each neuron would update the stimulus in a different direction. We propose an improved checkerboard stimulus where blocks are shifted randomly in space at fixed time steps. Theoretically, we show that the STA remains an unbiased estimator of the RF. Additionally, we show two major properties of this new stimulus: (i) For a fixed block sized, RF spatial resolution is improved as a function of the number of possible shifts; (ii) Targeting a given RF spatial resolution, our method converges faster than the standard one. Numerically, we perform an exhaustive analysis of the performance of the approach based on simulated spiked trains from LNP cascades neurons varying RF sizes and positions. Results show global improvements in the RF representation even after short stimulation times. This makes this approach a promising solution to improve RF estimation of large ensemble of neurons.

This work is ongoing and was submitted to COSYNE 2015.

5.7.6. *Shifting stimulus for faster receptive estimation of ensembles of neurons*

Participants: Kartheek Medathati, Fabio Solari [University of Genoa - DIBRIS, Italy], Manuela Chessa [University of Genoa - DIBRIS, Italy], Guillaume S. Masson [Institut des Neurosciences de la Timone, Team InVibe], Pierre Kornprobst [correspondent].

Motion estimation has been studied extensively in neurosciences in the last two decades. The general consensus that has evolved from the studies in the primate vision is that it is done in a two stage process involving cortical areas V1 and MT in the brain. Spatio temporal filters are leading contenders in terms of models that capture the characteristics exhibited in these areas. Even though there are many models in the biological vision literature covering the optical flow estimation problem based on the spatio-temporal filters little is known in terms of their performance on the modern day computer vision datasets such as Middlebury. In this paper, we start from a mostly classical feedforward V1-MT model introducing a additional decoding step to obtain an optical flow estimation. Two extensions are also discussed using nonlinear filtering of the MT response for a better handling of discontinuities. One essential contribution of this paper is to show how a neural model can be adapted to deal with real sequences and it is here for the first time that such a neural model is benchmarked on the modern computer vision dataset Middlebury. Results are promising and suggest several possible improvements.

This work is ongoing and was presented as a poster at the 1st Workshop of Visual Image Interpretation in Humans and Machine (VIIHM, EPSRC Network for Biological and Computer Vision in the UK). See [35] for more details.

5.7.7. *Exploring the richness of center-surround dynamics: A bifurcation study*

Participants: Kartheek Medathati, James Rankin [Center for Neural Sciences, NYU, USA], Guillaume S. Masson [Institut des Neurosciences de la Timone, Team InVibe], Pierre Kornprobst [correspondent].

The balance of excitatory and inhibitory interactions between neurons is one of the characteristic aspects of neural computation. In both neural network and neural field models these interactions have been modeled using center-surround connectivity kernels. Depending on the relative strength of excitation and inhibition these networks have been found to exhibit rich and interesting dynamical behavior. Although many models have been reported in the literature using center-surround connectivity kernels and many experimental studies have

shown evidence for changes in observed behavior from winner-take-all to gain control, a thorough bifurcation analysis of these networks in terms of sensitivity of the network to peak strength, discriminability of the peaks and speed of convergence has not been done. In our present work we visit this question in order to identify the parameter regimes where this important switch in the behavior of the network occurs and also establish the trade offs that arise with the choice of a particular connectivity kernel.

This work is ongoing and was presented as a poster at the conference "Nonlinear dynamics and stochastic methods: from neuroscience to other biological applications"

5.7.8. *From Habitat to Retina:Neural Population Coding using Natural Movies*

Participants: Bruno Cessac [correspondent], Ruben Herzog [Centro Interdisciplinario de Neurociencia de Valparaíso, Univ de Valparaíso, 2360102 Valparaíso, Chile], Joaquin Araya [Centro Interdisciplinario de Neurociencia de Valparaíso, Univ de Valparaíso, 2360102 Valparaíso, Chile], Michael Pizarro [Centro Interdisciplinario de Neurociencia de Valparaíso, Univ de Valparaíso, 2360102 Valparaíso, Chile], Cesar Ravello [Centro Interdisciplinario de Neurociencia de Valparaíso, Univ de Valparaíso, 2360102 Valparaíso, Chile], Maria Jose Escobar [Universidad Técnica Federico Santa María, Valparaiso, Chile], Adrian Palacios [Centro Interdisciplinario de Neurociencia de Valparaíso, Univ de Valparaíso, 2360102 Valparaíso, Chile].

We use a diurnal rodent retina (O. Degus), which has the advantage of present a 3:1 proportion of rods and cones, respectively, to study the RGC population responses to habitat-based natural stimuli. In order to do this, we have developed a mobile robot that is capable to record movies in the natural habitat of this rodent, simulating both his movements and the eye-ground distance, which allows us to stimulate and record an in vitro retina patch using MEA (multi electrode array) with a sequence of images taken from the animal natural habitat. The analysis of spike statistics has been done using the Enas software to characterize spatio-temporal pairwise correlation with Gibbs distributions. potential constitutes a useful tool for comparing pairwise spatio-temporal correlations between different conditions for the same RGC population. We show that correlated spiking patterns represents a major deviation between White Noise and Natural Movies conditions. We also conclude that population coding for this monophasic OFF RGC population is mostly based on spatial correlation when stimulated with Natural Movies.

This ongoing work has been presented in [48].

...

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

5. New Results

5.1. Highlights of the Year

- Publication of a joint work with RDP at ENS-Lyon in the journal 'Nature'. In December 2013, a joint work on phyllotaxy with the RDP lab from ENS-Lyon was published online in the journal Nature [2]. This paper obtained the 2014 prize "la Recherche" in the biology category http://www.leprixlarecherche.com. Based on the analysis of phyllotaxis perturbations in mutants, this study sheds a new light on our interpretation of phyllotaxis, revisiting the standard model and suggesting that several fields based on auxin and cytokinin with different properties are required to provide robustness to phyllotaxis.

- To study and model morphogenesis, the team has been working in the last 8 years on modeling mechanical forces and deformations in tissues in collaboration with the UMR RDP at ENS-Lyon. This work has given rise to the development of a 3D computational framework to model the mechanics of 3D plant tissues during growth at cellular resolution and has been finalized this year with a publication in PLoS Comp. Biology (to appear in 2015). This framework makes it possible to construct models of meristem development, showing how the regulation of regional identities can lead to realistic shape development by dynamically modulating the mechanical properties of cells. It has been used also to study the influence of a specific signalling cascade (the ABP1-Kat1 signalling pathway) and its putative mechanical consequences on primordium initiation [25]. The expertize gained by our groups on physical models of plant tissue development has been wrapped up in a review paper [12].

5.2. Analysis of structures resulting from meristem activity

5.2.1. *Acquisition and design of plant geometry*

Participants: Frédéric Boudon, Christophe Pradal, Christophe Godin, Christian Fournier, Ibrahim Chedaddi, Mathilde Balduzzi, Julien Diener.

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 (ligth, 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 [61], [52]. 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 and video.

- *Reconstruction of plant architecture from 3D laser scanner data.* (Chakkrit Preuksakarn, Mathilde Balduzzi, Frédéric Boudon, Christophe Godin, Pascal Ferraro [Labri, Bordeaux], Yassin Refahi)

  We investigate the possibility to use 3D laser scanners to automate plant digitizing. We are developing algorithms to reconstruct branching systems without leaves or foliage from scanner data or from scan simulated on plant mock-up obtained using different digitizing method.

For branching systems, a number of automatic methods were proposed in the literature to reconstruct plausible branching structures from laser scanner data. The question of their assessment and accuracy is however critical for further exploitation in biological applications. For this, we developed an evaluation pipeline to assess reconstruction accuracy. A laser scan database on which experts built reference reconstructions is used as a basis of the evaluation. A graphical editor has been developed to help experts to reconstruct semi automatically reference structures. The evaluation pipeline is given two plant structures and compares their organization. Similar elements are identified based on geometric criteria using an optimization algorithm. The organization of these elements is then compared and their similarity is quantified. Two indices of geometrical and structural similarities are defined, and automatic reconstructions can thus be compared to reference structures to assess their accuracy. The method is successful at capturing the variation of similarities between two structures as different levels of noise are introduced. A first comparative evaluation of the different methods of the literature has been designed and conducted. This work has been published in the special issue of Annals of Botany on FSPMs [16]. A procedure to automatically determine phyllotactic angles from scans of small plants has been added recently to the reconstruction pipeline.

Xu *et al.*, 07   Livny *et al.*, 10   Preuksakarn *et al.*, 10
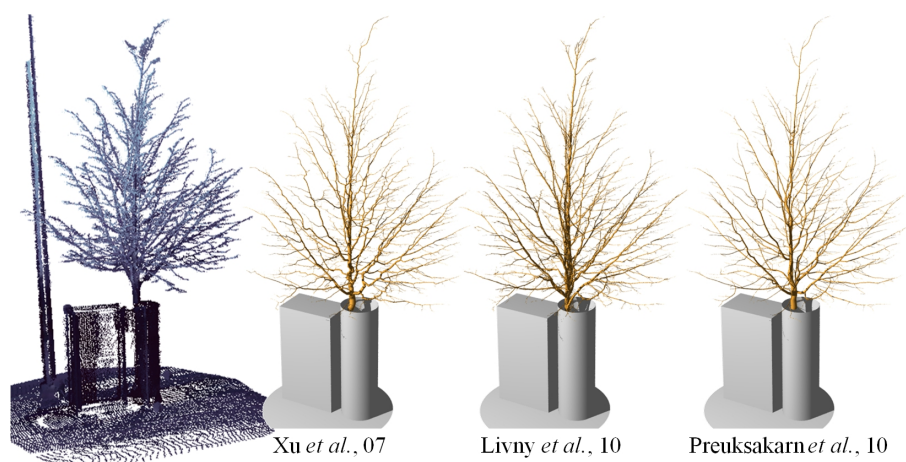
Figure 2. Comparison of different methods of the litterature to reconstruct plant architecture from laser scanner data. On the left, the original scan and on the rigth, the resulting reconstructions using three different methods. These reconstructions are quantitativelly assessed using our evaluation pipeline presented in [16]

In the context of the PhD of M. Balduzzi, we also investigated the reconstruction of tree foliage from 3D scans. Such elements are crucial to study the interaction of the plant with its environment. However, laser scans contain outliers in the silhouette of the scans that make the meshing of the pointset extremely difficult. New generation of laser scanners provide intensity of the laser reflected on the surface of scanned objects. This intensity depends on the distance to the object, its optical property and the incidence angle. A first work on this topic showed that after correcting the distance effect, the incidence angle can be deduced from the intensity. From this result, we developed a reconstruction pipeline using the scan intensities and based on Shape-From-Shading. Outliers being along the edge of the surface point cloud, we chose to develop a propagation SFS method initialized with points of the scans with high quality. We proved that surface with constant intensity are necessarily surfaces of constant slope or sand-pile surfaces. Using this result, a propagation method along iso-intensity regions was developed. These surfaces can then be sampled to provide a smooth point set without outliers. This work has been presented by M. Balduzzi for her thesis defense in november.

- *Reconstruction from video.* (Frédéric Boudon, Jerome Guenard [IRIT, Toulouse], Geraldine 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, dedicated methods for generating 3D plant models must be devised due to the high complexity of plant topology. In collaboration with our colleagues from IRIT, Toulouse, we developed an analysis-by-synthesis method which generates 3D models of plants from both images and a priori knowledge of the plant species.

  Our method is based on a skeletonisation algorithm which 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 organization of natural branching structure. 3D models are then generated. 2D projections 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 2D projection criterion. Realistic results are obtained on different species of plants, in particular vineyards. This work has been presented at the Mathematical Methods for Curves and Surfaces conference and published in LNCS [42].

- *Reconstruction of virtual fruits from pictures.* (Ibrahim Chedaddi, 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 MecaFruit3D.

  The aim of this work is to provide methods for generating fruit structure that can be integrated with models of fruit function. To this end, we used a modeling pipeline previously developped by Mik Cieslak in the OpenAlea platform. This methods 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. Previous studies demonstrated the possibility to create species-specific models of fruit structure with relatively low effort [49]. We focus now on validating the vasculature networks by quantitatively comparing them to experimental data from the litterature.

  These physiological data will be combined with a mechanical model of fruit growth, to investigate the effects of fruit structure on quality (see section 5.4.2).

- *Reconstruction of root structures.* (Julien Diener, Frédéric Boudon, Christophe Pradal, Christophe Godin, Philippe Nacry [BPMP, INRA], Christophe Périn [AGAP, CIRAD], Anne Dievart [AGAP, CIRAD], Xavier Draye [UCL, Belgium])

  This research theme is supported by the Agropolis through the Rhizopolis project and by NUMEV.

  Similarly to aerial part of plants, new needs for automatic digitizing of root systems emerge. Most existing methods focus only on semi-automatic approaches. This does not support the high-throughput capabilities of acquisition systems. In the context of the RhizoScan project, we previously designed a prototype of an automatic image analysis pipeline to extract root system architecture of branching systems grown in Petri boxes. This pipeline provides i) a set of model based image segmentation method, ii) the extraction of a graph representation of the root system, and iii) a method to identify the root axes organization. This year, we improved and extended the pipeline in the following way:

  1. We integrated a validation step in the workflow based on the comparison method presented in [16].

  2. We participated in an international collaboration with the Université Catholique de Louvain (Belgium), the CPIB of the University of Notthingham (UK), the University of Vienna (Austria), the Jülich research center (Germany) and INRA, to develop a standard file format for root architecture. The resulting format (RSML) is described in a publication to appear in *Plant Physiology* in 2015 (details can be found at rootsystemml.github.io).

3. We have initiated a collaboration to integrate visual data mining methods developed by the Zenith team in order to improve the automation of the image analysis pipeline.

4. In general, the robustness of the pipeline has been improved. In particular, an optimization method has been designed to select the root axes hierarchy that respect specific botanical constraints.

5.2.2. *Modeling the plant ontogenic programme*

Participants: Christophe Godin, Yann Guédon, Jean-Baptiste Durand, Pierre Fernique, Christophe Pradal, Jean Peyhardi.

This research theme is supported by two PhD programmes.

The remarkable organization of plants at macroscopic scales may be used to infer particular aspects of meristem functioning. The fact that plants are made up of the repetition of many similar components at different scales, and the presence of morphological gradients, e.g. [46], [54], [55], [51], provides macroscopic evidence for the existence of regularities and identities in processes that drive meristem activity at microscopic scales. Different concepts have been proposed to explain these specific organizations such as "morphogenetic programme" [59], "age state" [50] or "physiological age" [48]. 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 [50], [48]. Here we develop computational methods to decipher these rules.

- *Relating branching structure to the shoot properties* (Jean Peyhardi, Yann Guédon, Evelyne Coste [AGAP, AFEF team], Catherine Trottier [I3M], 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, we chose to develop a new family of multinomial GLMs called partitioned conditional GLMs [38] that enable to tackle hierarchically-structured 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 partitioned conditional GLMs and have been applied to apple and pear 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)

  A first study was published 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, which corresponds to 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 to quantify alternation from the yearly numbers of inflorescences at tree scale.

- *Identifying and characterizing patterns in tree-structured data* (Pierre Fernique, Jean-Baptiste Durand, Yann Guédon).

  In the context of Pierre Fernique's PhD (Montpellier 2 University and CIRAD), two complementary approaches were developed for analyzing patterns in tree-structured data:
    - multitype branching processes relying on local dependency properties for analyzing motifs.
    - multiple change-point models relying on long-term dependencies for segmenting trees in homogeneous zones.

  In multitype branching processes, the plant development is viewed as a demographic process, a parent entity of a given type generating child entities of different types (e.g. vegetative and flowering entities). Formally, the botanical entity properties are summarized as a categorical state variable. The number of child entities in each state is modeled through discrete multivariate distributions. Model selection procedures are necessary to specify parsimonious generation 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 [33]. To relax the strong constraints regarding dependencies induced by parametric distributions, mixture of graphical models were also considered [41]. Multitype branching processes were applied to the analysis of the patchiness pattern (consisting of canopies made of clumps of either vegetative or flowering growth units) in mango trees.

- *Simulating fruit tree phenology* (A.S. Briand, Frédéric Boudon, Frédéric Normand [CIRAD, Hort-Sys, Réunion Island], Anaëlle Dambreville, Jean-Baptiste Durand, Pierre Fernique, Yann Guédon, Christophe Pradal, Pierre-Eric Lauri [AFEF team, AGAP])

  Mango is a tropical tree characterized by strong asynchronisms within and between trees. To study more precisely the interplay between the plant structural components, we built an integrative model to simulate the plant development based on the L-system formalism and GLM to model the dependencies between events. This model has been started during the internship of A. Jestin last year and has been continued during the visit of F. Boudon at the Hortsys lab in the Réunion island. The model has been extended this year to model growth and phenology of shoots and inflorescences (internship of A.S. Briand). For this, the sizes of the different organs is modelled by statistical laws estimated from measurements that depends on their positions in the architecture. The growth speed of organs is modulated by the temperature. This model will serve as a basis for further ecophysiological study in silico.

- *Integrative developmental growth stages of shoots* (Anaëlle Dambreville, Yann Guédon, Pierre-Eric Lauri [AFEF team, AGAP], Frédéric Normand [CIRAD, HortSys, Réunion Island])

  Plant growth, i.e. the increase of organ dimensions over time, and development, i.e. the change in plant structure, are often studied as two separate processes. However, there is structural and functional evidence that these two processes are strongly related. Our aim was to investigate the co-ordination between growth and development using mango trees, which have well-defined developmental stages. Developmental stages, determined in an expert way, and organ sizes, determined from objective measurements, were collected during the vegetative growth and flowering phases of two cultivars of mango. For a given cultivar and growth unit type (either vegetative or flowering), a multi-stage model based on absolute growth rate sequences deduced from the measurements was first built, and then growth stages deduced from the model were compared with hand-annotated developmental stages. Strong matches were obtained between both stages, leading to a consistent definition of integrative developmental growth stages. The growth stages highlighted growth asynchronisms between

two topologically connected organs, namely the vegetative axis and its leaves. Integrative developmental growth stages emphasize that developmental stages are closely related to organ growth rates and can be interpreted in terms of the possible physiological processes (hydraulics, biomechanics and carbohydrate partitioning) underlying these stages.

- *Characterizing the successive flowering phases of strawberry in relation to genetic determinants* (Yann Guédon, Béatrice Denoyes [INRA, UMR BFP, Villenave d'Ornon], Justine Perrotte)

  Our aim was to characterize the successive flowering phases of perpetual flowering strawberry genotypes, which is of particular importance for better predicting fruit production. We applied multiple change-point models for the synchronous segmentation of the individuals of a given genotype in successive flowering phases. We identified two groups of genotypes that differ by the intensity of the flowering at the end of the flowering period. Using a genetic approach, we identified a locus controlling the flowering intensity at the end of the flowering period that likely explain these two groups of genotypes.

- *Self-nested structure of plants.* (Christophe Godin, Romain Azais, Farah Ben Naoum, Jean-Baptiste Durand, Alain Jean-Marie)

  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 (Directed Acyclic Graph, 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.

  The method proposed in [6] thus compresses a tree in width, but not in height. In a 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 has been submitted recently for publication.

  The degree of self-nestedness is defined in [6] as the edit-distance between the considered tree structure and its nearest embedded self-nested version. Indeed, finding the nearest self-nested tree of a structure without more assumptions is conjectured to be an NP-complete or NP-hard problem. We thus designed a heuristic method based on interacting simulated annealing algorithms to tackle this difficult question. This procedure is also a keystone in a new topological clustering algorithm for trees that we propose. In addition, we obtain new theoretical results on the combinatorics of self-nested structures. The redaction of an article is currently in progress.

5.2.3. *Analyzing the influence of the environment on the plant ontogenic programme*

Participants: Jean-Baptiste Durand, Christian Fournier, Christophe Godin, Yann Guédon, Christophe Pradal, Jean Peyhardi, Pierre Fernique, Guillaume Garin.

This research theme is supported by three PhD programs.

The ontogenetic programme of a plant is actually sensitive to environmental changes. If, in particular cases, we can make the assumption that the environment is a fixed control variable (see section 5.2.2), in general the structure produced by meristem results from a tight interaction between the plant and its environment, throughout its lifetime. Based on observations, we thus aim to trace back to the different components of the growth (ontogenetic development and its modulation by the environment). This is made using two types of approaches. On the one hand, we develop a statistical approach in which stochastic models are augmented with additional time-varying explanatory variables that represent the environment variations. The design of estimation procedures for these models make it possible to separate the plant ontogenetic programme from its modulation by the environment. On the other hand, we build reactive models that make it possible to simulate in a mechanistic way the interaction between the plant development and its environment.

- *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 [24] 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])

  We identified robust indicators that summarize the respective importance of ontogeny and environmental constraints (mainly related to light environment) in forest tree development. In this context, tree growth data correspond to the retrospective measurement of annual shoot characteristics (e.g. length, number of branches) along the main stem. We applied segmentation models to identify tree growth phases. These segmentation models, which are hidden semi-Markov chains, were compared with simple hidden Markov chains that correspond to the environment-driven development assumption. This statistical modelling approach was applied to both evergreen (Corsican pine and silver fir) and deciduous (sessile oak and Persian walnut) tree species growing in contrasted conditions ranging from managed forest stands to unmanaged understoreys. Growth phase duration distributions estimated within these segmentation models characterize the respective importance of ontogeny and environmental constraints in tree development at the population scale and have very contrasted characteristics in terms of shape and relative dispersion between ontogeny-driven and environment-driven tree development. These characteristics may change over tree life, reflecting changes in tree competition. Growth phase duration distributions summarize the joint trajectory of tree ontogeny and environment without requiring tree growth follow-up data for their estimation.

- *Analyzing fruit tree phenology in various climatic conditions from the shoot to the orchard scale* Yann Guédon, Juliano Dutra Schmitz [Universidade Federal de Pelotas, Brazil], Pierre-Eric Lauri [AFEF team, AGAP], Jean-Michel Legave [AFEF team, AGAP], Gustavo Malagui [Universidade Tecnológica Federal do Paraná]

  A first study was conducted at the shoot scale on four apple cultivars characterized by various chilling requirements and grown in two contrasting winter temperature conditions [20]. Our hypothesis was that shoot architecture is strongly affected by winter temperatures determining both the position and budburst of vegetative laterals with a lower effect on their outgrowth. A two-step approach was designed to quantify at the shoot scale i) the branching pattern and ii) two phenological stages of vegetative laterals, budburst and outgrowth. The branching pattern analysis combined various methods: branching sequence alignment and clustering, hidden semi-Markov chain and multiple change point model for segmenting branching sequences into homogeneous zones. A categorical variable, the branching zone, was built to summarize the lateral position along the shoot. It was integrated into the phenological analysis, based on a zero-inflated Poisson model, as a factor together with the cultivar and the winter temperature. We showed in this way that temperature had a main effect on the distribution of vegetative laterals along the shoot. It also strongly affected budburst, which was also affected by the cultivar and the branching zone. The outgrowth of the lateral was not significantly affected by temperature but was significantly affected by the cultivar and the branching zone. Furthermore, the delayed senescence and subsequent leaf persistence during winter, characterizing the apple tree in the mild winter temperature condition, had only a weak effect on the distribution of vegetative laterals and on budburst and lateral outgrowth. The actual shoot architecture and budburst result thus from an ordered sequence of events with a pivotal role of winter temperatures on the dormancy completion of individual lateral buds.

  A second study was conducted at the orchard scale. The time-course variation of dates of flowering stages was established for seventeen chronological sequences corresponding to various apple tree

cropping areas in Europe (Belgium, France, Germany, Italy, Switzerland), north Africa (Morocco) and southern Brazil. Our aim was to characterize the relationship between flowering advances in fruit trees and global warming and to compare the northern and the southern hemisphere situations. We applied piecewise constant and linear homoscedastic models to these phenological series. The sudden advance of flowering dates detected at the end of the 1980s in the European locations can be explained by changes in rates for completion of heat requirements, essential to the development of floral primordia within buds. No effect of the global warming could be detected in the Brazilian flowering series and we only found a direct effect of the chilling temperature on the flowering date the same year (the colder the Austral winter, the earlier the flowering date).

- *Investigating how architectural development interfer with epidemics and epidemic control* (Christian Fournier, Corinne Robert [EGC], Guillaume Garin [ITK], Bruno Andrieu [EGC], Christophe Pradal)

  Sustainable agriculture requires the identification of new, environmentally responsible strategies of crop protection. Modelling of pathosystems can allow a better understanding of the major interactions inside these dynamic systems and may lead to innovative protection strategies. In particular, functional–structural plant models (FSPMs) have been identified as a means to optimize the use of architecture-related traits. A current limitation lies in the inherent complexity of this type of modelling, and thus the purpose of this paper is to provide a framework to both extend and simplify the modelling of pathosystems using FSPMs.

  Different entities and interactions occurring in pathosystems were formalized in a conceptual model [21]. A framework based on these concepts was then implemented within the open-source OpenAlea modelling platform, using the platform's general strategy of modelling plant–environment interactions and extending it to handle plant interactions with pathogens. New developments include a generic data structure for representing lesions and dispersal units, and a series of generic protocols to communicate with objects representing the canopy and its microenvironment in the OpenAlea platform. Another development is the addition of a library of elementary models involved in pathosystem modelling. Several plant and physical models are already available in OpenAlea and can be combined in models of pathosystems using this framework approach.

  Two contrasting pathosystems are implemented using the framework and illustrate its generic utility. Simulations demonstrate the framework's ability to simulate multiscaled interactions within pathosystems, and also show that models are modular components within the framework and can be extended. This is illustrated by testing the impact of canopy architectural traits on fungal dispersal. This study provides a framework for modelling a large number of pathosystems using FSPMs. This structure can accommodate both previously developed models for individual aspects of pathosystems and new ones. Complex models are dissassembled into separate *knowledge sources* originating from different specialist areas of expertise and these can be shared and reassembled into multidisciplinary models. The framework thus provides a beneficial tool for a potential diverse and dynamic research community.

5.3. 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.

5.3.1. Data acquisition and design of meristem models

- *Improvement of the MARS-ALT pipeline robustness* Meristem, laser microscopy, image reconstruction, cell segmentation, automatic lineaging
  Participants: Léo Guignard, Christophe Godin, Christophe Pradal, Grégoire Malandain [Morpheme, Inria], Gaël Michelin [Morpheme, IPL Morphogenetics, Inria], Guillaume Baty, Sophie Ribes [IBC, UM2], Jan Traas [RDP, ENS], Patrick Lemaire [CRBM, CNRS], Yassin Refahi [RDP, ENS].
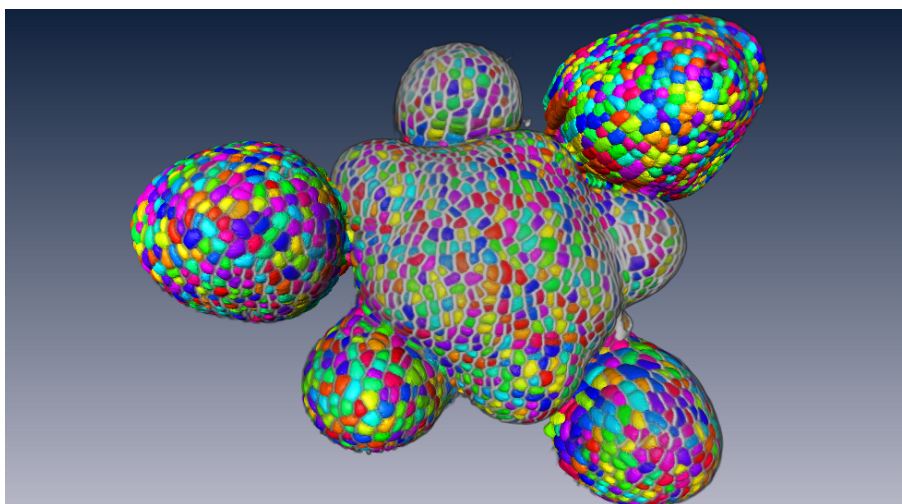
Participants: Guillaume Cerutti, Sophie Ribes, Christophe Godin, Géraldine Brunoud [RDP, ENS], Carlos Galvan-Ampudia [RDP, ENS], Teva Vernoux [RDP, ENS], Yassin Refahi [RDP, ENS, Sainsbury Lab].

This research theme is supported the HFSP project Biosensors.

To produce a more efficient data structure accounting for the geometry of cellular tissues, we studied the problem of reconstructing a mesh representation of cells in a complex, multi-layered tissue structure, based either on membrane/wall images segmented using MARS or on nuclei images of shoot apical meristems. The construction of such mesh structures for plant tissues is currently a missing step in the existing image analysis pipelines. We developed a set of tools to build a triangular mesh surface representing the tissue in 3D, to evaluate the quality of the tissue reconstruction over objective aspects, to optimize a low-quality mesh simultaneously along several criteria, and to go towards a higher-scale representation pulling away from the cell resolution [31]. These methods are used in particular on nuclei images of shoot apical meristems of *Arabidopsis thaliana* to project hormonal information at cell-level on a continuous 3D tissue geometry. This work is carried out in the context of the post-doc of Guillaume Cerutti within the HFSP project BioSensors (Collaboration RDP Lyon).

These tools can produce light discrete representations of the cell interfaces that enables fast visualization, information projection, and quantitative analysis of the tissue, and give way to *in silico* physical and mechanical simulations on real-world data.
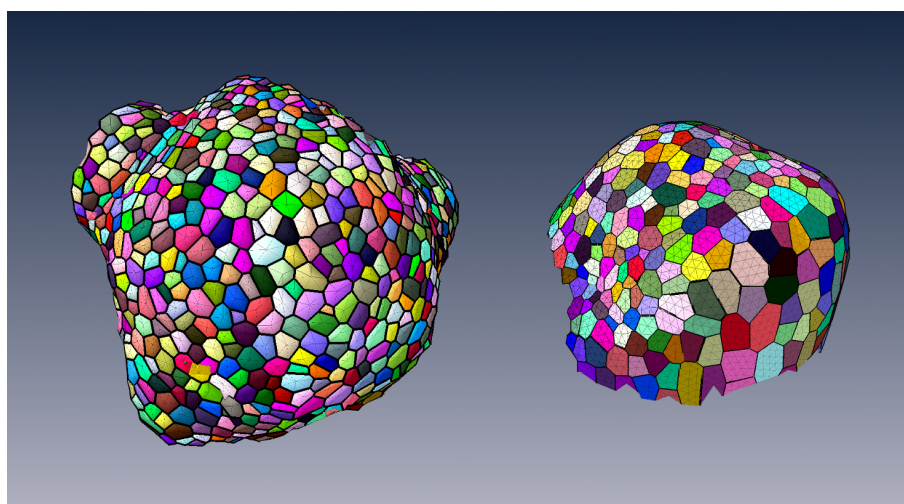
Figure 4. Triangular mesh representations of shoot apical meristem and flower meristem tissues obtained from MARS segmentations

- *Design of 3D digital atlases of tissue development*
  Participants: Sophie Ribes, Yassin Refahi [RDP, ENS, Sainsbury Lab], Guillaume Cerutti, Christophe Godin, Christophe Pradal, Christophe Pradal, Frédéric Boudon, Gregoire Malandain [RDP, ENS], Gaël Michelin [RDP, ENS], Guillaume Baty, Jan Traas [RDP, ENS], Teva Vernoux [RDP, ENS], Patrick Lemaire [CRBM, CNRS], Françoise Monéger [RDP, ENS].

  This research theme is supported the Inria Project Lab Morphogenetics, the ADT Mars-Alt and the HFSP project Biosensors.

  To organize the various genetic, physiological, physical, temporal and positional informations, we build a spatialized and dynamic database [56]. 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 digital atlas is mainly descriptive. However, like for classical databases, specific tools make it possible to explore the digital 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. Current developments of this tool consist in using directly the segmented images produced from laser microscopy to build the atlas. To better represent the development of a biological population, a method to compute an "average" structure is investigated.

5.3.2. *Shape analysis of meristems*

Participants: Jonathan Legrand, Pierre Fernique, Frédéric Boudon, Yann Guédon, Christophe Godin, Pradeep Das [RDP, ENS], Arezki Boudaoud [RDP, ENS].

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). Looking at homogeneous regions in the variable value space, cellular patterns can be identified. This work was developed in the context of the PhD of Jonathan Legrand with contributions of Pierre Fernique, another PhD student, that has been defended this year.

Considering the highly-structured nature of our data (time and space structuring) and the potential diversity and heterogeneity of possible cell descriptors, we developed two complementary approaches:

- A first one that favours the spatial structuring: In this approach, the cell neighbourhood and the cell descriptors are jointly taken into account in a clustering approach whose objective is to identify a small number of clusters corresponding to well-defined cell identities. Once the cells have been labelled using the clustering algorithm, cell generation distributions are estimated on the basis of the labelled lineage trees.

- A second one that favours the temporal structuring: In this approach, the data of interest are lineage forest and the only spatial structuring taken into account corresponds to siblings with respect to a given parent cell. In a first step, cell identities are inferred on the basis of the cell descriptors taking into account lineage relationships using hidden Markov tree models and the spatial regions that emerge from the cell identity labelling are then characterized. This second approach is supported by the fact that cell topology is only affected by division which makes highly relevant the local spatial information taken into account in this approach.

5.3.3. *Mechanical model*

Participants: Jean-Philippe Bernard, Olivier Ali, Christophe Godin, Benjamin Gilles, Frédéric Boudon, Ibrahim Cheddadi, Jan Traas [ENS-Lyon], Olivier Hamant [ENS-Lyon], Arezki Boudaoud [ENS-Lyon].

This research theme is supported by the Inria Project Lab Morphogenetics and the Jan Traas's ERC.

The rigid cell walls that surround plant cells are responsible for the acquisition of organ shapes. These walls are submitted to stresses due to cell turgor pressure. Wall deformation is caused by the turgor forces in the cell walls. Wall synthesis is triggered by these wall deformations when some specific threshold is exceeded. The final shape of the tissue integrates mechanically all the local deformations of each cell.

To quantify this growth process at the level of a multicellular tissue, we developed a model of growth that integrates mechanical forces development at cellular resolution. In this model, walls are characterized by their mechanical properties, e.g. elasticity, extensibility and anisotropy. For this, 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 will appear in PLoS Computational Biology in 2015. We used this framework to investigate the influence of a specific signalling cascade (the ABP1- Kat1 signalling pathway) and its putative mechanical consequences on primordium initiation [25]. A review of the different mechanical concepts underlying plant morphogenesis has been carried out in [12].

In our first approach, the mechanical model rely on a finite element method (FEM) to describe the deformation of the tissue. In FEM, the tissue is represented by a mesh. The positions of the vertices at each time step are estimated from a linear system. If the tissue is big or if the mesh is fine, the linear system can be large and thus leads to computational overheads. An alternative way to classical FEM is to use a meshless method where the deformation of the tissue can be characterized by a linear combination of deformations of a finite and small set of frames. Because shape functions are no longer defined on each element but on the whole tissue, they have to be updated at each growth step by estimating a new rest configuration. With meshless method, the discretization of the system can be dynamically updated parsimoniously according to the precision required to model the emergence of shapes (PhD work of Jean-Philippe Bernard).

5.3.4. *Gene regulatory networks: Design of a genetic model of inflorescence development.*

Participants: Eugenio Azpeitia, Christophe Godin, François Parcy, Etienne Farcot.

This research theme is supported by the Inria Project Lab Morphogenetics.

Modeling gene activities within cells is of primary importance since cell identities correspond to stable combination of gene expression.

We studied the regulatory network that controls the flowering transition 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. Three different network models were done and validate. A first Boolean version, a second fuzzy logic and an ODEs models were studied. The models are able to correctly recover the gene steady states observed in the meristems during the flower transitions, the gene transitions and the mutant effects. Importantly, the model is able to explain the cauliflower mutants. This work couples models at different scales, since the gene regulatory network 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. Finally, some links between gene regulation and plant growth have been identified. These links can be experimentally tested which could lead to a first integrated picture of flower development could be reached in the context of Eugenio Azpeitia postdoc.

5.3.5. *Model integration*

Participants: Frédéric Boudon, Christophe Godin, Guillaume Baty, Jan Traas.

This research theme is supported by the Morphogenetics Inria Project Lab.

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 group at ENS-Lyon [58] and the CPIB group in Nottingham, UK [47].

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*. In the past year, progress has been made in defining a generic tissue data structure that could be used in this platform. Currently, robust geometric operations such as division are implemented and tested. Moreover, a redesign of the structure based on more elaborated formalisms such as combinatorial maps is investigated.

5.4. Multi-scale models and analysis: from cells to plant architecture (and back)

5.4.1. *Modeling water transport 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.

5.4.2. *Mechanical modeling of fruit growth*

Participants: Ibrahim Cheddadi [Inra, Avignon], Mik Cieslak [U. Calgary], Frédéric Boudon, Valentina Baldazzi [Inra, Avignon], Nadia Bertin [Inra, Avignon], Michel Genard [Inra, Avignon], Christophe Godin.

This research theme is supported by the Agropolis project MecaFruit3D.

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, as well as mechanical interactions between tissue parts that are essential to properly model growth.

To help characterizing the 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 is being investigated, combining two tools previously developed in the team: on the one hand, a modeling pipeline that creates a 3D volumetric mesh of the internal fruit structure, including vasculature (see section 3), and couples it with water and carbon transport; on the other hand, a mechanical description of the growth of plant tissues (see section 5.3.3): growth is related to the extension of the cell walls, which is triggered when the so-called turgor pressure inside the cells exceeds a given threshold. The global shape of the tissue integrates mechanically all the local deformations of each cell.

In order to couple these two aspects of plant growth, we describe how volume variations are constrained by fluxes of matter, and how these fluxes depend on mechanical and physiological parameters. The corresponding set of equations are resolved thanks to the SOFA finite elements software.

This approach will be applied to study tomato fruit. 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.
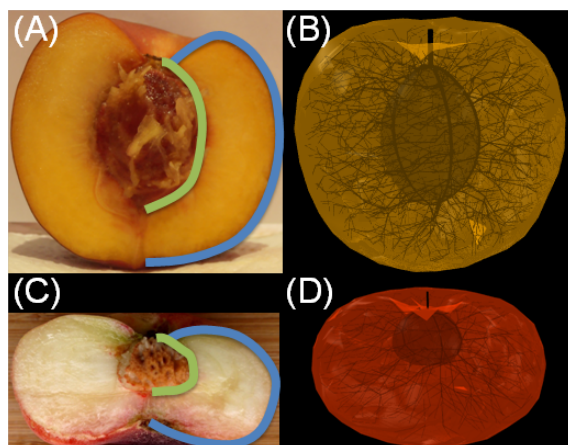
Figure 5. Virtual models of peaches reconstructed from images with simulated vasculatur to simulate carbon and water transport in the fruit

5.4.3. *Analyzing root growth and branching*

Participants: Beatriz Moreno Ortega, Sixtine Passot, Yann Guédon, Laurent Laplaze [IRD, DIADE], Mikaël Lucas [IRD, DIADE], Bertrand Muller [INRA, LEPSE].

This research theme is supported by two PhD programmes.

New 2D and 3D root phenotyping plateforms are emerging with associated image analysis toolbox (e.g. SmartRoot). The analysis of complex root phenotyping data is thus a new challenge in developmental biology.

We aim at developing a pipeline of methods for analyzing root systems at three scales:

1. tissular scale to identify and characterize the meristem, elongation and mature zones along a root using piecewise heteroscedastic linear models,
2. individual root scale to analyze the dynamics of root elongation,
3. root system scale to analyze the branching structure.

This pipeline of analysis methods will be applied to different species (maize, millet and *arabidopsis*) and for different biological objectives (study of genetic diversity for millet and of metabolic and hormonal controls of morphogenesis for maize).

5.4.4. *Analyzing shoot and leaf elongation*

Participants: Maryline Lièvre, Yann Guédon, Leo Guignard, Christine Granier [INRA, LEPSE].

This research theme is supported by one PhD programme.

This study is based on the observation of a lack of methods enabling the integrated analysis of the processes controlling the vegetative development in *Arabidopsis thaliana*. We developed a pipeline of analysis methods combining image analysis techniques and statistical models to integrate the measurements made at the leaf and shoot scales. Semi-Markov switching models were built for different genotypes, allowing a more thorough characterization of the studied mutants. These models validated the hypothesis that the rosette can be structured into successive developmental phases that could change depending on the genotype. They also highlighted the structuring role of the abaxial trichome trait, although the developmental phases cannot be explained entirely by this trait. We developed a second pipeline of analysis methods combining a semi-automatic method for segmenting leaf epidermis images based on the ilastik software, and the analysis of the obtained cell areas using a gamma mixture model whose parameters of gamma components are tied by a scaling rule. This model allowed us to estimate the distribution of the number of endocycles. We highlighted in this way that the mean number of endocycles changes drastically with leaf rank. Finally, we built a multi-scale model that integrates tissular, morphological, dynamical and dimension traits for each successive leaf along the shoot. This model gave us for the first time an integrative view of the development of the *Arabidopsis* rosette.

5.4.5. *Analyzing perturbations in Arabidopsis thaliana phyllotaxis*

Participants: Yassin Refahi, Fabrice Besnard, Yann Guédon, Christophe Godin, Etienne Farcot, Teva Vernoux [RDP, ENS].

This research theme has been supported by iSam, IBC and the Inria Project Lab Morphogenetics.

The geometric arrangement of lateral organs along plant stems, named phyllotaxis, shows a variety of striking patterns with remarkable regularities and symmetries. This has interested biologists, physicists, mathematicians and computer scientists for decades. These studies have lead to a commonly accepted standard interpretation of phyllotaxis that postulates that organs inhibit the formation of new organs in their vicinity. At a molecular scale, these inhibitory fields have been shown to result from the spatio-temporal distribution of the plant hormone auxin. This model theoretically explains a large part of the diversity of phyllotactic patterns observed in plants.

The cytokinin hormones are known to play a significant role in the regulation of phyllotaxis. Fabrice Besnard and Teva Vernoux realized that *Arabidopsis thalianaahp6* mutants, which are perturbed in the cytokinin signaling pathway, showed unusual chaotic perturbations of the phyllotaxis at macroscopic level.

In order to characterize these perturbations, we designed a pipeline of models and methods [60], [53] 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.

We then looked at the origin of these permutations using confocal microscopy and realized that organs were in fact frequently co-initiated in the mutant, leading after development randomly in half of the cases to permutations. We concluded that the mutant is actually perturbed in the time between consecutive organ initiation (i.e. the plastochrone), while relative angular positions are not affected. After closer inspection, we realized that the mutated gene encodes a protein diffusing from the organs and creating a field around the organs that regulates the plastochrone. We could demonstrate that in the mutant, the absence of this field leads to co-initiations and subsequently to the observed permutations.

Altogether, this study sheds a new light on our interpretation of phyllotaxis, revisiting the standard model and suggesting that several fields based on auxin and cytokinin with different properties are required to provide robustness to phyllotaxis. An overview of this work has been published in the journal Nature [15].

5.4.6. *A stochastic model of phyllotaxis*
Participants: Yassin Refahi, Christophe Godin, Etienne Farcot, Teva Vernoux [RDP, ENS].

This research theme has been supported by IBC and the Inria Project Lab Morphogenetics.

To proceed further and find a mechanistic interpretation of these results, we are currently developing a stochastic extension of the standard model of phyllotaxis. We first analyzed the properties of the inhibitory fields created by the existing primordia on the initiation of new promordia, and concluded that the angular positions of organs are very robust to perturbations while plastochrons may be dramatically affected. This suggested that there exists a strong decoupling between space and time in the patterning process. To account for this observation, we modeled the perception of the initiation signal by cells using stochastic processes coupled with the intensity of inhibitory fields and showed that the observed permutation patterns emerge spontaneously from this purely local processes. A paper describing this model will be submitted for publication in 2015.

5.4.7. *The role of auxin and sugar in rose bud outgrowth control*
Participants: Jessica Bertheloot [INRA, Angers], Frédéric Boudon, Christophe Godin.

Auxin in the stem is known to be a key regulator of apical dominance. Over the last decades, many studies have been undertaken to understand its action mode, which is indirect because auxin in the main stem does not enter into the bud. Recently, apical dominance over basal buds in pea has been related to low sugar availability caused by high sugar demand of growing apical organs. Auxin and sugar are two signals regulating the entrance of bud into sustained growth in opposite ways. Recently it has been demonstrated that sugar effect on bud outgrowth was preceded by a modification of the hormonal network involved in auxin effect, which suggests that auxin and sugar pathways do interact in a non-trivial way. But auxin and sugar effects have been studied separately until now. In this work, we investigate what is the combined effect of sugar and auxin on bud outgrowth, and how they integrate to control bud entrance into sustained growth. For this, a series of experiments has been carried on a single-node cuttings of Rosa hybrida grown in vitro in which different combinations of sugar and auxin levels have been tested. A model the regulatory networks controling stem-bud molecular interaction is currently being investigated.

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

6. New Results

6.1. Network Design and Management

Participants: Jean-Claude Bermond, David Coudert, Frédéric Giroire, Frédéric Havet, Alvinice Kodjo, Aurélien Lancin, Bi Li, Fatima Zahra Moataz, Christelle Molle-Caillouet, Joanna Moulierac, Nicolas Nisse, Stéphane Pérennes, Truong Khoa Phan.

More information on several results presented in this section may be found in the PhD thesis of A. Kodjo [13], B. Li [15] and T. K. Phan [18].

6.1.1. Optimization in backbone networks

6.1.1.1. Shared Risk Link Group

The notion of Shared Risk Link Groups (SRLG) captures survivability issues when a set of links of a network may fail simultaneously. The theory of survivable network design relies on basic combinatorial objects that are rather easy to compute in the classical graph models: shortest paths, minimum cuts, or pairs of disjoint paths. In the SRLG context, the optimization criterion for these objects is no longer the number of edges they use, but the number of SRLGs involved. Unfortunately, computing these combinatorial objects is NP-hard and hard to approximate with this objective in general. Nevertheless some objects can be computed in polynomial time when the SRLGs satisfy certain structural properties of locality which correspond to practical ones, namely the star property (all links affected by a given SRLG are incident to a unique node) and the span 1 property (the links affected by a given SRLG form a connected component of the network). The star property is defined in a multi-colored model where a link can be affected by several SRLGs while the span property is defined only in a mono-colored model where a link can be affected by at most one SRLG. In [52], we extend these notions to characterize new cases in which these optimization problems can be solved in polynomial time or are fixed parameter tractable. We also investigate on the computational impact of the transformation from the multi-colored model to the mono-colored one. Experimental results are presented to validate the proposed algorithms and principles.

6.1.1.2. Dynamic Routing and Spectrum Assignment in Optical Networks

Elastic Optical Networks (EONs) promises a better utilization of the spectrum in optical networks. In fact, as the optical transmission spectrum is carved into fixed-length bands in the traditional WDM networks, small bit rates are over-provisioned and very high bit rates do not fit. EONs are moving away from this fixed-grid and allow the spectrum to be divided flexibly: each request is allocated exactly the resources it needs. In [34], we present two exact algorithms to route and allocate spectrum to a new request in an EON using only Non-Disruptive Defragmentation (Push-Pull). In the first algorithm, we find the shortest routing path for the new request (i.e., the shortest path from source to destination where contiguous spectrum to satisfy the request can be freed) and then find the position that gives the overall minimum delay on that path. In the second algorithm, we find at the same time a routing path and a position in the spectrum, that minimize the delay of insertion (over all other paths and positions). Both algorithms are polynomial in the size of the network, its bandwidth and the number of provisioned requests.

6.1.2. Microwave Backhaul networks

6.1.2.1. Chance-Constrained Optimization of Reliable Backhaul networks

In [25], we extend our former investigation on conceiving reliable fixed point-to-point wireless networks under outage probability constraints. We consider the problem of determining the minimum cost bandwidth assignment of a network, while guaranteeing a reliability level of the solution. If the optimal bandwidth assignment and routing of traffic demands are accomplished, the reliability criterion requires that network flows remain feasible with high probability, regarding that the performance of microwave links is prone to

variations due to external factors, e.g., weather. We introduce a *chance-constrained programming* approach to tackle this problem and we present reformulations to standard integer linear programming models, including a budget-constrained formulation. To improve the solving performance, we propose new valid inequalities and a primal heuristic. Computational results present a performance analysis of the valid inequalities and the heuristic. Further, the outperformance of the novel model compared to more traditional approaches is documented.

6.1.2.2. Robust optimization in multi-operators microwave backhaul networks

In [41], we consider the problem of sharing the infrastructure of a backhaul network for routing. We investigate on the revenue maximization problem for the physical network operator (PNO) when subject to stochastic traffic requirements of multiple virtual network operators (VNO) and prescribed service level agreements (SLA). We use robust optimization to study the tradeoff between revenue maximization and the allowed level of uncertainty in the traffic demands. This mixed integer linear programming model takes into account end-to-end traffic delays as example of quality-of-service requirement in a SLA. To show the effectiveness of our model, we present a study on the price of robustness, i.e. the additional price to pay in order to obtain a feasible solution for the robust scheme, on realistic scenarios.

6.1.3. Energy efficiency

6.1.3.1. Robust Optimization for Energy-aware Routing with Redundancy Elimination

Many studies in literature have shown that energy-aware routing (EAR) can significantly reduce energy consumption for backbone networks. Also, as an arising concern in networking research area, the protocol-independent traffic redundancy elimination (RE) technique helps to reduce (a.k.a compress) traffic load on backbone network. In [35], [50], we present an extended model of the classical multi-commodity flow problem with compressible flows. Our model is robust with fluctuation of traffic demand and compression rate. In details, we allow any set of a predefined size of traffic flows to deviate simultaneously from their nominal volumes or compression rates. As an applicable example, we use this model to combine redundancy elimination and energy-aware routing to increase energy efficiency for a backbone network. Using this extra knowledge on the dynamics of the traffic pattern, we are able to significantly increase energy efficiency for the network. We formally define the problem and model it as a Mixed Integer Linear Program (MILP). We then propose an efficient heuristic algorithm that is suitable for large networks. Simulation results with real traffic traces on Abilene, Geant and Germany50 networks show that our approach allows for 16-28% extra energy savings with respect to the classical EAR model.

6.1.3.2. Optimizing IGP Link Weights for Energy-efficiency in a Changing World

Recently, due to the increasing power consumption and worldwide gas emissions in ICT (Information and Communication Technology), energy efficient ways to design and operate backbone networks are becoming a new concern for network operators. Since these networks are usually overprovisioned and since traffic load has a small influence on power consumption of network equipments, the most common approach to save energy is to put unused line cards that drive links between neighbouring routers into sleep mode. To guarantee QoS, all traffic demands should be routed without violating capacity constraints and the network should keep its connectivity. From the perspective of traffic engineering, we argue that stability in routing configuration also plays an important role in QoS. In details, frequent changes in network configuration (link weights, slept and activated links) to adapt with traffic fluctuation in daily time cause network oscillations. We propose in [62] a novel optimization method to adjust the link weights of Open Shortest Path First (OSPF) protocol while limiting the changes in network configurations when multi-period traffic matrices are considered. We formally define the problem and model it as Mixed Integer Linear Program (MILP). We then propose an efficient heuristic algorithm that is suitable for large networks. Simulation results with real traffic traces on three different networks show that our approach achieves high energy saving while keeping the networks in stable state (less changes in network configuration).

6.1.3.3. Grid spanners with low forwarding index for energy efficient networks

A routing R of a connected graph G is a collection that contains simple paths connecting every ordered pair of vertices in G. The edge-forwarding index with respect to R (or simply the forwarding index with respect to R) $\pi(G, R)$ of G is the maximum number of paths in R passing through any edge of G. The forwarding index $\pi(G)$ of G is the minimum $\pi(G, R)$ over all routings R's of G. This parameter has been studied for different graph classes. Motivated by energy efficiency, we look in [57], for different numbers of edges, at the best spanning graphs of a square grid, namely those with a low forwarding index.

6.1.4. Software-Defined Networks

6.1.4.1. Rule Placement in Software-Defined Networks for Energy-aware Routing

Software-defined Networks (SDN), in particular OpenFlow, is a new networking paradigm enabling innovation through network programmability. Over past few years, many applications have been built using SDN such as server load balancing, virtual-machine migration, traffic engineering and access control. We focus on using SDN for energy-aware routing (EAR). SDN can collect traffic matrix and then computes routing solutions satisfying QoS while being minimal in energy consumption (with minimal number of active links). However, prior works on EAR have assumed that the table of OpenFlow switch can hold an infinite number of rules. In practice, this assumption does not hold since the flow table is implemented with Ternary Content Addressable Memory (TCAM) which is expensive and power-hungry. In [39], [56], we propose an optimization method to minimize energy consumption for a backbone network while respecting capacity constraints on links and rule space constraints on routers. In details, we present an exact formulation using Integer Linear Program (ILP) and introduce efficient greedy heuristic algorithm. Based on simulations, we show that using this smart rule space allocation, it is possible to save almost as much power consumption as the classical EAR approach.

6.1.4.2. Compressing Two-dimensional Routing Tables with Order

A communication in a network is a pair of nodes (s, t). The node s is called the source source and t the destination. A communication set is a set of distinct communications, i.e. two communications might have the same source or the same destination, but they cannot have both same source and same destination. A routing of a communication (s, t) is a path in the network from s to t. A routing of a communication set is a union of routings of its communications. At each node, there is a set X of communications whose routing path goes through this node. The node needs to be able to find for each communication (s, t) in X, the port that the routing path of (s, t) uses to leave it. An easy way of doing it is to store the list of all triples (s, t, k), where $(s, t) \in X$ and k is the port used by the (s, t)-path to leave the node. Such triples are called communication triples. However, such a list might be very large. Motivated by routing in telecommunication network using Software Defined Network Technologies, we consider in [55] the problem of compacting this list using aggregation rules. Indeed, SDN routers use specific memory which is expensive and of small capacity. Hence, in addition, we can use some additional triples, called *-triples. As an example, a t-destination triple $(*, t, p)$, means that every communication with destination t leaves on port p. We carry out in this work a study of the problem complexity, providing results of NP-completeness, of Fixed-Parameter Tractability and approximation algorithms.

6.1.5. Data gathering in radio networks

In the gathering problem, a particular node in a graph, the base station, aims at receiving messages from some nodes in the graph. At each step, a node can send one message to one of its neighbors (such an action is called a call). However, a node cannot send and receive a message during the same step. Moreover, the communication is subject to interference constraints ; more precisely we consider a binary interference model where two calls interfere in a step, if the sender of one call is at distance at most d_I from the receiver of the other call. Given a graph with a base station and a set of nodes having some messages, the goal of the gathering problem is to compute a schedule of calls for the base station to receive all messages as fast as possible, i.e., minimizing the number of steps (called makespan). The gathering problem is equivalent to the personalized broadcasting problem where the base station has to send messages to some nodes in the graph, with same transmission constraints. In [23], we focus on the gathering and personalized broadcasting problem in grids. Moreover, we

consider the non-buffering model: when a node receives a message at some step, it must transmit it during the next step. In this setting, though the problem of determining the complexity of computing the optimal makespan in a grid is still open, we present linear (in the number of messages) algorithms that compute schedules for gathering with $d_I \in \{0, 1, 2\}$. In particular, we present an algorithm that achieves the optimal makespan up to an additive constant 2 when $d_I = 0$. If no messages are "close" to the axes (the base station being the origin), our algorithms achieve the optimal makespan up to an additive constant 1 when $d_I = 0$, 4 when $d_I = 2$, and 3 when both $d_I = 1$ and the base station is in a corner. Note that, the approximation algorithms that we present also provide approximation up to a ratio 2 for the gathering with buffering. All our results are proved in terms of personalized broadcasting.

6.2. Graph Algorithms

Participants: Julio Araújo, Jean-Claude Bermond, David Coudert, Guillaume Ducoffe, Frédéric Giroire, Aurélien Lancin, Bi Li, Fatima Zahra Moataz, Christelle Molle-Caillouet, Nicolas Nisse, Stéphane Pérennes.

COATI 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. More information on several results presented in this section may be found in PhD thesis of B. Li [15] and A. Lancin [14], and in the Habilitation thesis of N. Nisse [17].

6.2.1. *Complexity and Computation of Graph Parameters*

We use graph theory to model various network 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 from which the problems are polynomial-time solvable.

6.2.1.1. Hyperbolicity

The Gromov hyperbolicity is an important parameter for analyzing complex networks since it expresses how the metric structure of a network looks like a tree. In other words, it provides bounds on the stretch resulting from the embedding of a network topology into a weighted tree. It is therefore used to provide bounds on the expected stretch of greedy-routing algorithms in Internet-like graphs. However, the best known algorithm for computing this parameter has time complexity in $O(n^{3.69})$, which is prohibitive for large-scale graphs.

In [47], we investigate some relations between the hyperbolicity of a graph and the hyperbolicity of its *atoms*, that are the subgraphs resulting from the decomposition of the graph according its clique minimal separators. More precisely, we prove that the maximum hyperbolicity taken over all the atoms is at least the hyperbolicity of *G minus one*. We also give an algorithm to slightly modify the atoms, which is at no extra cost than computing the atoms themselves, and so that the maximum hyperbolicity taken over all the resulting graphs is *exactly* the hyperbolicity of G. An experimental evaluation of our methodology is provided for large collaboration networks. Finally, we deduce from our theoretical results the first *linear-time* algorithm to compute the hyperbolicity of an outerplanar graph.

The shortest-path metric d of a connected graph G is 1/2-hyperbolic if, and only if, it satisfies $\mathrm{d}(u,v) + \mathrm{d}(x,y) \leq \max\{\mathrm{d}(u,x) + \mathrm{d}(v,y), \mathrm{d}(u,y) + \mathrm{d}(v,x)\} + 1$, for every 4-tuple u, x, v, y of G. We show in [26], [48] that the problem of deciding whether an unweighted graph is 1/2-hyperbolic is subcubic equivalent to the problem of determining whether there is a chordless cycle of length 4 in a graph. An improved algorithm is also given for both problems, taking advantage of fast rectangular matrix multiplication. In the worst case it runs in $O(n^{3.26})$-time.

6.2.1.2. Branch and Bound Algorithm for computing Pathwidth

It is well known that many NP-hard problems are tractable in the class of bounded pathwidth graphs. In particular, path-decompositions of graphs are an important ingredient of dynamic programming algorithms for solving such problems. Therefore, computing the pathwidth and associated path-decomposition of graphs has both a theoretical and practical interest. In [36], [51], we design a Branch and Bound algorithm that computes the exact pathwidth of graphs and a corresponding path-decomposition. Our main contribution consists of

several non-trivial techniques to reduce the size of the input graph (pre-processing) and to cut the exploration space during the search phase of the algorithm. We evaluate experimentally our algorithm by comparing it to existing algorithms of the literature. It appears from the simulations that our algorithm offers a significative gain with respect to previous work. In particular, it is able to compute the exact pathwidth of any graph with less than 60 nodes in a reasonable running-time (10 min.). Moreover, our algorithm also achieves good performance when used as a heuristic (i.e., when returning best result found within bounded time-limit). Our algorithm is not restricted to undirected graphs since it actually computes the vertex-separation of digraphs (which coincides with the pathwidth in case of undirected graphs).

6.2.1.3. To satisfy impatient Web surfers is hard

Prefetching is a basic mechanism for faster data access and efficient computing. An important issue in prefetching is the tradeoff between the amount of network's resources wasted by the prefetching and the gain of time. For instance, in the Web, browsers may download documents in advance while a Web surfer is surfing. Since the Web surfer follows the hyperlinks in an unpredictable way, the choice of the Web pages to be prefetched must be computed online. The question is then to determine the minimum amount of resources used by prefetching that ensures that all documents accessed by the Web surfer have previously been loaded in the cache. In [28], we model this problem as a two-player game similar to Cops and Robber Games in graphs. Let $k \geq 1$ be any integer. The first player, a fugitive, starts on a marked vertex of a (di)graph G. The second player, an observer, marks at most k vertices, then the fugitive moves along one edge/arc of G to a new vertex, then the observer marks at most k vertices, etc. The fugitive wins if it enters an unmarked vertex, and the observer wins otherwise. The surveillance number of a (di)graph is the minimum k such that the observer marking at most k vertices at each step can win against any strategy of the fugitive. We also consider the connected variant of this game, i.e., when a vertex can be marked only if it is adjacent to an already marked vertex. We study the computational complexity of the game. All our results hold for both variants, connected or unrestricted. We show that deciding whether the surveillance number of a chordal graph is at most 2 is NP-hard. We also prove that deciding if the surveillance number of a DAG is at most 4 is PSPACE-complete. Moreover, we show that the problem of computing the surveillance number is NP-hard in split graphs. On the other hand, we provide polynomial-time algorithms computing surveillance numbers of trees and interval graphs. Moreover, in the case of trees, we establish a combinatorial characterization of the surveillance number.

6.2.2. *Tree-decompositions*

6.2.2.1. Minimum Size Tree-Decompositions

Tree-Decompositions are the corner-stone of many dynamic programming algorithms for solving graph problems. Since the complexity of such algorithms generally depends exponentially on the width (size of the bags) of the decomposition, much work has been devoted to compute tree-decompositions with small width. However, practical algorithms computing tree-decompositions only exist for graphs with treewidth less than 4. In such graphs, the time-complexity of dynamic programming algorithms based on tree-decompositions is dominated by the size (number of bags) of the tree-decompositions. It is then interesting to try to minimize the size of the tree-decompositions. In [42], [60], we consider the problem of computing a tree-decomposition of a graph with width at most k and minimum size. More precisely, we focus on the following problem: given a fixed $k \geq 1$, what is the complexity of computing a tree-decomposition of width at most k with minimum size in the class of graphs with treewidth at most k? We prove that the problem is NP-complete in planar graphs for any fixed $k \geq 4$ and polynomial for $k \leq 2$. We also show that for $k = 3$ the problem can be solved in polynomial time in the class of trees and 2-connected outerplanar graphs.

6.2.2.2. Exclusive Graph Searching vs. Pathwidth

In Graph Searching, a team of searchers aims at capturing an invisible fugitive moving arbitrarily fast in a graph. Equivalently, the searchers try to clear a contaminated network. The problem is to compute the minimum number of searchers required to accomplish this task. Several variants of Graph Searching have been studied mainly because of their close relationship with the pathwidth of a graph. Blin et al. defined the Exclusive Graph Searching where searchers cannot "jump" and no node can be occupied by more than one searcher. In [61], we study the complexity of this new variant. We show that the problem is NP-hard in

planar graphs with maximum degree 3 and it can be solved in linear time in the class of cographs. We also show that monotone Exclusive Graph Searching is NP-complete in split graphs where Pathwidth is known to be solvable in polynomial time. Moreover, we prove that monotone Exclusive Graph Searching is in P in a subclass of star-like graphs where Pathwidth is known to be NP-hard. Hence, the computational complexities of monotone Exclusive Graph Searching and Pathwidth cannot be compared. This is the first variant of Graph Searching for which such a difference is proved.

6.2.2.3. Diameter of Minimal Separators in Graphs

In [49], we establish general relationships between the topological properties of graphs and their metric properties. For this purpose, we upper-bound the diameter of the *minimal separators* in any graph by a function of their sizes. More precisely, we prove that, in any graph G, the diameter of any minimal separator S in G is at most $\lfloor \frac{\ell(G)}{2} \rfloor \cdot (|S| - 1)$ where $\ell(G)$ is the maximum length of an isometric cycle in G. We refine this bound in the case of graphs admitting a *distance preserving ordering* for which we prove that any minimal separator S has diameter at most $2(|S| - 1)$. Our proofs are mainly based on the property that the minimal separators in a graph G are connected in some power of G.

Our result easily implies that the *treelengthtl*(G) of any graph G is at most $\lfloor \frac{\ell(G)}{2} \rfloor$ times its *treewidthtw*(G). In addition, we prove that, for any graph G that excludes an *apex graphH* as a minor, $tw(G) \leq c_H \cdot tl(G)$ for some constant c_H only depending on H. We refine this constant when G has bounded genus. As a consequence, we obtain a very simple $O(\ell(G))$-approximation algorithm for computing the treewidth of n-node m-edge graphs that exclude an apex graph as a minor in $O(nm)$-time.

6.2.3. Distributed computing with mobile agents

6.2.3.1. Stigmergy of Anonymous Agents in Discrete Environments

Communication by stigmergy consists, for agents/robots devoid of other dedicated communication devices, in exchanging information by observing each other's movements, similar to how honeybees use a dance to inform each other on the location of food sources. Stigmergy, while a popular technique in soft computing (e.g., swarm intelligence and swarm robotics), has received little attention from a computational viewpoint, with only one study proposing a method in a continuous environment. An important question is whether there are limits intrinsic to the environment on the feasibility of stigmergy. While it is not the case in a continuous environment, we show that the answer is quite different when the environment is discrete. In [53], [37], we consider stigmergy in graphs and identifies classes of graphs in which robots can communicate by stigmergy. We provide two algorithms with different tradeoffs. One algorithm achieves faster stigmergy when the density of robots is low enough to let robots move independently. This algorithm works when the graph contains some particular pairwise-disjoint subgraphs. The second algorithm, while slower solves the problem under an extremely high density of robots assuming that the graph admits some large cycle. Both algorithms are described in a general way, for any graph that admits the desired properties and with identified nodes. We show how the latter assumption can be removed in more specific topologies. Indeed, we consider stigmergy in the grid which offers additional orientation information not available in a general graphs, allowing us to relax some of the assumptions. Given an $N \times M$ anonymous grid, we show that the first algorithm requires $O(\mathcal{M})$ steps to achieve communication by stigmergy, where $\mathcal{M}$ is the maximum length of a communication message, but it works only if the number of robots is less than $\lfloor \frac{N \cdot M}{9} \rfloor$. The second algorithm, which requires $O(k^2)$ steps, where k is the number of robots, on the other hand, works for up to $N \cdot M - 5$ robots. In both cases, we consider very weak assumptions on the robots capabilities: i.e., we assume that the robots are anonymous, asynchronous, uniform, and execute deterministic algorithms.

6.2.3.2. Gathering and Exclusive Searching on Rings under Minimal Assumptions

Consider a set of mobile robots with minimal capabilities placed over distinct nodes of a discrete anonymous ring. Asynchronously, each robot takes a snapshot of the ring, determining which nodes are either occupied by robots or empty. Based on the observed configuration, it decides whether to move to one of its adjacent nodes or not. In the first case, it performs the computed move, eventually. The computation also depends on the required task. In [38], we solve both the well-known Gathering and Exclusive Searching tasks. In the former problem, all robots must simultaneously occupy the same node, eventually. In the latter problem, the

aim is to clear all edges of the graph. An edge is cleared if it is traversed by a robot or if both its endpoints are occupied. We consider the exclusive searching where it must be ensured that two robots never occupy the same node. Moreover, since the robots are oblivious, the clearing is perpetual, i.e., the ring is cleared infinitely often. In the literature, most contributions are restricted to a subset of initial configurations. Here, we design two different algorithms and provide a characterization of the initial configurations that permit the resolution of the problems under minimal assumptions.

6.2.4. Enhancing the Web's Transparency

Today's Web services – such as Google, Amazon, and Facebook – leverage user data for varied purposes, including personalizing recommendations, targeting advertisements, and adjusting prices. At present, users have little insight into how their data is being used. Hence, they cannot make informed choices about the services they choose.

To increase transparency, we developed *XRay* [40], the first fine-grained, robust, and scalable personal data tracking system for the Web. XRay predicts which data in an arbitrary Web account (such as emails, searches, or viewed products) is being used to target which outputs (such as ads, recommended products, or prices). XRay's core functions are service agnostic and easy to instantiate for new services, and they can track data within and across services. To make predictions independent of the audited service, XRay relies on the following insight: by comparing outputs from different accounts with similar, but not identical, subsets of data, one can pinpoint targeting through correlation. We show both theoretically, and through experiments on Gmail, Amazon, and YouTube, that XRay achieves high precision and recall by correlating data from a surprisingly small number of extra accounts.

6.2.5. Algorithm design in biology

In COATI, we have recently started a collaboration with EPI ABS (Algorithms Biology Structure) from Sophia Antipolis on minimal connectivity complexes in mass spectrometry based macro-molecular complex reconstruction [63]. This problem turns out to be a minimum color covering problem (minimum number of colors to cover colored edges with connectivity constraints on the subgraphs induced by the colors) of the edges of a graph, and is surprizingly similar to a capacity maximization problem in a multi-interfaces radio network we were studying.

Consider a set of oligomers listing the subunits involved in sub-complexes of a macro-molecular assembly, obtained e.g. using native mass spectrometry or affinity purification. Given these oligomers, connectivity inference (CI) consists of finding the most plausible contacts between these subunits, and minimum connectivity inference (MCI) is the variant consisting of finding a set of contacts of smallest cardinality. MCI problems avoid speculating on the total number of contacts, but yield a subset of all contacts and do not allow exploiting a priori information on the likelihood of individual contacts. In this context, we present in [43] two novel algorithms, ALGO-MILP-W and ALGO-MILP-WB. The former solves the minimum weight connectivity inference (MWCI), an optimization problem whose criterion mixes the number of contacts and their likelihood. The latter uses the former in a bootstrap fashion, to improve the sensitivity and the specificity of solution sets. Experiments on the yeast exosome, for which both a high resolution crystal structure and a large set of oligomers is known, show that our algorithms predict contacts with high specificity and sensitivity, yielding a very significant improvement over previous work. The software accompanying this paper is made available, and should prove of ubiquitous interest whenever connectivity inference from oligomers is faced.

6.3. Structural Graph Theory

Participants: Jean-Claude Bermond, Frédéric Havet, Nicolas Nisse, Ana Karolinna Maia de Oliveira, Stéphane Pérennes.

More information on several results presented in this section may be found in PhD thesis of A. K. Maia de Oliveira [16], and in the Habilitation thesis of N. Nisse [17].

6.3.1. *Graph colouring and applications*

Graph colouring is a central problem in graph theory and it has a huge number of applications in various scientific domains (telecommunications, scheduling, bio-informatics, ...). We mainly study graph colouring problems that model ressource allocation problems.

6.3.1.1. Backbone colouring

A well-known channel assignment problem is the following: we are given a graph G, whose vertices correspond to transmitters, together with an edge-weighting 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 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 in this case is called the q-backbone chromatic number and is denoted $BBC_q(G, H)$. In [30] and [45], we focus on the case when G is planar and H is a forest. In [30], 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). We also discuss a circular version of the problem. In [45], we give some upper bounds when G is planar and has no cycles of length 4 an 5, and G is a tree, and we relate those results to the celebrated Steinberg's Conjecture stating that every planar graphs with no cycles of length 4 or 5 is 3-colourable.

In [29], we consider the list version of this problem (in which each vertex is given a particular list of admissible colours), with particular focus on colours in $\mathbb{Z}_p$ – this problem is closely related to the problem of circular choosability. We first prove that the list circular q-backbone chromatic number of a graph is bounded by a function of the list chromatic number. We then consider the more general problem in which each edge is assigned an individual distance between its endpoints, and provide bounds using the Combinatorial Nullstellensatz. Through this result and through structural approaches, we achieve good bounds when both the graph and the backbone belong to restricted families of graphs.

6.3.1.2. On-line colouring graphs with few P_4s

Various on-line colouring procedures are used. The most widespread ones is the greedy one, which results in a greedy colouring. Given a graph $G = (V; E)$, a *greedy colouring* of G is a proper colouring such that, for each two colours $i < j$, every vertex of $V(G)$ coloured j has a neighbour with colour i. A second optimization procedure consists from time to time to consider the present colouring and to free some colour when possible: if each vertex of a colour class has another colour that is not used by its neighbours, we can recolour each vertex in the calls by another colour. This procedure results in a b-colouring of the graph. A *b-colouring* of a graph G is a proper colouring such that every colour class contains a vertex which is adjacent to at least one vertex in every other colour class. One of the performance measure of such graph is the maximum numbers of colours they could possibly use. The greatest k such that G has a greedy colouring with k colours is the *Grundy number* of G. The greatest integer k for which there exists a b-colouring of G with k colours is its *b-chromatic number*. Determining the Grundy number and the b-chromatic number of a graph are NP-hard problems in general. For a fixed q, the $(q; q - 4)$-graphs are the graphs for which no set of at most q vertices induces more than $q - 4$ distinct induced P_4s paths of order 4). In [24], we obtain polynomial-time algorithms to determine the Grundy number and the b-chromatic number of $(q; q - 4)$-graphs, for a fixed q. They generalize previous results obtained for cographs and P_4-sparse graphs, classes strictly contained in the $(q; q - 4)$-graphs.

6.3.1.3. Weighted colouring

We also studied weighted colouring which models various problems of shared resources allocation. Given a vertex-weighted graph G and a (proper) r-colouring $c = \{C_1, ..., C_r\}$ of G, the weight of a colour class C_i is the maximum weight of a vertex coloured i and the weight of c is the sum of the weights of its colour classes. The objective of the Weighted Colouring Problem is, given a vertex-weighted graph G, to determine the

minimum weight of a proper colouring of G, that is, its *weighted chromatic number*. In [21], [33], we prove that the Weighted Coloring Problem admits a version of the Hajós' Theorem and so we show a necessary and sufficient condition for the weighted chromatic number of a vertex-weighted graph G to be at least k, for any positive real k. The Weighted Colouring Problem problem remains NP-complete in some particular graph classes as bipartite graphs. In their seminal paper, Guan and Zhu asked whether the weighted chromatic number of bounded tree-width graphs (partial k-trees) can be computed in polynomial-time. Surprisingly, the time-complexity of computing this parameter in trees is still open. We show in [21] that, assuming the Exponential Time Hypothesis (3-SAT cannot be solved in sub-exponential time), the best algorithm to compute the weighted chromatic number of n-node trees has time-complexity $n^{\Theta(\log n)}$. Our result mainly relies on proving that, when computing an optimal proper weighted colouring of a graph G, it is hard to combine colourings of its connected components, even when G is a forest.

6.3.1.4. Inducing proper colourings

Frequently, the proper colouring of the graph must be induced by some other parameters that a vertex can compute locally, for example on looking on the labels assigned to its incident edges or to their orientations.

For a connected graph G of order $|V(G)| \geq 3$ and a k-labelling $c : E(G) \to \{1, 2, ..., k\}$ of the edges of G, the *code* of a vertex v of G is the ordered k-tuple $(\ell_1, \ell_2, ..., \ell_k)$, where ℓ_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* $det(G)$ of G. In [31], 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.

An *orientation* of a graph G is a digraph D obtained from G by replacing each edge by exactly one of the two possible arcs with the same endvertices. For each $v \in V(G)$, the *indegree* of v in D, denoted by $d_D^-(v)$, is the number of arcs with head v in D. An orientation D of G is *proper* if $d_D^-(u) \neq d_D^-(v)$, for all $uv \in E(G)$. The *proper orientation number* of a graph G, denoted by $po(G)$, is the minimum of the maximum indegree over all its proper orientations. In [32], [44], we prove that $po(G) \leq \left(\Delta(G) + \sqrt{\Delta(G)}\right)/2 + 1$ if G is a bipartite graph, and $po(G) \leq 4$ if G is a tree. It is well-known that $po(G) \leq \Delta(G)$, for every graph G. However, we prove that deciding whether $po(G) \leq \Delta(G) - 1$ is already an NP-complete problem on graphs with $\Delta(G) = k$, for every $k \geq 3$. We also show that it is NP-complete to decide whether $po(G) \leq 2$, for planar *subcubic* graphs G. Moreover, we prove that it is NP-complete to decide whether $po(G) \leq 3$, for planar bipartite graphs G with maximum degree 5.

6.3.2. 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 undirected graphs.

6.3.2.1. Finding a subdivision of a digraph

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 F, detecting a subdivision of F 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 F. In [16], a number of examples of polynomial instances, several NP-completeness proofs as well as a number of conjectures and open problems are given. In addition, it is conjectured that, for every integer k greater than 1, the directed cycles of length at least k have the Erdős-Pósa Property : for every n, there exists an integer t_n such that for every digraph D, either D contains n disjoint directed cycles of length at least k, or there is a set T of t_n vertices that meets every directed cycle of length at least k. This generalizes a celebrated result of Reed, Robertson, Seymour and Thomas which is the case $k = 2$ of this conjecture. We prove the conjecture

for $k = 3$. We also show that the directed k-Linkage problem is polynomial-time solvable for digraphs with circumference at most 2. From these two results, we deduce that if F is the disjoint union of directed cycles of length at most 3, then one can decide in polynomial time if a digraph contains a subdivision of F.

6.3.2.2. The complexity of finding arc-disjoint branching flows

The concept of arc-disjoint flows in networks is a very general framework within which many well-known and important problems can be formulated. In particular, the existence of arc-disjoint branching flows, that is, flows which send one unit of flow from a given source s to all other vertices, generalizes the concept of arc-disjoint out-branchings (spanning out-trees) in a digraph. A pair of out-branchings $B_{s,1}^+$, $B_{s,2}^+$ from a root s in a digraph $D = (V, A)$ on n vertices corresponds to arc-disjoint branching flows x_1, x_2 (the arcs carrying flow in x_i are those used in $B_{s,i}^+$, $i = 1, 2$) in the network that we obtain from D by giving all arcs capacity $n-1$. It is then a natural question to ask how much we can lower the capacities on the arcs and still have, say, two arc-disjoint branching flows from the given root s. In [46], we prove that for every fixed integer ≥ 2 it is

- an NP-complete problem to decide whether a network $\mathcal{N} = (V, A, u)$ where $u_{ij} = k$ for every arc ij has two arc-disjoint branching flows rooted at s.

- a polynomial problem to decide whether a network $\mathcal{N} = (V, A, u)$ on n vertices and $u_{ij} = n - k$ for every arc ij has two arc-disjoint branching flows rooted at s.

The algorithm for the later result generalizes the polynomial algorithm, due to Lovász, for deciding whether a given input digraph has two arc-disjoint out-branchings rooted at a given vertex. Finally we prove that under the so-called Exponential Time Hypothesis (ETH), for every $\epsilon > 0$ and for every $k(n)$ with $(\log(n))^{1+\epsilon} \leq k(n) \leq \frac{n}{2}$ (and for every large i we have $k(n) = i$ for some n) there is no polynomial algorithm for deciding whether a given digraph contains two arc-disjoint branching flows from the same root so that no arc carries flow larger than $n - k(n)$.

6.3.2.3. Splitting a tournament into two subtournaments with given minimum outdegree

A (k_1, k_2)-*outdegree-splitting* of a digraph D is a partition (V_1, V_2) of its vertex set such that $D[V_1]$ and $D[V_2]$ have minimum outdegree at least k_1 and k_2, respectively. In [58], we show that there exists a minimum function f_T such that every tournament of minimum outdegree at least $f_T(k_1, k_2)$ has a (k_1, k_2)-outdegree-splitting, and $f_T(k_1, k_2) \leq k_1^2/2 + 3k_1/2 + k_2 + 1$. We also show a polynomial-time algorithm that finds a (k_1, k_2)-outdegree-splitting of a tournament if one exists, and returns 'no' otherwise. We give better bound on f_T and faster algorithms when $k_1 = 1$.

6.3.2.4. Eulerian and Hamiltonian dicycles in directed hypergraphs

In [19], we generalize the concepts of Eulerian and Hamiltonian digraphs to directed hypergraphs. A *dihypergraphH* is a pair $(\mathcal{V}(H), \mathcal{E}(H))$, where $\mathcal{V}(H)$ is a non-empty set of elements, called *vertices*, and $\mathcal{E}(H)$ is a collection of ordered pairs of subsets of $\mathcal{V}(H)$, called *hyperarcs*. It is Eulerian (resp. Hamiltonian) if there is a dicycle containing each hyperarc (resp. each vertex) exactly once. We first present some properties of Eulerian and Hamiltonian dihypergraphs. For example, we show that deciding whether a dihypergraph is Eulerian is an NP-complete problem. We also study when iterated line dihypergraphs are Eulerian and Hamiltonian. Finally, we study when the generalized de Bruijn dihypergraphs are Eulerian and Hamiltonian. In particular, we determine when they contain a complete Berge dicycle, i.e. an Eulerian and Hamiltonian dicycle.

DIANA Team

5. New Results

5.1. Highlights of the Year

Arnaud Legout and Thierry Parmentelat designed and realized the very first Inria Mooc hosted on the FUN platform. This Mooc is devoted to the study of the Python language, and targets undergrandudate students. The objective of the course is to give students a thorough understanding of the internal mechanisms of language, and lead them to small and realistic applications. This Mooc was a big success: 9166 persons registered to the course, out of them five hundred followed the whole course and more than a hundred finished the project. For more details on this Mooc see https://www.france-universite-numerique-mooc.fr/courses/inria/41001/Trimestre_4_2014/about.

5.2. From network-level measurements to expected QoE: the Skype use case

Contributors: Salim Afra, Chadi Barakat and Damien Saucez. Applications rely on rich multimedia contents and experience of end users is sensitive to network conditions. Consequently, network operators must design their infrastructure to ensure high Quality of Experience (QoE) for their customers. However, applications are usually over-the-top services on which network operators have no control and users have no mean to tune the network when they undergo poor QoE. In this project, called ACQUA for Application for the Prediction of Quality of Experience at Internet Access, we propose a new approach that allows network operators to determine how their network performance will influence QoE and end users to predict the QoE even before launching their applications. We predict the subjective QoE users will undergo based on the knowledge of objective network performance parameters obtained with active measurements (e.g., delay, loss) and machine learning. With the particular case of Skype calls and using a decision tree, we show that our approach achieves 83% of accuracy when estimating QoE from the delay, bandwidth, and loss. Our approach can be seen as a new way of performing measurements at the Internet access, where instead of expressing the expected performance in terms of network-level measurements, the performance of the access is expressed in clear terms related to the expected quality for the main applications of interest to the end user. The strength of the approach is in its capacity of expressing directly the QoE as a function of network-level measurements, which is an enabler for QoE prediction, and in reusing the same network-level measurements as input to different models for the QoE of end user applications. More details on this approach and on our application ACQUA can be found in section 4.5 , in the report summarizing the results [24] and on the application web page http://team.inria.fr/diana/acqua/.

5.3. Understanding of modern web traffic

Contributors: Salim Afra, Chadi Barakat, Byungchul Park and Damien Saucez.

Mobile devices are everywhere nowadays but little is known about the way they differ from traditional non-mobile devices in terms of usage and the characteristics of the web traffic they generate. In this contribution, we propose a first study of the differences that exist between mobile and non-mobile Web traffic seen from the lorgnette of a university campus network. The study is performed at different levels starting from users' behavior to transport protocol configurations. Our main findings are that mobile users often browse websites tailored to their devices. They show a significant adoption of Apps to browse the web and a preference for multimedia content. The different way of conceiving the web for mobiles is reflected at the HTTP and TCP levels with much less HTTP redirections and abrupt TCP connection terminations. Interestingly, mobile traffic carries larger contents and have larger TCP flows than non-mobile traffic. By cross-analysis of protocols and users' behavior, we explain why TCP flows in mobile traffic are larger than those of non-mobiles. Further details on this study can be found in [30].

5.4. Characterizing ICMP Rate Limitation on Routers

Contributors: Chadi Barakat and Ricardo Ravaioli.

In the last decade, path discovery has been extensively covered in the literature. In its simplest form, it generally works by sending probes that expire along the path from a host to a destination. It is also known that network administrators often configure their routers to limit the amount of ICMP replies sent, a common practice typically referred to as ICMP rate limitation. In this contribution we attempt to characterize the responsiveness of routers to expiring ICMP echo-request packets. Our contribution is twofold: first, we provide a detailed analysis of how routers are most commonly configured to respond to expiring packets; next, we show that for the vast majority of routers the measured round-trip time is not affected by the probing rate. This contribution is published in ICC'2015 [21]. It is the result of a collaboration with the SIGNET group at I3S in the context of a PhD thesis funded by the UCN@SOPHIA Labex.

5.5. Studying Social Networks at Scale: Macroscopic Anatomy of the Twitter Social Graph

Contributors: Maksym Gabielkov and Arnaud Legout.

Twitter is one of the largest social networks using exclusively directed links among accounts. This makes the Twitter social graph much closer to the social graph supporting real life communications than, for instance, Facebook. Therefore, understanding the structure of the Twitter social graph is interesting not only for computer scientists, but also for researchers in other fields, such as sociologists. However, little is known about how the information propagation in Twitter is constrained by its inner structure. We have performed an in-depth study of the macroscopic structure of the Twitter social graph unveiling the highways on which tweets propagate, the specific user activity associated with each component of this macroscopic structure, and the evolution of this macroscopic structure with time for the past 6 years. For this study, we crawled Twitter to retrieve all accounts and all social relationships (follow links) among accounts; the crawl completed in July 2012 with 505 million accounts interconnected by 23 billion links. Then, we proposed a methodology to unveil the macroscopic structure of the Twitter social graph. This macroscopic structure consists of 8 components defined by their connectivity characteristics. Each component group users with a specific usage of Twitter. For instance, we identified components gathering together spammers, or celebrities. Finally, we presented a method to approximate the macroscopic structure of the Twitter social graph in the past, validate this method using old datasets, and discuss the evolution of the macroscopic structure of the Twitter social graph during the past 6 years. This study was published in ACM Sigmetrics 2014 [17].

5.6. When AIMD meets ICN: a bandwidth sharing perspective

Contributors: Chadi Barakat and Damien Saucez.

Information-centric networking (ICN) leverages content demand redundancy and proposes in-network caching to reduce network and servers load and to improve quality of experience. In this contribution, we study the interaction between in-network caching of ICN and Additive Increase Multiplicative Decrease (AIMD) end-to-end congestion control with a focus on how bandwidth is shared, as a function of content popularity and cache provisioning. As caching shortens AIMD feedback loop, the download rate of AIMD is impacted. Supported by an analytical model based on Discriminatory Processor Sharing and real experiments, we observe that popular contents benefit from caching and realize a shorter download time at the expense of unpopular contents, which see their download time inflated by a factor bounded by $1/(1-\rho)$, where ρ is the network load. This bias can be removed by redefining congestion control to be delay independent or by over-provisioning link capacity at the edge so that to compensate for the greediness of popular contents. Further details on this study, which is the result of a collaboration with Politecnico di Bari, can be found in [23].

5.7. On the incentives and incremental deployments of ICN technologies for OTT services

Contributors: Chadi Barakat and Damien Saucez.

With the explosion of broadband Over-The-Top (OTT) services, the Internet is autonomously migrating toward overlay and incrementally deployable content distribution infrastructures. Information-Centric Networking (ICN) technologies are the natural candidates to efficiently distribute popular content to users. However, the strategic incentives in exploiting ICN, for both users and ISPs, are much less understood to date. We hence studied in [15] the strategic incentives for ICN overlay adoption in OTT services based on a game theoretical approach and discussed how OTTs shall shape their prices to motivate ICN overlay usages.

5.8. On ICN Cache Allocation to Content Providers

Contributor: Damien Saucez
Cross-Team Contributors: Mahmoud El Chamie (Maestro)
External contributors: Sahar Hoteit and Stefano Secci from Sorbonne Universités, UPMC Univ Paris 06.

Information Centric Networks (ICNs) allow offloading content distribution from content service providers by means of in-network caching. Despite a rather high maturation in the definition of ICN forwarding techniques, minor attention has been given to the strategic interaction among the multiple ICN stakeholders. We decided to focus on situations involving multiple Content Providers (CPs) and one ICN provider having to give them access to its caches. Intuitively, this situation is prone to high cache contention, in particular at the appealing topology cross-points. To address this problem we propose a resource allocation and pricing framework to support the network provider in the cache allocation to multiple CPs, for situations where CPs have non-overlapping sets of files and untruthful demands need to be avoided. As cache imputations to CPs need to be fair and robust against overclaiming, we evaluated common proportional and max-min fairness (PF, MMF) allocation rules, as well as coalitional game rules, the Nucleolus and the Shapley value. We found that the naive least-recently-used-based ICN approach provides proportional fairness. Moreover, the game-theoretic rules outperform in terms of content access latency the naive ICN approach as well as PF and MMF approaches, while sitting in between PF and MMF in terms of fairness. This paper is under submission [27].

5.9. Demonstrating a unified ICN development and evaluation framework

Contributors: Walid Dabbous, Alina Quereilhac, Damien Saucez and Thierry Turletti.

Information-Centric Networking solutions target world-wide deployment in the Internet. It is hence necessary to have access to a development and evaluation environment which enables both controllable and realistic experimentation to thoroughly understand how ICN solutions would behave in real life deployment. Such solution can be obtained with NEPI that we demonstrated at the ACM Information Centric Networking 2014 conference. In this demonstration, we presented a development and evaluation framework that combines emulation and live prototyping environments to provide ICN designers and implementers the means to build beyond-prototype ICN solutions. This framework is built upon NEPI. We demonstrated the benefits of such integrated approach by showing how complete experimental studies can be carried out with minimum manual intervention and experiment set-up overhead, in both emulation and live environments. More precisely, we demonstrated how to deploy the same experiment in different environment and how NEPI can help to minimise the implementation and operational overhead. This demonstration is summarised in [31].

5.10. Optimizing rules placement in OpenFlow networks: trading routing for better efficiency

Contributors: Chadi Barakat, Xuan Nam Nguyen, Damien Saucez and Thierry Turletti

The idea behind Software Defined Networking (SDN) is to conceive the network as one programmable entity rather than a set of devices to manually configure, and OpenFlow meets this objective. In OpenFlow, a centralized programmable controller installs forwarding rules onto switches to implement policies. However, this flexibility comes at the expense of extra overhead as the number of rules might exceed the memory capacity of switches, which raises the question of how to place most profitable rules on board. Solutions proposed so far strictly impose paths to be followed inside the network. We advocate instead that we can relax routing requirements within the network to concentrate on the final destination to which the traffic should forwarded, not how to route to this destination. In [19] we illustrate the concept, with an optimization problem that gets the maximum amount of traffic delivered according to policies and the actual dimensioning of the network. The traffic that cannot be accommodated is forwarded to the controller that has the capacity to process it further. [19] also demonstrates that our approach permits a better utilization of scarce resources in the network. We extended the work by stating that in many situations (e.g., data-center networks), the exact path followed by packets has not significant impact on performances as long as packets are delivered to their final destination decided by the endpoint policy. It is thus possible to deviate part of the traffic to alternative paths so as to better use network resources without violating the endpoint policy. In [20], we propose a linear optimization model of the rule allocation problem in resource constrained OpenFlow networks with loose routing policies. We show that the general problem is NP-hard and propose a polynomial time heuristic, called OFFICER, that aims at maximizing the amount of carried traffic in under-provisioned networks. Our numerical evaluation on four different topologies show that exploiting various paths allows to increase the amount of traffic supported by the network without significantly increasing the path length.

5.11. A Survey of Software-Defined Networking

Contributors: Bruno Astuto Arouche Nunes, Xuan Nam Nguyen and Thierry Turletti.

We wrote a survey of the emerging field of Software-Defined Networking (SDN). SDN is currently attracting significant attention from both academia and industry. Its field is quite recent, yet growing at a very fast pace. Still, there are important research challenges to be addressed. We look at the history of programmable networks, from early ideas until recent developments. In particular we described the SDN architecture in detail as well as the OpenFlow standard. We provided an overview of current SDN implementations and testing platforms and examined network services and applications that have been developed based on the SDN paradigm. We concluded with a discussion of future directions enabled by SDN ranging from support for heterogeneous networks to Information Centric Networking (ICN). The survey has been published in the IEEE Surveys and Tutorials journal [9]. This paper is among the top downloads on IEEE Exlpore in December 2014. See http://ieeexplore.ieee.org/xpl/browsePopular.jsp?reload=true.

5.12. Software-Defined Networking Enabled Capacity Sharing in User Centric Networks

Contributors: Bruno Astuto Arouche Nunes and Thierry Turletti.

We proposed to use SDN to deploy capacity sharing mechanisms in the context of User Centric Networking (UCN). We consider user-centric networks as a way of considerably mitigating the problem of sharing limited network capacity and resources efficiently and in a fairly manner. UCNs are self-organizing networks where the end user plays an active role in delivering networking functions such as providing Internet access to other users. We propose to leverage the SDN paradigm to enable cooperation between wireless nodes and to provide capacity sharing services in UCNs. Our proposed approach allows coverage of existing network infrastructure (e.g., WiFi or 3GPP) to be extended to other end users or ad hoc networks that would otherwise not be able to have access to network connectivity and services. Moreover, it takes into account current network load and conditions, and QoS requirements of applications. This work has been published in a special issue of Communications Magazine [14].

5.13. Decentralizing SDN's Control Plane

Contributors: Bruno Nunes Astuto and Thierry Turletti.

Motivated by the internets of the future that will likely be considerably larger in size as well as highly heterogeneous and decentralized, we sketched out a framework aiming to enable not only physical, but also logical distribution of the Software-Defined Networking (SDN) control plane. This framework will accomplish network control distribution by defining a hierarchy of controllers that can "match" an internet's organizational– and administrative structure. The main idea is to delegate control between main controllers and secondary controllers in order to accommodate administrative decentralization and autonomy. This work has been presented in a short paper at the IEEE LCN conference [22].

5.14. Extending DCE to emulate Wireless Software Defined Networks.

Participants: Emilio Mancini, Hardik Soni, Thierry Turletti and Walid Dabbous.

Today it is not possible to simulate and evaluate in a realistic way wireless Software Defined Networking solutions. Indeed, the most used SDN emulator tool, Mininet, can only emulate point-to-point physical links using virtual Ethernet pairs (e.g., MAC layer is ignored), and it cannot provide mobility models for wireless nodes. To make the Direct Code Execution module (DCE) able to run Software Defined Networks we started to support OpenFlow NOX controller and Open vSwitch. The actual NOX binary is executed on a simulated ns-3 node. OpenFlow wireless routers are simulated using the Open vSwitch distribution with data-path kernel module support as it is widely used. DCE provides a mechanism to incorporate such a kernel module based application execution. A demonstration has been done at the 17th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems [29].

5.15. On the Performance of the LISP Beta Network

Contributor: Damien Saucez

The future Internet has been a hot topic during the past decade and many approaches towards this future Internet, ranging from incremental evolution to complete clean slate ones, have been proposed. One of the proposition, LISP, advocates for the separation of the identifier and the locator roles of IP addresses to reduce BGP churn and BGP table size. Up to now, however, most studies concerning LISP have been theoretical and, in fact, little is known about the actual LISP deployment performance. In [16], we report the measurement campaigns carried out on the LISP Beta Network. More precisely, we evaluated the performance of the two key components of the infrastructure: the control plane (i.e., the mapping system) and the interworking mechanism (i.e., communication between LISP and non-LISP sites). Our measurements highlight that performance offered by the LISP interworking infrastructure is strongly dependent on BGP routing policies. If we exclude misconfigured nodes, the mapping system typically provides reliable performance and relatively low median mapping resolution delays. Although the bias is not very important, control plane performance favors USA sites as a result of its larger LISP user base but also because European infrastructure appears to be less reliable.

This work resulted in a collaboration with Telecom ParisTech starting in mid-2014 a PhD thesis on the feasibility of large scale measurement of LISP networks with Luigi Iannone as advisor and Damien Saucez as co-advisor.

5.16. Standardization: Contributions to the IETF LISP WG

Contibutor: Damien Saucez

In the context of the LISP WG, we contributed to an Internet-draft called "An Architectural Introduction to the LISP Location-Identity Separation System" [25] that describes the architecture of the Locator/ID Separation Protocol (LISP), making it easier to read the rest of the LISP specifications and providing a basis for discussion about the details of the LISP protocols. This document is used for introductory purposes, more details can be found in RFC6830, the protocol specification. This internet-draft is in RFC queue, for imminent publication as RFC.

In the context of the LISP WG, we contributed to an Internet-draft called "LISP Threats Analysis" [33] that proposes a threat analysis of the Locator/Identifier Separation Protocol (LISP). This internet-draft is under discussion in the Working Group.

In the context of the LISP WG, we contributed to an Internet-draft called "LISP-Security (LISP-SEC)" [28] that specifies LISP-SEC, a set of security mechanisms that provides origin authentication, integrity and anti-replay protection to LISP's EID-to-RLOC mapping data conveyed via mapping lookup process. LISP-SEC also enables verification of authorization on EID-prefix claims in Map-Reply messages. This internet-draft is under discussion in the Working Group.

In the context of the LISP WG, we contributed to an Internet-draft called "LISP Impact" [32]. The Locator/Identifier Separation Protocol (LISP) aims at improving the Internet scalability properties leveraging on three simple principles: address role separation, encapsulation, and mapping. In this internet-draft, based on implementation, deployment, and theoretical studies, we discuss the impact that deployment of LISP can have on both the Internet in general and for the end-users in particular. This internet-draft is adopted as Working Group document on December 2014.

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

6. New Results

6.1. Highlights of the Year

Valeria Vignudelli has received the AILA (Associazione Italiana di Logica e sue Applicazioni) award for her 2014 master thesis.

6.2. Service-oriented computing

Participants: Maurizio Gabbrielli, Elena Giachino, Saverio Giallorenzo, Claudio Guidi, Mario Bravetti, Ivan Lanese, Michael Lienhardt, Jacopo Mauro, Fabrizio Montesi, Gianluigi Zavattaro.

6.2.1. *Orchestrations*

Orchestration models and languages in the context of Service-Oriented Architectures (SOA) are used to describe the composition of services focusing on their interactions. Coloured Petri nets (CPN) offer a formal yet easy tool for modelling interactions in SOAs, however mapping abstract SOAs into executable ones requires a non-trivial and time-costly analysis. In [34], we propose a methodology that maps CPN-modelled SOAs into Jolie SOAs (our target language), exploiting a collection of recurring control-flow patterns, called Workflow Patterns, as composable blocks of the translation. We validate our approach with a realistic use case. In addition, we pragmatically assess the expressiveness of Jolie with respect to the considered WPs.

6.2.2. *Choreographies*

Choreographies are high-level descriptions of distributed interacting systems featuring as basic unit a communication between two participants. A main feature of choreographies is that they ensure deadlock-freedom by construction. From a choreography one can automatically derive a description of the behaviour of each participant using a notion of projection. Choreographies can be used both at the level of types (multiparty session types) or as a programming language. In [18] we surveyed the work on choreographies and behavioural contracts in multiparty interactions. In [28] we explored the notion of deadlock freedom (the system never gets stuck), and the related notions of lock freedom (each action is eventually executed under a fair scheduling) and progress (each session never gets stuck). Previous work studied how to define progress in an open setting by introducing the notion of catalysers, execution contexts generated from the type of a process. We refined the notion of catalysers leading to a novel characterization of progress in terms of the standard notion of lock-freedom. We applied our results both to binary session types and in an untyped session-based setting. We combined our results with existing techniques for lock-freedom, obtaining a new methodology for proving progress. Our methodology captures new processes w.r.t. previous progress analysis based on session types. The two following works consider the extension of choreographies, which traditionally have a static structure, to deal with adaptation, i.e., dynamic changes of the structure of choreographies. A preliminary analysis of adaptable choreographies at the level of types is presented in [27]. This work considers both updates from inside the system (self-adaptation), and external updates. Adaptable choreographies as a programming language are considered in [33], where we presented AIOCJ, a framework for programming distributed adaptive applications. AIOCJ allows the programmer to specify which parts of the application can be adapted. Adaptation takes place at run-time by means of rules, which can change during the execution to tackle possibly unforeseen adaptation needs. AIOCJ relies on a solid theory that ensures applications to be deadlock free by construction also after adaptation.

6.3. Models for reliability

Participants: Mario Bravetti, Elena Giachino, Ivan Lanese, Michael Lienhardt, Gianluigi Zavattaro.

6.3.1. *Reversibility*

We have continued the study of causal-consistent reversibility started in the past years. In [17] we presented an overview of causal-consistent reversibility, summarizing the main approaches in the literature, and the related results and applications. An interesting application is debugging. Reversible debugging provides developers with a way to execute their applications both forward and backward, seeking the cause of a misbehaviour. In a concurrent setting, reversing actions in the exact reverse order they have been executed may lead to undo many actions that were not related to the bug under analysis. On the other hand, undoing actions in some order that violates causal dependencies may lead to states that could not be reached in a forward execution. In [36] we proposed a new approach, where each action can be reversed if all its consequences have already been reversed. The main feature of the approach is that it allows the programmer to easily individuate and undo exactly the actions that caused a given misbehaviour till the corresponding bug is reached. We discussed the appropriate primitives for causal-consistent reversible debugging and presented their prototype implementation in the CaReDeb tool.

6.3.2. *Fault models*

We have continued the study of primitives for fault handling in a concurrent setting. In [19] we critically discussed the different choices that have to be made when defining a fault model for a concurrent object-oriented programming language. We consider in particular the ABS language, and analyse the interplay between the fault model and the main features of ABS, namely the cooperative concurrency model, based on asynchronous method invocations whose return results via futures, and its emphasis on static analysis based on invariants.

6.4. Cloud Computing

Participants: Roberto Amadini, Maurizio Gabbrielli, Elena Giachino, Saverio Giallorenzo, Claudio Guidi, Cosimo Laneve, Michael Lienhardt, Tudor Alexandru Lascu, Jacopo Mauro, Gianluigi Zavattaro.

6.4.1. *Cloud application deployment*

Configuration and management of applications in the cloud is a complex task that requires novel methodologies and tools. In [16] we have performed a foundational study of the complexity boundaries for the automatic deployment problem, showing that in the general case this problem is undecidable, it is decidable but non-primitive recursive if capacity constraints are not taken into account, while it turns out to be polynomial time if also conflicts between software components are not considered. Starting from these foundational observations, we have investigated the exploitability in this specific context of state-of-the-art constraint optimization techniques, a well established approach for the modeling and solution of complex optimization problems. In particular, in [23], [24] we have studied how the "portfolio technique" approach can be applied to optimization problems, combining and exploiting the performances of existing solvers to get a global, more robust and fast solver. Encouraged by these results, we have developed SUNNY-CP [13], [22]: a porftolio constraint solver for constraint satisfaction and optimization problems. SUNNY-CP has proven to have remarkable performances, ranking 4th in the annual MiniZinc challenge (i.e., the international competition to evaluate the performances of constraint solvers) and receiving a 'honorable' mention by the challenge organizers.

6.4.2. *Cloud resource management*

The management of cloud resources from client programs requires the definition of Application Programming Interfaces (APIs) that expose specific functionalities to external invokers. Programs can be built that compose existing APIs in order to obtain new functionalities. However API composition easily becomes a frustrating and time-costly task that hinders API reuse. The issue derives from technology-dependent features of API composition such as the need of extensive documentation, protocol integration, security issues, etc.. In [39] we introduce the perspective of the API-as-a-Service (APIaaS) layer as tool to ease the development and deployment of applications based on API composition, abstracting communication protocols and message formats. We elicit the desirable features of such a layer and provide a proof-of-concept prototype implemented using a service-oriented language.

Another critical aspect in this context deals with the problem of dynamic reallocation of resources. In [38] we study a type-based technique for modeling and analysis of systems in which concurrent object-oriented programs dynamically create and move resources. The type of a program is behavioural, namely it expresses the resource deployments over periods of (logical) time. Our technique admits the inference of types and may underlie the optimisation of the costs and consumption of resources.

6.5. Resource Control

Participants: Michele Alberti, Alberto Cappai, Ugo Dal Lago, Simone Martini, Giulio Pellitta, Davide Sangiorgi, Marco Solieri, Valeria Vignudelli.

6.5.1. *Probabilistic higher-order calculi*

The first results of our efforts on probabilistic higher-order systems and languages have started to appear in 2014. In particular, we have focused our attention on the impact of probability to the classical notion of context equivalence for the lambda-calculus, showing that applicative bisimilarity continues to be a congruence [31], and that it even coincides with context equivalence when evaluation is done in the call-by-value order [29]. The expressive power of higher-order concurrent contexts has been compared to the expressive power of lambda-calculi contexts and put in relation with other equivalences when the observed process is either an ordinary Labelled Transition Systems (LTS) or a reactive probabilistic transition system [25]. The obtained spectrum of equivalences for reactive probabilistic processes has been shown to be finer than the one for classic LTSs. We have also analysed the expressive power of different first-order testing equivalences (with nondeterministic tests, probabilistic tests, and both nondeterministic and probabilistic tests) in the spectrum for reactive probabilistic processes [26].

6.5.2. *Resource consumption*

The main result about resource consumption has been about an open problem on the λ-calculus: we proved that the number of leftmost-outermost steps to normal form is indeed an invariant cost model in the sense of Slot and van Emde Boas' weak invariance thesis [21]. We also introduced a new recursion theoretic framework for probabilistic computation in which one is able to capture probabilistic polynomial time through Leivant's Tiering [32].

6.5.3. *Geometry of interaction*

Novel results have been obtained for Geometry of Interaction (GoI), itself a semantics framework for linear logic introduced by Jean-Yves Girard thirty years ago. In particular, we have shown how the most concrete presentations of GoI, namely so-called token machines, can go *parallel*, thus exploiting the potential parallelism in functional programs (through the Curry-Howard Correspondence). This has been made concrete by studying extensions of multiplicative linear logic in which synchronization becomes an operator where tokens can indeed synchronize [30]. This has been later shown to be necessary to model quantum computation [44]. A simple, minimalistic GoI model of the resource λ-calculus has also been introduced [43].

6.6. Verification techniques for extensional properties

Participants: Daniel Hirschkoff, Elena Giachino, Michael Lienhardt, Cosimo Laneve, Jean-Marie Madiot, Davide Sangiorgi.

Extensional refers to properties that have to do with behavioural 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. Related to techniques for extensional properties are the issues of decidability (the problem of establishing whether certain properties are computationally feasable).

6.6.1. Coinductive techniques

Coinductive techniques, notably those based on bisimulation, are widely used in concurrency theory to reason about systems of processes. The bisimulation proof method can be enhanced by employing 'bisimulations up-to' techniques. A comprehensive theory of such enhancements has been developed for first-order (i.e., CCS-like) LTSs and bisimilarity, based on the notion of compatible function for fixed-point theory. We have transported this theory onto languages whose bisimilarity and LTS go beyond those of first-order models [40]. The approach consists in exhibiting fully abstract translations of the more sophisticated LTSs and bisimilarities onto the first-order ones. This allows us to reuse directly the large corpus of up-to techniques that are available on first-order LTSs. We have investigated the method on the π-calculus, the Higher-Order π-calculus, and a (call-by-value) λ-calculus with references.

In [20], mostly a tutorial paper, a few forms of bisimulation and of coinductive techniques that have been proposed for higher-order languages are discussed, beginning with the pure lambda-calculus and then moving to extensions of it, notably those with non-determinism and probabilities.

6.6.2. Deadlock detection

Deadlock detection in concurrent programs that create networks with an arbitrary number of nodes is extremely complex and solutions either give imprecise answers or do not scale. To enable the analysis of such programs, we have studied an algorithm for detecting deadlocks [37], [35], in a basic model featuring recursion and fresh name generation, called Lam. We then have designed a type system that associates Lams to processes. As a byproduct of these two techniques, we have an algorithm that is more powerful than previous ones and that can be easily integrated into the current release of TyPiCal, a type-based analyser for π-calculus.

6.6.3. Expressiveness and decidability in actor-like systems

Refining work in previous years, we have studied [15] the expressive power of an actor-like language, featuring concurrent objects and asynchronous message-passing. We have identified the presence/absence of fields as a crucial 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. This decidability result still holds for actors with states when the number of actors is bounded and the state is read-only.

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

6. New Results

6.1. Web programming

Participants: Yoann Couillec, Vincent Prunet, Tamara Rezk, Manuel Serrano [correspondant].

6.1.1. *Hop.js*

Multitier programming languages unify within a single formalism and a single execution environment the programming of the different tiers of distributed applications. On the Web, this programming paradigm unifies the client tier, the server tier, and, when one is used, the database tier. This homogenization offers several advantages over traditional Web programming that rely on different languages and different environments for the two or three tiers of the Web application: programmers have only one language to learn, maintenance and evolution are simplified by the use of a single formalism, global static analyses are doable as a single semantics is involved, debugging and other runtime tools are more powerful as they access global informations about the execution [17].

The three first multitier platforms for the Web all appeared in 2006: GWT (a.k.a., Google Web Toolkit), Links, and Hop [6], [5]. Each relied on a different programming model and languages. GWT maps the Java programming model on the Web, as it allows, Java/Swing likes programs to be compiled and executed on the Web; Links is functional language with experimental features such as the storing of the whole execution context on the client; Hop is based on the Scheme programming language. These three pioneers have open the path for the other multitier languages such as, Ocsigen for Ocaml, UrWeb, js-scala, etc.

In spite of their interesting properties, multitier languages have not become that popular on the Web. Today, only GWT is widely used in industrial applications but arguably GWT is not a fully multitier language as developing applications with GWT requires explicit JavaScript and HTML programming. This lack of popularity of other systems is likely due to their core based languages than to the programming model itself.

JavaScript is the *defacto* standard on the Web. Since the mid 90's, it is the language of the client-side programming and more recently, with systems like nodejs, it is also a viable solution for the server-side programming. As we are convinced by the virtues of multitier programming we have started a new project consisting of enabling multitier programming JavaScript. We have created a new language called HopScript, which is a minimalist extension of JavaScript for multitier programming, and we have implemented a brand new runtime environment called Hop.js. This environment contains a builtin Web server, on-the-fly HopScript compilers, and many runtime libraries.

HopScript is a super set of JavaScript, *i.e.,* all JavaScript programs are legal HopScript programs. Hop.js is a compliant JavaScript execution environment as it succeeds at 99% of the Ecma 262 tests suite. The Hop.js environment also aims at Node.js compatibility. In its current version it supports about 70% of the Node.js runtime environment. In particular, it fully supports the Node.js modules, which lets Hop programs reuse existing Node.js modules as is.

A prototype version of Hop.js is currently used by several academic and SME R&D teams to jointly develop an assistive robotic platform and a set of distributed applications.

We plan to release the first public Hop.js version by the end of the first semester of 2015, as we plan to start describing in forthcoming papers.

6.1.2. *Multitier Debugging*

Debugging Web applications is difficult because of their distributed nature and because the server-side and the client-side of the application are generally treated separately. The multitier approach, which reunifies the two ends of the application inside a unique execution environment, helps the debugging process because it lets the debugger access more runtime informations.

Based on our previous work on the Hop multitier debugger [17], we have built a multitier debugger for Hop.js, our multitier extension of JavaScript. Its advantage over most debuggers for the Web is that it reports the full stack trace containing all the server-side and client-side frames that have conducted to an error. Errors are reported on their actual position on the source code, wherever they occur on the server or on the client. This paper presents this debugger and sketches its implementation. This work is described in a yet unpublished paper, which will appear in 2015.

6.1.3. *Datasource*

We extended the HOP.JS language with an embedded language, inspired by PLINQ and ORC, called DATA-SOURCE. It allows programmers request multiple data sources with queries written in a unique language. We used a plinq-like language to express queries and an orc-like language to orchestrate them. Our query language and the orchastration languages can be used simultaneously or separately. We implemented bindings between DATASOURCE and some representative types of data sets such as SPARQL endpoints, relationnal databases, WEB services, and WEB pages. We are extending HOP.JS by supporting EcmaScript 6 array comprehensions in order to write a unique query over multiple data sources in a unified formalism. The query is then compiled into database specific queries. We linked all the bindings made for HOP with HOP.JS. We implemented another binding for a document oriented data base, MONGODB.

6.2. Distributed programming

Participant: Bernard Serpette [correspondant].

6.2.1. *Logical behavioural semantics of Esterel*

We have formalised, with the Coq system, the logical behavioural semantics of Esterel as described in Gérard Berry's book. In order to define the properties of reactivity and determinism, we have defined a new semantics using contexts with a proven correspondence between the two semantics.

The specification and the proofs of the correspondence take 3500 lines of Coq.

6.2.2. *Abstract distributed machine*

We have experimented an abstract machine composed of distributed nodes. Each node has exactly two named links to other nodes and an instruction able to modify one link of a reachable node. This instruction is executed when a token is received, once the instruction is achieved the token is transmitted to another reachable node.

This abstract machine is turing complete. The λ-calculus and the π-calculus can be compiled to the instruction set of this machine.

The execution of one individual node may involve paths of arbitrary length, for example, when compiling the λ-calculus or the π-calculus, the path length for accessing a variable is proportional to it's de Bruijn index and therefore is not bounded. Given a machine with instructions of unbounded paths, we can build an equivalent machine where all the paths are bounded by two: a node is only able to access it's own links and the links of its neighbour. Moreover, this transformation uses only 6 different instructions.

6.3. Security and Privacy

Participants: Ilaria Castellani, José Fragoso Santos, Nataliia Bielova, Tamara Rezk [correspondant].

6.3.1. *Security of Dynamically Evolving Systems of Communicating Processes*

We have started to address security issues in the context of dynamically evolving systems of communicating processes, which are able to adapt themselves in reaction to particular events (for instance, security attacks or changes in security policies). We present initial results on a simple model of processes communicating via structured interactions (sessions), in which self-adaptation and security concerns are jointly addressed. In this model, security violations occur when processes attempt to read or write messages of inappropriate security level within a structured interaction. Such violations trigger adaptation mechanisms that prevent the violations to occur and/or to propagate their effect in the choreography. Our model is equipped with local and global mechanisms for reacting to security violations; type soundness results ensure that the global protocols are still correctly executed while the system adapts itself to preserve its security.

6.3.2. Browser Randomisation against Fingerprinting: a Quantitative Information Flow Approach

Web tracking companies use device fingerprinting to distinguish the users of the websites by checking the numerous properties of their machines and web browsers. One way to protect the users' privacy is to make them switch between different machine and browser configurations. We propose a formalisation of this privacy enforcement mechanism.

We use information-theoretic channels to model the knowledge of the tracker and the fingerprinting program, and show how to synthesise a randomisation mechanism that defines the distribution of configurations for each user. This mechanism provides a strong guarantee of *privacy* (the probability of identifying the user is bounded by a given threshold) while maximising *usability* (the user switches to other configurations rarely). To find an optimal solution, we express the enforcement problem of randomisation by a linear program. We investigate and compare several approaches to randomisation and find that more efficient privacy enforcement would often provide lower usability. Finally, we relax the requirement of knowing the fingerprinting program in advance, by proposing a randomisation mechanism that guarantees privacy for an arbitrary program.

This work has been published and presented at the Nordic Conference on Secure IT Systems (NordSec 2014) [12]. The extended version of the paper has been published as a technical report [20].

6.3.3. Crying Wolf? On the Price Discrimination of Online Airline Tickets

Price discrimination refers to the practice of dynamically varying the prices of goods based on a customer's purchasing power and willingness to pay. Motivated by several anecdotal accounts, we report on a three week experiment, conducted in search of price discrimination in airline tickets. Despite presenting the companies with multiple opportunities for discriminating us, and contrary to our expectations, we did not find any evidence for systematic price discrimination. At the same time, we witnessed the highly volatile prices of certain airlines which make it hard to establish cause and effect. Finally, we provided alternative explanations for the observed price differences.

This work has been published and presented at the Workshop on Hot Topics in Privacy Enhancing Technologies (HotPETs 2014) [19].

6.3.4. Stateful Declassification Policies for Event-Driven Programs

We propose a novel mechanism for enforcing information flow policies with support for declassification on event-driven programs. Declassification policies consist of two functions. First, a projection function specifies for each confidential event what information in the event can be declassified directly. This generalizes the traditional security labelling of inputs. Second, a stateful release function specifies the aggregate information about all confidential events seen so far that can be declassified. We provide evidence that such declassification policies are useful in the context of JavaScript web applications. An enforcement mechanism for our policies is presented and its soundness and precision is proven. Finally, we give evidence of practicality by implementing and evaluating the mechanism in a browser. This work has been published at Computer Security Foundations (CSF'14) [18].

6.3.5. An Information Flow Monitor for a Core of DOM

We propose and prove sound a novel, purely dynamic, flow sensitive monitor for securing information flow in an imperative language extended with DOM-like tree operations, that we call Core DOM. In Core DOM, as in the DOM API, tree nodes are treated as first-class values. We take advantage of this feature in order to implement an information flow control mechanism that is finer-grained than previous approaches in the literature. Furthermore, we extend Core DOM with additional constructs to model the behavior of live collections in the DOM Core Level 1 API. We show that this kind of construct effectively augments the observational power of an attacker and we modify the proposed monitor so as to tackle newly introduced forms of information leaks. This work has been published at the 9th International Symposium on Trustworthy Global Computing (TGC) [11].

6.3.6. An Information Flow Monitor-Inlining Compiler for Securing a Core of JavaScript

Web application designers and users alike are interested in isolation properties for trusted JavaScript code in order to prevent confidential resources from being leaked to untrusted parties. Noninterference provides the mathematical foundation for reasoning precisely about the information flows that take place during the execution of a program. Due to the dynamicity of the language, research on mechanisms for enforcing noninterference in JavaScript has mostly focused on dynamic approaches. We present the first information flow monitor inlining compiler for a realistic core of JavaScript. We prove that the proposed compiler enforces termination-insensitive noninterference and we provide an implementation that illustrates its applicability.

This work has been published at the 29th IFIP International Information Security and Privacy Conference (IFIP SEC) [14].

6.3.7. From Static to Hybrid Typing Secure Information Flow in a Core of JavaScript

We propose a novel type system for securing information flow in a core of JavaScript. This core takes into account the defining features of the language, such as prototypical inheritance, extensible objects, and constructs that check the existence of object properties. We design a hybrid version of the proposed type system. This version infers a set of assertions under which a program can be securely accepted and instruments it so as to dynamically check whether these assertions hold. By deferring rejection to runtime, the hybrid version can typecheck secure programs that purely static type systems cannot accept.

<div align="center">

MAESTRO Project-Team

</div>

6. New Results

6.1. Highlights of the Year

E. Altman has received the "Isaacs' Award" granted by the International Society on Dynamic Games in recognition for his research on dynamic game theory.

M. El Chamie got the Best Session Presentation Award at the IEEE American Control Conference ACC 2014 for the paper "Newton's method for constrained norm minimization and its application to weighted graph problems," co-authored with G. Neglia.

THANES is a new French-Brazilian joint-team between MAESTRO and researchers from Univ. Federal do Rio de Janeiro (Brazil) and Carneggie Mellon Univ. (USA). The team investigates network science problems with a particular focus on Online Social Networks.

BEST PAPERS AWARDS :

[43] **6th IEEE INFOCOM International Workshop on Network Science for Communication Networks (NetSciCom)**. K. AVRACHENKOV, P. BASU, G. NEGLIA, B. RIBEIRO, D. TOWSLEY.

[70] **4th IEEE Online Conference on Green Communications (GreenComm)**. C. ROTTONDI, G. NEGLIA, G. VERTICALE.

6.2. Network Science

Participants: Eitan Altman, Konstantin Avrachenkov, Mahmoud El Chamie, Julien Gaillard, Arun Kadavankandy, Jithin Kazhuthuveettil Sreedharan, Hlib Mykhailenko, Philippe Nain, Giovanni Neglia, Yonathan Portilla, Alexandre Reiffers, Vikas Singh, Marina Sokol.

6.2.1. *Epidemic models of propagation of content*

Epidemic models have received significant attention in the past few decades to study the propagation of viruses, worms and ideas in computer and social networks. In the case of viruses, the goal is to understand how the topology of the network and the properties of its nodes impact the spread of the epidemics. In [38], E. Altman, A. Avritzer and L. Pfleger de Aguiar (Siemens Corporation, Princeton, USA), R. El-Azouzi (Univ. of Avignon), and D. S. Menasche (Federal Univ. of Rio de Janeiro, Brazil) propose rejuvenation as a way to cope with epidemics. Reformatting a computer may solve the problem of virus contamination (but it might be a costly operation) while less dramatic actions may render the computer operational again (even in the presence of the virus). In this work they evaluate the performance gain of such measures as well as sampling for early detection of viruses while these incubate. During incubation, contaminated terminals are infectious and yet, if not detected to be so, they cannot be isolated and treated.

In [60], Y. Hayel (Univ. of Avignon), S. Trajanovski and P. Van Mieghem (Delft Univ. of Technology, The Netherlands), E. Altman, and H. Wang (Delft Institute of Applied Mathematics, The Netherlands), compare solutions involving vaccination to those that involve healing from a selfish point of view of an individual networked user. A game theoretical model is presented and the obtained equilibrium is computed for various types of topologies including the fully connected one, the bipartite graph and a community structure. A novel use of potential games is presented to compute the equilibria.

In [61], L. Maggi and F. De Pellegrini (CREATE-NET, Italy), A. Reiffers, J. J. Herings (Maastricht Univ., The Netherlands) and E. Altman, study a viral diffusion of a content in a multi-community environment. Exploiting time scale separation, the authors are able to reduce the dimensionality of the problem and to compute its limiting behavior in closed form. They further study regulation and cooperative approaches for sharing the cost for fighting the spread of the infection among the communities.

Social networks can have asymmetric relationships. In the online social network Twitter, a follower receives tweets from a followed person but the followed person is not obliged to subscribe to the channel of the follower. Thus, it is natural to consider the dissemination of information in directed networks. In [44], K. Avrachenkov in collaboration with B. Prabhu (LAAS-CNRS), K. De Turck and D. Fiems (Ghent Univ., Belgium) use the mean-field approach to derive differential equations that describe the dissemination of information in a social network with asymmetric relationships. In particular, their model reflects the impact of the degree distribution on the information propagation process. They further show that for an important subclass of their model, the differential equations can be solved analytically.

6.2.2. *Bio-Inspired Models for Characterizing YouTube Viewcount*

Bio-inspired models have long been advocated for the dissemination of content in the Internet. How good are such models and how representative are they? In [69], C. Richier, R. El-Azouzi, T. Jimenez, G. Linares (all with Univ. of Avignon), E. Altman and Y. Portilla propose six different epidemic models. These are classified according to various criteria: (i) the size of the target population, which may be constant, or linearly increasing or infinite, (ii) the virality of the content: it is said to be viral if nodes that receive the content participate in retransmitting it (by sharing or embedding). They then collected data on the viewcounts of videos in youtube and examined how well they fit their models. They showed that their six models cover 90% of the videos with an average mean square error of less than 5%. They further studied the capability of using these models to predict the evolution of the viewcount.

6.2.3. *Network centrality measures*

Finding quickly top-k lists of nodes with the largest degrees in large complex networks is a basic problem of recommendation systems. If the adjacency list of the network is known (not often the case in complex networks), a deterministic algorithm to solve this problem requires an average complexity of $O(n)$, where n is the number of nodes in the network. Even this modest complexity can be excessive for large complex networks. In [18], K. Avrachenkov and M. Sokol in collaboration with N. Litvak (Twente Univ., The Netherlands) and D. Towsley (Univ. of Massachusetts, Amherst, USA) propose to use a random-walk-based method. They 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 that requires very little knowledge about the structure of the network.

In [46], K. Avrachenkov in collaboration with N. Litvak (Twente Univ., the Netherlands) and L. Ostroumova and E. Suyargulova (both from Yandex, Russia) address the problem of quick detection of high-degree entities in large online social networks. The practical importance of this problem is attested by a large number of companies that continuously collect and update statistics about popular entities, usually using the degree of an entity as an approximation of its popularity. They suggest a simple, efficient, and easy to implement two-stage randomized algorithm that provides highly accurate solutions for this problem. For instance, their algorithm needs only one thousand API requests in order to find the top-100 most followed users in Twitter, a network with approximately a billion of registered users, with more than 90% precision. Their algorithm significantly outperforms existing methods and serves many different purposes, such as finding the most popular users or the most popular interest groups in social networks. They show that the complexity of the algorithm is sublinear in the network size, and that high efficiency is achieved in networks with high variability among the entities, expressed through heavy-tailed distributions.

Personalized PageRank is an algorithm to classify the importance of web pages on a user-dependent basis. In [48], K. Avrachenkov and M. Sokol in collaboration with R. van der Hofstad (EURANDOM, The Netherlands) introduce two generalizations of Personalized PageRank with node-dependent restart. The first generalization is based on the proportion of visits to nodes before the restart, whereas the second generalization is based on the proportion of time a node is visited just before the restart. In the original case of constant restart probability, the two measures coincide. They discuss interesting particular cases of restart probabilities and restart distributions. They show that both generalizations of Personalized PageRank have an elegant expression

connecting the so-called direct and reverse Personalized PageRanks that yield a symmetry property of these Personalized PageRanks.

Along with K. Avrachenkov and N. M. Markovich (Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia), J. K. Sreedharan investigated distribution and dependence of extremes in network sampling processes [47]. This is one of the first studies associating extremal value theory to sampling of large networks. The work showed that for any general stationary samples from the graph (function of node samples) meeting two mixing conditions, the knowledge of bivariate distribution or bivariate copula is sufficient to derive many of its extremal properties. The work proved the usage of a single parameter to find many relevant extremes in networks like order statistics, first hitting time, mean cluster size etc. In particular, correlation in degrees of adjacent nodes are modelled and different random walks, such as PageRank, are studied in detail. This work has been done in the context of Inria Alcatel-Lucent Bell Labs joint laboratory's ADR "Network Science" (see §7.1.2).

6.2.4. *Influence maximization in complex networks*

Efficient marketing or awareness-raising campaigns seek to recruit a small number, w, of influential individuals—where w is the campaign budget—that are able to cover the largest possible target audience through their social connections. In [43] K. Avrachenkov and G. Neglia in collaboration with P. Basu (BBN Technologies, US), B. Ribeiro (CMU, US) and D. Towsley (Univ. of Massachusetts, Amherst, USA) assume that the topology is gradually discovered thanks to recruited individuals disclosing their social connections. They analyze the performance of a variety of online myopic algorithms (i.e. that do not have a priori information on the topology) currently used to sample and search large networks. They also propose a new greedy online algorithm, Maximum Expected Uncovered Degree (MEUD). Their proposed algorithm greedily maximizes the expected size of the cover, but it requires the degree distribution to be known. For a class of random power law networks they show that MEUD simplifies into a straightforward procedure, denoted as MOD because it requires only the knowledge of the Maximum Observed Degree. This work has been done in the context of THANES Joint team (see §8.3.1.1) and Inria Alcatel-Lucent Bell Labs joint laboratory's ADR "Network Science" (see §7.1.2).

In [66] G. Neglia, in collaboration with X. Ye (Politecnico di Torino, Italy), M. Gabielkov and A. Legout (from the DIANA team) consider how to maximize users influence in Online Social Networks (OSNs) by exploiting social relationships only. Their first contribution is to extend to OSNs the model of Kempe, Kleinberg and Tardös on the propagation of information in a social network and to show that a greedy algorithm is a good approximation of the optimal algorithm that is NP- hard. However, the greedy algorithm requires global knowledge, which is hardly practical. Their second contribution is to show on simulations on the full Twitter social graph that simple and practical strategies perform close to the greedy algorithm.

6.2.5. *Clustering*

Clustering of a graph is the task of grouping its nodes in such a way that the nodes within the same cluster are well connected, but they are less connected to nodes in different clusters. In [45] K. Avrachenkov, M. El Chamie and G. Neglia propose a clustering metric based on the random walks' properties to evaluate the quality of a graph clustering. They also propose a randomized algorithm that identifies a locally optimal clustering of the graph according to the metric defined. The algorithm is intrinsically distributed and asynchronous. If the graph represents an actual network where nodes have computing capabilities, each node can determine its own cluster relying only on local communications. They show that the size of clusters can be adapted to the available processing capabilities to reduce the algorithm's complexity.

6.2.6. *Average consensus protocols*

In [54], [82], M. El Chamie in collaboration with J. Liu and T. Başar (Univ. of Illinois at Urbana Champaign, USA) studies the performance of a subclass of distributed averaging algorithms where the information exchanged between neighboring nodes (agents) is subject to deterministic uniform quantization. They give the convergence properties of linear averaging due to such quantization (which is a practical concern for many applications) that cause nonlinearity in the system. This is the first attempt to solve the exact model.

In [53], M. El Chamie in collaboration with T. Başar (Univ. of Illinois at Urbana Champaign, USA) considers optimal design strategies in consensus protocols for networks vulnerable to adversarial attacks. They provide a game theoretical model for the problem of a network with an adversary corrupting the control signal with noise. They derive the optimal strategies for both players (the adversary and the network designer) of the resulting game using a saddle point equilibrium solution in mixed strategies.

6.3. Wireless Networks

Participants: Eitan Altman, Abdulhalim Dandoush, Majed Haddad, Jithin Kazhuthuveettil Sreedharan.

6.3.1. *Localization in ad-hoc wireless sensors networks*

Range-based localization algorithms in wireless sensor networks are more accurate but also more computationally complex than the range-free algorithms. In collaboration with M. S. Elgamel (Univ. of Louisiana, USA), A. Dandoush has revised the Trigonometric based Ad-hoc Localization System (TALS) proposed in the literature. In [83], they propose a new technique to optimize the system: by eliminating the need of solving a linear system of equations via least square methods or its variants or the need for any square root operations, the computational overhead is reduced. Also, a novel modified Manhattan distance is proposed and used in the elimination process ensuring thereby a very good accuracy with less complexity than the basic TALS. Through a mathematical analysis and intensive simulations, the optimized TALS is shown to present superior performance and accuracy results compared to other localization techniques.

6.3.2. *Channel management*

The enhanced Inter Cell Interference Coordination (eICIC) feature has been introduced to solve the interference problem in small cells. It involves two parameters which need to be optimized, namely the Cell Range Extension (CRE) of the small cells and the ABS ratio (ABSr) which defines a mute ratio for the macro cell to reduce the interference it produces. In [72], A. Tall, Z. Altman (Orange Labs, Issy les Moulineaux) and E. Altman propose self-optimizing algorithms for the eICIC. The CRE is adjusted by means of a load balancing algorithm. The ABSr parameter is optimized by maximizing a proportional fair utility of user throughputs. The convergence of the algorithms is proven using Stochastic Approximation theorems. Numerical simulations illustrate the important performance gain brought about by the different algorithms.

Cognitive Radios are proposed as a solution to scarcity of wireless spectrum and one of the main challenges here is to gain knowledge about the spectrum usage by the licensed users, termed as spectrum sensing. In [29], Vinod Sharma (Indian Institute of Science, Bangalore, India) and J. K. Sreedharan study novel algorithms for spectrum sensing which minimize the expected time for spectrum sensing with stringent constraints on the probability of wrong detection. Algorithms are distributed in nature and the work proves that the algorithms are asymptotically optimal distributed sequential hypothesis tests. Along with theoretical guarantees, many practical scenarios in Cognitive Radios are also investigated.

6.3.3. *Self-Organizing Network (SON)*

The fast development of SON technology in mobile networks renders critical the problem of coordinating SON functionalities operating simultaneously. SON functionalities can be viewed as control loops that may need to be coordinated to guarantee conflict free operation, to enforce stability of the network and to achieve performance gain. In [30], A. Tall and Z. Altman (Orange Labs, Issy les Moulineaux), R. Combes (SUPELEC), and E. Altman propose a distributed solution for coordinating SON functionalities. It uses Rosen's concave games framework in conjunction with convex optimization. The SON functionalities are modeled as linear Ordinary Differential Equation (ODE)s. The stability of the system is first evaluated using a basic control theory approach together with strict diagonal concavity notion that originates from game theory. The coordination solution consists in finding a linear map (called coordination matrix) that stabilizes the system of SON functionalities. It is proven that the solution remains valid in a noisy environment using Stochastic Approximation.

6.4. Network Engineering Games

Participants: Eitan Altman, Ilaria Brunetti, Majed Haddad, Alexandre Reiffers.

6.4.1. *The association problem*

In [57], M. Haddad, S. Habib (Orange Labs, Issy les Moulineaux), and P. Wiecek (Wroclaw Univ. of Technology, Poland) and E. Altman develop a hierarchical Bayesian game framework for automated dynamic offset selection. Users compete to maximize their throughput by picking the best locally serving radio access network (RAN) with respect to their own measurement, their demand and a partial statistical channel state information of other users. In particular, they investigate the properties of a Stackelberg game, in which the base station is a player on its own. They derive analytically the utilities related to the channel quality perceived by users to obtain the equilibria. They study the Price of Anarchy of such system, which is defined as the ratio of the social welfare attained when a network planner chooses policies to maximize social welfare versus the social welfare attained at a Nash/Stackeleberg equilibrium when users choose their policies strategically.

6.4.2. *Cognitive radio*

In [26], M. Haddad, P. Wiecek (Wroclaw Univ. of Technology, Poland), O. Habachi and Y. Hayel (both with Univ. of Avignon) propose a game theoretical approach that allows cognitive radio pairs, namely the primary user (PU) and the secondary user (SU), to update their transmission powers and frequencies simultaneously. Specifically, a Stackelberg game model in which individual users attempt to hierarchically access to the wireless spectrum while maximizing their energy efficiency was addressed. A thorough analysis of the existence, uniqueness and characterization of the Stackelberg equilibrium was conducted. In particular, it was shown that a spectrum coordination naturally occurs when both actors in the system decide sequentially about their powers and their transmitting carriers. As a result, spectrum sensing in such a situation turns out to be a simple detection of the presence/absence of a transmission on each sub-band. An algorithmic analysis on how the PU and the SU can reach such a spectrum coordination using an appropriate learning process is provided.

In [59], the same authors present a hierarchical game to model distributed joint power and channel allocation for multi-carrier energy efficient cognitive radio systems. A thorough analysis of the existence, uniqueness and characterization of the Stackelberg equilibrium is conducted. It was proved that, at the Stackelberg equilibrium, each of the two users transmits on only one carrier depending on the fading channel gains. This results contrast with capacity-based approaches in which a certain number of carriers is exploited depending on the channel gains. Interestingly, it was shown that, for the vast majority of cases, introducing a certain degree of hierarchy in a multi-carrier system induces a natural coordination pattern where users have incentive to choose their transmitting carriers in such a way that they always transmit on orthogonal channels. Analytical results were provided for assessing and improving the performances in terms of energy efficiency between the non-cooperative game with synchronous decision makers and the proposed Stackelberg game.

6.4.3. *Routing Games*

In [39], E. Altman, J. Kuri (Indian Institute of Science, Bangalore, India) and R. El-Azouzi (Univ. of Avignon) study a routing game that models competition over a simple network with losses. Packets may be lost in the network due to either congestion losses or to channel random losses. They compute the equilibrium and establish its properties. They identify a Braess type paradox in which by adding a link the loss probabilities of all players increase.

G. Accongiagioco (Institute for Advanced Studies, Lucca, Italy), E. Altman, E. Gregori (Italian National Research Council, Italy) and L. Lenzini (Univ. of Pisa, Italy) analyze in [36] the decisions taken by an Autonomous System (AS) when joining the Internet. They first define a realistic model for the interconnection costs incurred and then they use this cost model to perform a game theoretic analysis of the decisions related to the creation of new links in the Internet. The proposed model does not fall into the standard category of routing games, hence they devise new tools to solve it by exploiting peculiar properties of the game. They prove analytically the existence of multiple equilibria for specific cases, and provide an algorithm to compute the stable ones. The analysis of the model's outcome highlights the existence of a Price of Anarchy and a Price of Stability.

6.4.4. Network neutrality and collusion

Representatives of several Internet access providers have expressed their wish to see a substantial change in the pricing policies of the Internet. In particular, they would like to see content providers pay for use of the network, given the large amount of resources they use. This would be in clear violation of the "network neutrality" principle that had characterized the development of the wireline Internet. In [14], E. Altman, M. K. Hanawal (former PhD student in MAESTRO) and R. Sundaresan (Indian Institute of Science, Bangalore, India) proposed and studied possible ways of implementing such payments and of regulating their amount. The results were reported already in a previous report, but were substantially revised during the period of this project.

6.4.5. Competition over popularity in social networks

We have pursued our analysis of competition over popularity and visibility in social networks. In [68], A. Reiffers and E. Altman, together with Y. Hayel (Univ. of Avignon) study a game model that arises when the rate of transmission of packets of each source can be accelerated in order to optimize a weighted sum of its acceleration cost and the expected number of its contents on the timelines of those who follow that content. While this paper considers equilibrium within static policies (in which the acceleration rate does not change in time), the same authors study in [51] the structure of dynamic equilibrium policies which are allowed to change as a function of the time (or of the state). A problem with a similar tradeoff is studied by E. Altman in a mobile context in [13] where the question of accelerating the transmission rate of content arises in a context of competition over content where it is assumed that if a content reaches a given destination then that destination will not be interested any more in receiving competing content.

In [67], A. Reiffers, E. Altman and Y. Hayel (Univ. of Avignon) extend the work in [68], and model the situation in which several social networks are available and a source may control not only the rate of transmission (acceleration) but may also decide how to split its content to the various social networks.

A competition over the timing of the transmission of a content was studied by E. Altman and N. Shimkin (Israel Institute of Technology, Israel) in [41]. Uniqueness of a symmetric equilibrium was established under the assumption of Poisson arrival of requests.

6.5. Green Networking and Smart Grids

Participants: Sara Alouf, Eitan Altman, Alberto Benegiamo, Ioannis Dimitriou, Majed Haddad, Alain Jean-Marie, Giovanni Neglia.

6.5.1. Energy efficiency in wireless networks

In [25], M. Haddad, P. Wiecek (Wroclaw Univ. of Technology, Poland), O. Habachi and Y. Hayel (both with Univ. of Avignon) investigated the achievable performances of multi-carrier energy efficient power control game. Both the simultaneous-move and the hierarchical games were addressed. For the first time, the analytical closed-form expressions of the spectrum coordination and the spectral efficiency of such models was derived. Results indicate that the spectrum coordination capability induced by the power control game model enables the wireless network to enjoy the energy efficiency improvement while still achieving a high spectral efficiency.

In [58], the same authors studied energy efficiency of heterogeneous networks for both sparse and dense (two-tier and multi-tier) small cell deployments. The problem is formulated as a hierarchical (Stackelberg) game in which the macro cell is the leader whereas the small cell is the follower. Both players want to strategically decide on their power allocation policies in order to maximize the energy efficiency of their registered users. A backward induction method has been used to obtain a closed-form expression of the Stackelberg equilibrium. It was shown that the energy efficiency is maximized when only one sub-band is exploited for the players of the game depending on their fading channel gains.

In [34], R. A. Vaca Ramirez and J. S. Thompson (Univ. of Edinburgh, UK), E. Altman and V. M. Ramos Ramos (Univ. Autonoma Metropolitana, Mexico) aim to reduce the power expenditure in the reverse link during low network load periods, by allocating extra resource blocks (RBs) to the mobile users. This is in contrast with other approaches in which resources are reduced in hours of low energy consumption. The user's rate demands are split among its allocated RBs in order to transmit in each of them by using a low level modulation order. In this low SINR regime the transmission is much more energy efficient since the log appearing in Shannon formula is in close to linear. We model the bandwidth expansion (BE) process by a game theory framework derived from the concept of stable marriage with incomplete lists (SMI).

P. Wiecek (Wroclaw Univ. of Technology, Poland) and E. Altman consider in [42] dynamic Multiple Access games between a random number of players competing over collision channels. Each of several mobiles involved in an interaction determines whether to transmit at a high or at a low power. High power decreases the lifetime of the battery but results in smaller collision probability. They formulated this game as an anonymous sequential game with undiscounted reward and computed the equilibrium [42]. The internal state of a player corresponds to the amount of energy left in the battery and the actions correspond to the transmission power.

I. Dimitriou investigated in [52] the power management of mobile devices, using a variant of an M/G/1 queue with probabilistic inhomogeneous multiple vacations and generalized service process. Under the vacation scheme, at the end of a vacation the server goes on another vacation, with a different probability distribution, if during the previous vacation there have been no arrivals. The modified vacation policy depends on the initial vacation interval and the server selects randomly over M such vacation policies. The theoretical system can be applied for modeling the power saving mode of mobile devices in modern wireless systems. Moreover, the form of the service process properly describes the incremental redundancy retransmission scheme that provides different types of retransmissions in such systems. Steady state analysis is investigated, energy and performance metrics are obtained and used to provide numerical results that are also validated against simulations.

6.5.2. *Energy efficiency in delay tolerant networks*

Energy efficiency in mobile networks is further studied in [28] where L. Sassatelli (Univ. of Nice Sophia Antipolis), A. Ali, M. Panda and T. Chahed (all with Telecom SudParis) and E. Altman tackle the issue of reliable transport in Delay-Tolerant mobile ad hoc Networks, that are operated by some opportunistic routing algorithm. We propose a reliable transport mechanism that relies on Acknowledgements (ACK) and coding at the source. The various versions of the problem depending on buffer management policies are formulated, and a fluid model based on a mean-field approximation is derived for the designed reliable transport mechanism. This model allows to express both the mean file completion time and the energy consumption up to the delivery of the last ACK at the source.

6.5.3. *Modeling of a smart green base station*

S. Alouf, I. Dimitriou A. Jean-Marie have considered the modeling of wireless communication base stations with autonomous energy supply (solar, wind). They proposed and analyzed a queueing model to assess performance of a base station fully powered by renewable energy sources. The system operates in a finite state Markovian random environment that properly describes the intermittent nature of renewable energy sources and the data traffic. The base station is considered to be "smart" in the sense that it is able to dynamically adjust its coverage area, controlling thereby the traffic rate and its energy consumption. They show how the matrix-analytic formalism enables to construct and study the performance of a smart green base station operating in random environment. More precisely, the behavior of such a system is described by a five-dimensional Markov process, which is a homogeneous finite Quasi Birth-Death (QBD) process. Several existing algorithms can be used in order to obtain the stationary probability vector, which is the basis for the calculation of interesting performance metrics. This work is on-going and has not been submitted for publication yet.

6.5.4. *Direct Load Control*

Balancing energy demand and production is becoming a more and more challenging task for energy utilities also because of the larger penetration of renewable energies which are more difficult to predict and control.

While the traditional solution is to dynamically adapt energy production to follow the time-varying demand, a new trend is to drive the demand itself. Most of the ongoing actions in this direction involve greedy energy consumers, like industrial plant, supermarkets or large buildings. Pervasive communication technologies may allow in the near future to push further the granularity of such approach, by having the energy utility interacting with residential appliances. In [65] and in its extension [64], G. Neglia, in collaboration with G. Di Bella, L. Giarré and I. Tinnirello (Univ. of Palermo, Italy) study large scale direct control of inelastic home appliances whose energy demand cannot be shaped, but simply deferred. Their solution does not suppose any particular intelligence at the appliances. The actuators are rather smart plugs (simple devices with local communication capabilities that can be inserted between appliances plugs and power sockets) and are able to interrupt/reactivate power flow through the plug. A simple control message can be broadcast to a large set of smart plugs for probabilistically enabling or deferring the activation requests of a specific load type in order to satisfy a probabilistic bound on the aggregated power consumption. The control law and the most important performance metrics can be easily derived analytically.

6.5.5. *Charge of Electric Vehicles*

The massive introduction of Electric Vehicles (EVs) is expected to significantly increase the power load experienced by the electrical grid, but also to foster the exploitation of renewable energy sources: if the charge process of a fleet of EVs is scheduled by an intelligent entity such as a load aggregator, the EVs' batteries can contribute in flattening energy production peaks due to the intermittent production patterns of renewables by being recharged when energy production surpluses occur. To this aim, time varying energy prices are used, which can be diminished in case of excessive energy production to incentivize energy consumption (or increased in case of shortage to discourage energy utilization). In [70] G. Neglia, in cooperation with C. Rottondi and G. Verticale (Politecnico di Milano, Italy), evaluate the complexity of the optimal scheduling problem for a fleet of EVs aimed at minimizing the overall cost of the battery recharge in presence of time-variable energy tariffs. The scenario under consideration is a fleet owner having full knowledge of customers' traveling needs at the beginning of the scheduling horizon. They prove that the problem has polynomial complexity, provide complexity lower and upper bounds, and compare its performance to a benchmark approach which does not rely on prior knowledge of customers' requests, in order to evaluate whether the additional complexity required by the optimal scheduling strategy w.r.t. the benchmark is worthy the achieved economic advantages. Numerical results show considerable cost savings obtained by the optimal scheduling strategy.

6.6. Content-Oriented Systems

Participants: Sara Alouf, Eitan Altman, Konstantin Avrachenkov, Nicaise Choungmo Fofack, Abdulhalim Dandoush, Majed Haddad, Alain Jean-Marie, Philippe Nain, Giovanni Neglia, Marina Sokol.

6.6.1. *Modeling modern DNS caches*

N. Choungmo Fofack and S. Alouf have pursued their study of the modern behavior of DNS (Domain Name System) caches. The entire set of traces collected in 2013 by Inria's IT services in Sophia Antipolis at one of the Inria's DNS caches have been processed and analyzed with the help of N. Nedkov (4-month intern in MAESTRO). This allowed to strengthen the validation of the theoretical models developed in 2013 (see [86]). On the other hand, parts of [86] have been revisited and derived under more general assumptions. As a direct consequence, the exact analysis derived on linear cache networks is extended to a large class of hierarchical cache networks called *linear-star* networks which include linear and two-level tree/star networks. In addition, closed-form expressions for the cache consistency measures (refresh rate and correctness probability) are provided under the assumption that contents requests and updates occur according to two independent renewal processes.

6.6.2. *Analysis of general and heterogeneous cache networks*

There has been considerable research on the performance analysis of *on-demand* caching replacement policies like Least-Recently-Used (LRU), First-In-First-Out (FIFO) or Random (RND). Much progress has been made

on the analysis of a single cache running these algorithms. However it has been almost impossible to extend the results to networks of caches. In [22], N. Choungmo Fofack, P. Nain and G. Neglia, in collaboration with D. Towsley (Univ. of Massachusetts, Amherst, USA), introduce a Time-To-Live (TTL) based caching model, that assigns a timer to each content stored in the cache and redraws it every time the content is requested (at each hit/miss). They derive the performance metrics (hit/miss ratio and rate, occupancy) of a TTL-based cache in isolation fed by stationary and ergodic request processes with general TTL distributions. Moreover they propose an iterative procedure to analyze TTL-based cache networks under the assumptions that requests are described by *renewal processes* (that generalize Poisson processes or the standard IRM assumption). They validate the theoretical findings through event-driven and Monte-Carlo simulations based on the Fourier Amplitude Sensitivity Test to explore the space of the input parameters. The analytic model predicts remarkably well all metrics of interest with relative errors smaller than 1%.

Jointly with M. Badov, M. Dehghan, D. L. Goeckel and D. Towsley (all with the Univ. of Massachusetts, Amherst, USA), N. Choungmo Fofack proposes in [81] approximate models to assess the performance of a cache network with arbitrary topology where nodes run the Least Recently Used (LRU), First-In First-Out (FIFO), or Random (RND) replacement policies on arbitrary size caches. The authors take advantage of the notions of "cache characteristic time" and "Time-To-Live (TTL)-based cache" to develop a unified framework for approximating metrics of interest of interconnected caches. This approach is validated through event-driven simulations, and when possible, compared to the existing *a-NET* model.

6.6.3. *Data placement and retrieval in distributed/peer-to-peer systems*

Distributed systems using a network of peers have become an alternative solution for storing data. These systems are based on three pillars: data fragmentation and dissemination among the peers, redundancy mechanisms to cope with peers churn and repair mechanisms to recover lost or temporarily unavailable data. In previous years, A. Dandoush, S. Alouf and P. Nain have studied the performance of peer-to-peer storage systems in terms of data lifetime and availability using the traditional redundancy schemes. This work has now been published in [23].

A. Jean-Marie and O. Morad (Univ. Montpellier 2) have proposed a control-theoretic model for the optimization of prefetching in the context of hypervideo or, more generally, connected documents. The user is assumed to move randomly from document to document, and the controller attempts at downloading in advance the documents accessed. A penalty is incurred when the document is not completely present. The model is flexible in the sense that it allows several variants for the network model and the cost metric [63]. They have proposed exact algorithms and heuristics for the solution of this problem, and compared them on a benchmark of different user behaviors [62].

The question of whether it is possible to prefetch documents so that the user never experiences blocking, has been modeled with a "cops-and-robbers" game jointly with F. Fomin (Univ. Bergen), F. Giroire and N. Nisse (both from Inria project-team COATI) and D. Mazauric (former PhD student in MAESTRO and MASCOTTE) [24] (see also MAESTRO's 2011 activity report).

6.6.4. *Streaming optimization*

In streaming applications such as youtube, packets have to be played at the destination at the same rate they were created. If a packet is not available at the destination when it has to be played then a starvation occurs. This results in an unpleasant frozen screen and in an interruption in the video. To decrease the probability of a starvation the destination first waits till it has received some target number of packets and only then starts to play them. In [32], E. Altman and M. Haddad together with Y. Xu (Fudan Univ. China), R. El-Azouzi and T. Jimenez (Univ. of Avignon), and S.-E. Elayoubi (Orange Labs, Issy les Moulineaux) compute the starvation probability as a function of the initial buffering and study tradeoffs between the two performance measures: starvation probabilities and the pre-buffering delay.

6.6.5. *Stochastic geometry and network coding for distributed storage*

In [37] E. Altman and K. Avrachenkov in collaboration with J. Goseling (Twente Univ., The Netherlands) consider storage devices located in the plane according to a general point process and specialize the results for

the homogeneous Poisson process. A large data file is stored at the storage devices, which have limited storage capabilities. Hence, they can only store parts of the data. Clients can contact the storage devices to retrieve the data. The expected costs of obtaining the complete data under uncoded or coded data allocation strategies are compared. It is shown that for the general class of cost measures where the cost of retrieving data is increasing with the distance between client and storage devices, coded allocation outperforms uncoded allocation. The improvement offered by coding is quantified for two more specific classes of performance measures. Finally, the results are validated by computing the costs of the allocation strategies for the case that storage devices coincide with currently deployed mobile base stations.

6.7. Advances in Methodological Tools

Participants: Eitan Altman, Konstantin Avrachenkov, Ilaria Brunetti, Ioannis Dimitriou, Mahmoud El Chamie, Majed Haddad, Alain Jean-Marie, Philippe Nain, Giovanni Neglia.

6.7.1. *Queueing theory*

In [21] K. Avrachenkov and P. Nain in collaboration with U. Yechiali (Tel Aviv Univ., Israel) study a retrial queueing system with two independent Poisson streams of jobs flowing into 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. They study the queueing system using multi-dimensional probability generating functions, and derive its necessary and sufficient stability conditions while solving a Riemann-Hilbert boundary value problem. Various performance measures are calculated and numerical results are presented. In particular, numerical results demonstrate that the proposed multiple access system with two types of jobs and constant retrial rates provides incentives for the users to respect their contracts.

In [19] K. Avrachenkov in collaboration with E. Morozov (Petrozavodsk State Univ., Russia) consider a finite buffer capacity GI/GI/c/K-type retrial queueing system with constant retrial rate. The system consists of a primary queue and an orbit queue. The primary queue has c identical servers and can accommodate up to K jobs (including c jobs under service). If a newly arriving job finds the primary queue to be full, it joins the orbit queue. The original primary jobs arrive to the system according to a renewal process. The jobs have i.i.d. service times. The head of line job in the orbit queue retries to enter the primary queue after an exponentially distributed time independent of the length of the orbit queue. Telephone exchange systems, medium access protocols, optical networks with near-zero buffering and TCP short-file transfers are some telecommunication applications of the proposed queueing system. The model is also applicable in logistics. They establish sufficient stability conditions for this system. In addition to the known cases, the proposed model covers a number of new particular cases with the closed-form stability conditions. The stability conditions that they obtained have clear probabilistic interpretation.

In [20] K. Avrachenkov in collaboration with E. Morozov and R. Nekrasova (Petrozavodsk State Univ., Russia) and B. Steyaert (Ghent Univ., Belgium) study a retrial queueing system with N classes of customers, where a class-i blocked customer joins orbit i. Orbit i works like a single-server queueing system with (exponential) constant retrial time (with rate μ_{0i}) regardless of the orbit size. Such a system is motivated by multiple telecommunication applications, for instance wireless multi-access systems, and transmission control protocols. First, they present a review of some corresponding recent results related to a single-orbit retrial system. Then, using a regenerative approach, they deduce a set of necessary stability conditions for such a system. They will show that these conditions have a very clear probabilistic interpretation. They also performed a number of simulations to show that the obtained conditions delimit the stability domain with a remarkable accuracy, being in fact the (necessary and sufficient) stability criteria, at the very least for the 2-orbit M/M/1/1-type and M/Pareto/1/1-type retrial systems that they focus on.

In [75], I. Dimitriou investigates a single server system accepting two types of retrial customers and paired services. The service station can handle at most one customer, and if upon arrival a customer finds the server busy it is routed to an infinite capacity orbit queue according to its type. Upon a service completion epoch, if at least one orbit queue is non-empty, the server seeks to find customers from the orbits. If both orbit queues are non-empty, the seeking process will bring to the service area a pair of customers, one from each orbit. If only one is non-empty, then a customer from this orbit queue will be brought to the service area. However, if a primary customer arrives during the seeking process it will occupy the server immediately. It is shown that the joint stationary orbit queue length distribution at service completion epochs, can be determined via transformation to a Riemann boundary value problem. Stability condition is investigated, while an extension of the model is also discussed and analyzed. Numerical results are obtained and yield insight into the behavior of the system. The theoretical system can be used to model a relay node for two connections in wireless communication, where network coding is used.

When individuals have to take a decision on whether or not to join a queue, one may expect to have threshold equilibria in which customers join the queue if its size is smaller than a threshold and do not join if it exceeds the threshold. In [74], P. Wiecek (Wroclaw Univ. of Technology, Poland), E. Altman and A. Ghosh (Univ. of Pennsylvania, USA) have studied queueing in which the congestion cost per user decreases in the queue size. An example for such a situation is multicast communication where all individuals that participate in the multicast session share the transmission cost. They showed that many equilibria exist and computed the asymptotic system behavior as the arrival rate of individuals grows.

6.7.2. *Markov processes*

In [16] K. Avrachenkov in collaboration with A. Eshragh (Univ. of Adelaide, Australia) and J. Filar (Flinders Univ., Australia) present some algebraic properties of a particular class of probability transition matrices, namely, Hamiltonian transition matrices. Each matrix P in this class corresponds to a Hamiltonian cycle in a given graph G on n nodes and to an irreducible, periodic, Markov chain. They show that a number of important matrices traditionally associated with Markov chains, namely, the stationary, fundamental, deviation and the hitting time matrix all have elegant expansions in the first $n-1$ powers of P, whose coefficients can be explicitly derived. They also consider the resolvent-like matrices associated with any given Hamiltonian cycle and its reverse cycle and prove an identity about the product of these matrices. As an illustration of these analytical results, they exploit them to develop a new heuristic algorithm to determine a non-Hamiltonicity of a given graph.

6.7.3. *Control theory*

In [17] K. Avrachenkov and O. Habachi (former post-doc in MAESTRO) in collaboration with A. Piunovskiy and Y. Zhang (both from the Univ. of Liverpool, UK) investigate infinite-horizon deterministic optimal control problems with both gradual and impulsive controls, where any finitely many impulses are allowed simultaneously. Both discounted and long-run time-average criteria are considered. They establish very general and at the same time natural conditions, under which the dynamic programming approach results in an optimal feedback policy. The established theoretical results are applied to the Internet congestion control, and by solving analytically and non-trivially the underlying optimal control problems, they obtain a simple threshold-based active queue management scheme, which takes into account the main parameters of the transmission control protocols, and improves the fairness among the connections in a given network.

6.7.4. *Game theory*

6.7.4.1. *Estimating the Shapley-Shubik index*

In [15] K. Avrachenkov in collaboration with L. Cottatellucci (EURECOM) and L. Maggi (CREATE-NET, Italy) consider simple Markovian games, in which several states succeed each other over time, following an exogenous discrete-time Markov chain. In each state, a different simple static game is played by the same set of players. They investigate the approximation of the Shapley-Shubik power index in simple Markovian games (SSM). They prove that an exponential number of queries on coalition values is necessary for any deterministic algorithm even to approximate SSM with polynomial accuracy. Motivated by this, they propose and study three

randomized approaches to compute a confidence interval for SSM. They rest upon two different assumptions, static and dynamic, about the process through which the estimator agent learns the coalition values. Such approaches can also be utilized to compute confidence intervals for the Shapley value in any Markovian game. The proposed methods require a number of queries, which is polynomial in the number of players in order to achieve a polynomial accuracy.

6.7.4.2. Evolutionary games

Evolutionary games attempt to explain the evolution of species and the dynamics of competition. The player's utility is called "fitness" and a larger fitness indicates a larger rate of reproducibility. In standard evolutionary games, one studies interactions between individuals each of which is consider as a player. In [49], I. Brunetti, E. Altman, and R. El-Azouzi (Univ. of Avignon) argue that in many situations both in biology as well as in networking, one cannot attribute a fitness to an individual but rather to a group of individuals that behaves as an altruistic entity. For example, in a hive of bees it is only the queen that reproduces and thus one cannot model a single bee as a selfish player. They present new definitions for evolutionary games for such situations and study their equilibrium.

This, as well as other considerations in multi-population evolutionary games, is applied in [56] by H. Gaiech and R. El-Azouzi (Univ. of Avignon), M. Haddad, E. Altman and I. Mabrouki (Univ. of Manouba, Tunisia) to Multiple Access Control for which the equilibrium is explicitly computed.

In [84] E. Altman presents a summary of the foundations of classical evolutionary games addressed to a wide public. Both the equilibrium notion of ESS (Evolutionary Stable Strategy) as well as the replicator dynamics (which describes the non-equilibrium behavior) are presented.

6.7.4.3. Sequential Anonymous Games

Stationary anonymous sequential games are a special class of games that combines features from both population games (infinitely many players) with stochastic games. It allows studying competition in complex systems where each individual belongs to a community (which we call individual state) which may change in time as a result of actions taken by the individual. Unlike standard evolutionary games, a player does not just optimize its immediate reward (fitness) but some long term reward over the time. P. Wiecek (Wroclaw Univ. of Technology, Poland) and E. Altman proved in [42] the existence of an equilibrium for the general model and studied the two applications. The first one is described in §6.5.1 .

The second application is a maintenance repair problem: each of a large number of cars can decide whether to behave gently or to drive fast. By driving fast it takes larger risks for having an accident. The probability of an accident depends on the fraction of drivers that drive fast. An internal state of the car is either good (g) or bad (b). A car gets to a state b as a result of an accident and then it has some penalty and costs for repair. The advantage of driving fast is reducing delay costs. This problem is formulated as a sequential anonymous game and its equilibrium is computed. They computation makes use of the linear structure of both the transition probabilities and the immediate fitness in the global state.

6.7.5. Optimization

In [55] M. El Chamie and G. Neglia provide a methodology for solving smooth norm optimization problems under some linear constraints using the Newton's method. This problem arises in many machine learning and graph optimization applications. They show how Newton's method significantly outperforms gradient methods both in terms of convergence speed and in term of robustness to the step size selection.

SCALE Team

6. New Results

6.1. Programming Languages for Distributed Systems

One of the objectives of the Scale team is to design programming models easing the development and safe execution of distributed systems. This section describes our results in this direction.

6.1.1. Multi-active Objects

Participants: Ludovic Henrio, Fabrice Huet, Justine Rochas, Vincenzo Mastandrea.

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:

- We published the extension of multi-active objects to support scheduling and thread limitation [12].
- We developed a compiler from ABS language into ProActive multi-active objects. This translation can be generalised to many other active object languages. This work has been published as a research report [25], and is under submission to a conference.
- We started to work on static detection of deadlocks for multi-active object. This is the work of Vincenzo Mastandrea who is starting a Labex PhD in collaboration with the FOCUS EPI (Univ of Bologna).
- Extensive use of multiactive objects in our CAN P2P network and implementation of usecases [2].
- We formalised in Isabelle/HOL a first version of the semantics of multiactive objects. This work was done in collaboration with Florian Kammuller

We plan to continue to improve the model, especially about compile-time checking of annotations and about fault tolerance of multiactive objects.

6.1.2. Autonomic Monitoring and Management of Components

Participants: Françoise Baude, Ludovic Henrio.

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. 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.

- This year, we published a journal paper summarising our approach in the GCM/ProActive framework and our contribution on componentised membranes for autonomic computing [3].
- We also improved, in the context of the SCADA associate team and during the internship of Matias Ibañez, the support for autonomic components, providing all the architecture and API so that the programmer of autonomic aspect can do them in a DSL reconfiguration language, called GCMScript. This was implemented and experimented, a publication is under submission on this work.

6.1.3. Algorithmic skeletons

Participant: Ludovic 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. In the previous year we designed the possibility for a skeleton to output events, which increases the control and monitoring capabilities. This year we published our previous results in [14] and realised additional steps:

- Study of different ways to predict the execution time for a skeleton, inspired from simple statistic functions. This improvement together with the distributed execution of skeletons should allow us to publish a journal paper on this subject in 2015

6.1.4. *Optimization of data transfer in event-based programming models*

Participants: Iyad Alshabani, Françoise Baude, Laurent Pellegrino.

In [6], we extended a previous work with conceptual and experimental performance evaluations. This previous and collaborative work [1] developed an innovative approach of "lazy copy and transfer" of the data parts of event objects exchanged by peers in the context of event-driven architecture applications.

While event notifications are routed in a conventional manner through an event service, data parts of the events are directly and transparently transferred from publishers to subscribers. The theoretical analysis shows that we can reduce the average event delivery time by half, compared to a conventional approach requiring the full mediation of the event service. The experimental analysis confirms that the proposed approach outperforms the conventional one (both for throughput and delivery time) even though the middleware overhead, introduced by the specific adopted model, slightly reduces the expected benefits.

6.1.5. *Behavioural Semantics*

Participants: Ludovic Henrio, Eric Madelaine, Min Zhang.

We have studied Parameterised Networks of Automata (pNets) from a theoretical perspective. We started with some 'pragmatic' expressiveness of the pNets formalism, showing how to express a wide range of classical constructs of (value-passing) process calculi, but also complex interaction patterns used in modern distributed systems. Our framework can model full systems, using (closed) hierarchies of pNets; we can also build (open) pNet systems expressing composition operators. Concerning more fundamental aspects, we defined a strong bisimulation theory specifically for the pNet model, proved its properties, and illustrated it on some examples. One of the original aspects of the approach is to relate the compositional nature of pNets with the notion of bisimulation; this was exemplified by studying the properties of a flattening operator for pNets. This work has been accepted for publication at PDP'2015 ([24]).

6.1.6. *A Time-sensitive Heterogeneous Behavioural Model*

Participants: Eric Madelaine, Yanwen Chen.

This work concludes the PhD research of Yanwen Chen, targetting a timed-sensitive extension of the pNets model with logical clocks inherited from the CCSL language. The main results of this year are: 1) a new notion of Time Specification (TS), used to handle the abstract properties of each level of processes in a pNet structure, 2) algorithms to compute such TSs for basic parameterized and timed processes, and from composition of timed-pNets, 3) conditions for checking the compatibility of composition, 4) a use-case from the area of intelligent transportation systems, illustrating the whole chain of modeling, upto a symbolic simulation of the full composed system, with the TimeSquare tool. This work was published as [4], [23], and in the PhD thesis of Y. Chen, defended on 2014, Nov. 30th.

6.1.7. *Structure and structural correctness for GCM components*

Participants: Ludovic Henrio, Oleksandra Kulankhina, Eric Madelaine.

We have defined a set of rules characterizing the well-formed composition of components in order to guarantee their safe deployment and execution. This work focuses on the structural aspects of component composition; it puts together most of the concepts common to many component models, but never formalized as a whole. Our formalization characterizes correct component architectures made of functional and non-functional aspects, both structured as component assemblies. So-called 'Interceptor chains' can be used for a safe and controlled interaction between the two aspects. Our well-formed components guarantee a set of properties ensuring that the deployed component system has a correct architecture and can run safely. Those definitions constitute the formal basis for VerCors tool. This work was done in the context of O. Kulankhina phd research, and in collaboration with Dongqian Liu (ECNU Shanghai), as part of the Associated Team DAESD.

6.2. Run-time/middle-ware level

6.2.1. *Scalable and robust Middleware for distributed event based computing*
Participants: Françoise Baude, Fabrice Huet, Laurent Pellegrino, Maeva Antoine.

In the context of the FP7 STREP PLAY and French SocEDA ANR research projects terminated late 2013, we initiated and pursued the design and development of the Event Cloud. This has been the core content of Laurent Pellegrino PhD thesis [2], and the corresponding software deposit at the APP for this middleware.

As a distributed system, this middleware can suffer from failures. To resist to such situations, we have added a capability of checkpointing. In [18] we present how to design an adaptation of the famous Chandy and Lamport algorithm for distributed snapshot taking, to the case of the Event Cloud. Indeed, as the Event Cloud peers are multi-active objects, we need to take care when and how to serve the chekpointing request and so, when to apply the Chandy Lamport protocol operations. Consequently, we have make sure that the obtained distributed snapshot is indeed consistent. As publication of events are triggered from the outside of the Event Cloud, we however are not able to recover them from the last saved snapshot in case of peer crash and subsequent whole Event Cloud recovery. However, we ensure any event injected through a peer, before this peer was participating in the last global checkpoint taking is safely part of it.

As a distributed system handling huge amount of information, this middleware can suffer from data imbalances. In [22], [8], we have reviewed the litterature of structured peer to peer systems regarding the way they handle load imbalance. We have generalized those popular approaches by proposing a core API that we have proved to be indeed also applicable to the Event Cloud middleware way of implementing a load balancing policy.

Storing highly skewed data in a distributed system has become a very frequent issue, in particular with the emergence of semantic web and big data. This often leads to biased data dissemination among nodes. Addressing load imbalance is necessary, especially to minimize response time and avoid workload being handled by only one or few nodes. We have proposed a protocol which allows a peer to change its hash function at runtime, without a priori knowledge regarding data distribution. This provides a simple but efficient adaptive load balancing mechanism. Moreover, we have shown that a structured overlay can still be consistent event when all peers do not apply the same hash function on data [7].

6.2.2. *Virtual Machines Placement Algorithms*
Participants: Fabien Hermenier, Vincent Kherbache.

In [21], [19], we present BtrPlace as an application of the dynamic bin packing problem with a focus on its dynamic and heterogeneous nature. We advocate flexibility to answer these issues and present the theoretical aspects of BtrPlace and its modeling using Constraint Programming. In [5] we rely on BtrPlace to achieve energy efficiency. To maintain an energy footprint as low as possible, data centres manage their VMs according to conventional and established rules. Each data centre is however made unique due to its hardware and workload specificities. This prevents the *ad-hoc* design of current VM schedulers from taking these particularities into account to provide additional energy savings. In this paper, we present Plug4Green, an application that relies on BtrPlace to customize an energy-aware VM scheduler. This flexibility is validated through the implementation of 23 SLA constraints and 2 objectives aiming at reducing either the power

consumption or the greenhouse gas emissions. On a heterogeneous test bed, Plug4Green specialization to fit the hardware and the workload specificities allowed to reduce the energy consumption and the gas emission by up to 33% and 34%, respectively. Finally, simulations showed that Plug4Green is capable of computing an improved placement for 7,500 VMs running on 1,500 servers within a minute.

Finally, we started to investigate on easing the jobs of data centre operators using BtrPlace. For example, server maintenance is a common but still critical operation. A prerequisite is indeed to relocate elsewhere the VMs running on the production servers to prepare them for the maintenance. When the maintenance focuses several servers, this may lead to a costly relocation of several VMs so the migration plan must be chose wisely. This however implies to master numerous human, technical, and economical aspects that play a role in the design of a quality migration plan. In[13], we study migration plans that can be decided by an operator to prepare for an hardware upgrade or a server refresh on multiple servers. We exhibit performance bottleneck and pitfalls that reduce the plan efficiency. We then discuss and validate possible improvements deduced from the knowledge of the environment peculiarities.

6.3. Application level

6.3.1. Simulation Software Architecture

Participant: Olivier Dalle.

In general purpose software engineering (as opposed to simulation software engineering), the motivations for reuse have long been advocated and demonstrated: lower risks of defects, collective support of potentially larger user community, lower development costs, and so on. In simulation software architectures, we can also cite business-specific motivations, such as providing a better reproducibility of simulation experiments, or avoiding a complex validation process. In [20], we show that although it is rarely discussed, reuse is a problem that may be considered in two opposite directions: reusing and being reused.

6.3.2. DEVS-based Modeling & Simulation

Participants: Olivier Dalle, Damian Vicino.

DEVS is a formalism for the specification of discrete-event simulation models, proposed by Zeigler in the 70's, that is still the subject of many research in the simulation community. Surprisingly, the problem of representing the time in this formalism has always been somehow neglected, and most DEVS simulators keep using Floating Point numbers for their arithmetics on time values, which leads to a range of systematic errors, including severe ones such as breaking the causal relations in the model. In [16] we propose a new data type for discretized time representation in DEVS, based on rational numbers. Indeed, we show that rational numbers offer good stability properties for the arithmetics used in DEVS, with a limited impact on the simulation execution performance.

6.3.3. GPU-based High Performance Cloud Computing

Participants: Michael Benguigui, Françoise Baude, Fabrice Huet.

To address HPC, GPU devices are now considered as unavoidable cheap, energy efficient, and very efficient alternative computing units. 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.

As a challenging example, we have pursued our work on pricing American multi-dimensional (so very computation intensive) options in finance. From our previous work that achieved pricing a 40-assets based american option within 8 hours of computation on a single GPU, the work in [9] allows us to reach approximatively one hour of computation time. For this, we run using active objects coupled with OpenCL codes, on 18 GPU nodes acquired from the Grid'5000 platform (the maximum amount of available GPU on Grid'5000 that we could book at once).

Moreover, the balancing of work is taking in consideration the heterogeneous nature of the involved GPUs, and is capable to harness the computing power of multi-core CPUs that also support running OpenCL codes. This parallel and distributed pricing approach is also extended in the forthcoming PhD thesis of Michael Benguigui: it successfully tackles the Value At Risk computation of a portofolio composed of such complex financial products.

6.3.4. *Simulation of Software-Defined Networks*

Participants: Olivier Dalle, Damian Vicino.

Software Defined Networks (SDN) is a new technology that has gained a lot of attention recently. It introduces programmatic ways to reorganize the network logical topology. To achieve this, the network interacts with a set of controllers, that can dynamically update the configuration of the network routing equipments based on the received events. As often with new network technologies, discrete-event simulation proves to be an invaluable tool for understanding and analzing the performance and behavior of the new systems. In [17], we use such smulations for evaluating the impact of Software-Defined Networks' Reactive Routing on BitTorrent performance. Indeed, BitTorrent uses choking algorithms that continuously open and close connections to different peers. Software Defined Networks implementing Reactive Routing may be negatively affecting the performances of the system under specific conditions because of it lack of knowledge of BitTorrent strategies.

<p align="center" style="color:red">**AYIN Team**</p>

5. New Results

5.1. Highlights of the Year

- Yuliya Tarabalka was nominated CR1 since 1 January 2015.
- Josiane Zerubia was elected for a duration of 6 years at the board of directors of the French Society of Photogrammetry and Remote Sensing (SFPT, http://www.sfpt.fr/).
- Josiane Zerubia was invited by Technion to give a plenary talk at SIMA'14 in Ein Gedi, Israel organized for the 60th birthday of Prof. Alfred Bruckstein in May, http://www.cs.technion.ac.il/SIMA14/.

5.2. Markov Random Fields

5.2.1. *Fusion of multitemporal and multiresolution remote sensing data and application to natural disasters*

Participants: Ihsen Hedhli, Josiane Zerubia [contact].

This work was carried out in collaboration with Prof. Gabriele Moser and Prof. Sebastiano Serpico from DITEN departement, University of Genoa, Italy.

Multitemporal data, Multiresolution data, Supervised classification, Hierarchical Markov random fields.

The capabilities to monitor the Earth surface, and especially urban and built-up areas, from environmental disasters such as floods or earthquakes, and to assess the ground impact and damage of such events play primary roles from multiple social, economic, and human viewpoints. In this framework, accurate and time-efficient classification methods are especially important tools to support rapid and reliable assessment of the ground changes and damages induced by a disaster, in particular when an extensive area has been affected. Given the huge amount and variety of data available currently from last-generation very-high resolution (VHR) satellite missions, (such as Pléiades, COSMO-SkyMed, or WorldView-2), the main difficulty is to develop a classifier that can take benefit of multiband, multiresolution, multidate, and possibly multisensor input imagery. In such a context, Markov random field (MRF) models are widely used to solve classification problems as they permit one to integrate contextual information into the classification scheme. Due to their non-causal nature, these models generally lead to iterative inference algorithms that are computationally demanding (e.g., optimization via simulated annealing), thereby justifying the choice of a hierarchical structure, with good methodological and application-oriented properties such as: (i) the causality in scale, under Markovian assumption, which allows the use of a non-iterative algorithm with acceptable computational time and (ii) the possibility to incorporate images acquired at multiple resolutions in the hierarchy for multiresolution and multisensor fusion purposes [10]. In the proposed method, multidate and multiresolution fusion is based on explicit statistical modeling through hierarchical Markov random field modeling. The model allows both input data collected at multiple resolutions and additional multiscale features derived through wavelets to be fused. The proposed approach consists of a supervised Bayesian classifier that combines: (i) a joint class-conditional statistical model for pixelwise information and (ii) a hierarchical MRF for spatio-temporal and multiresolution contextual information. Step (i) deals, first, with the modeling of the marginal statistics of the spectral channels acquired at each resolution and conditioned to each class. Step (ii) consists in the integration of this statistical modeling in a hierarchical Markov random field for each date. An especially novel element of the proposed approach is the use of multiple quad-trees in cascade (see Figure 1), each associated with each new available image at different dates, with the aim to characterize the temporal correlations associated with distinct images in the input time series and to support the joint analysis of multitemporal, multiresolution, and possibly multisensor imagery. The transition probabilities between scales and between different dates determine the hierarchical MRF since they formalize the causality of the statistical interactions involved [11].
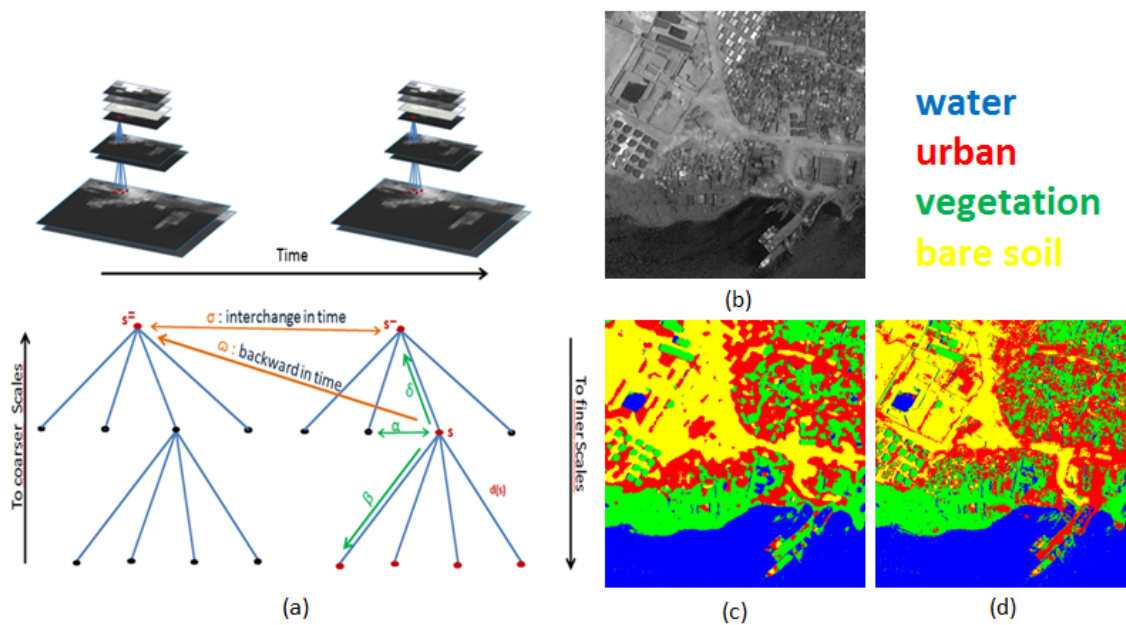
Figure 1. a) Multitemporal hierarchical structure; b) Panchromatic image of Port au Prince (Pléiades, ©CNES distribution Airbus DS, 2013); c) Classification map using single date hierarchical structure; d) Classification map obtained through the proposed multitemporal method.

5.2.2. *A multi-layer Markov model for change detection in temporally separated aerial image pairs*

Participant: Josiane Zerubia [contact].

This work was carried out in collaboration with Prof. Zoltan Kato from Institute of Informatics, University of Szeged, Hungary [http://www.inf.u-szeged.hu/~kato/], and Praveer Singh from Institut Mines-Telecom.

Multilayer Markov Random Fields (MRF), Histogram of Gradients (HOG), change detection, graph-cut optimization, aerial/satellite images.

In the proposed approach developed last year, we have tried to include both texture as well as pixel level information to build a three layer Markov model using the *Histogram of Oriented Gradients (HOG)* and the *Gray Level Difference* features on the topmost and bottommost layer respectively. Using a ground truth (GT) mask defined manually by an expert for each of the image pairs in the data set (obtained from the Hungarian Institute of Geodesy, Cartography and Remote Sensing), we employ a supervised technique to mark the initial set of pixels / sites as foreground or background. On the basis of the *HOG difference* and the *Gray level difference* feature vector corresponding to all the pixels in the image pair, a probability density function is fitted individually for the binary label set comprising of foreground and background labels using the GT. The probabilistic estimate is calculated using one training image pair for each data set. Using this probabilistic measure, a negative log likelihood is computed for each pixel (for both the features as well as the binary label set) which is then passed to the energy function of the proposed 3-layer MRF model. The final segmentation is obtained by minimizing the energy using a graph-cut algorithm, and subsequently a final foreground and background labelling is obtained over the combined layer. Figure 2 , shows aerial image pairs, one of them captured in 1984 by FOMI, Hungary (a) and the other one by GoogleEarth in 2007 (c). (b) is the ground truth and (d) is a combination of the hierarchical MRF based change detection (in red), ground truth (in green) and changes detected correctly (in yellow). This year, we have made a comparison of this method with two other multilayer MRFs for change detection developed at MTA-SZTAKI in Budapest, Hungary.
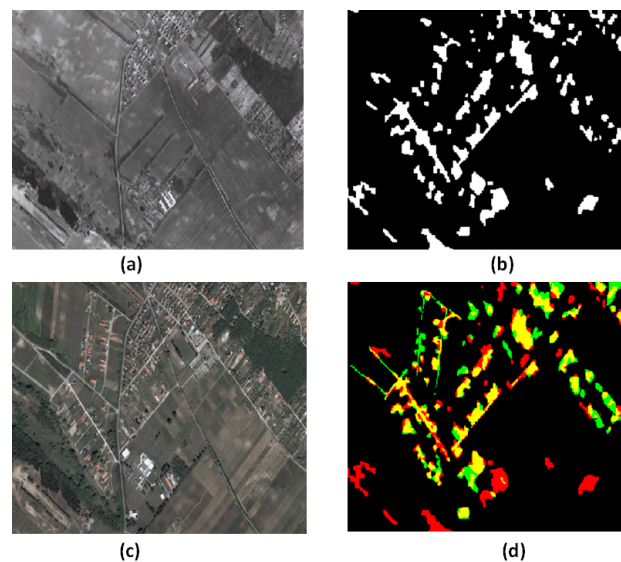

Figure 2. Change detection in an aerial image pair using a hierarchical MRF. a) Aerial image captured in 1984 by ©FOMI; b) Ground truth; c) Aerial image captured by ©GoogleEarth in 2007; d) Combination of the hierarchical MRF based change detection (in red), ground truth (in green), and changes detected correctly (in yellow).

5.2.3. *Graph-cut-based model for spectral-spatial classification of hyperspectral images*
Participant: Yuliya Tarabalka [contact].

This work has been done in collaboration with Aakanksha Rana (Institut Mines-Telecom/EURECOM).

Hyperspectral images, graph cut, multi-label alpha expansion, contextual information, energy minimization

The very high spatial and spectral resolution of the last generation of remote sensors provides rich information about every pixel in an image scene, hence opening new perspectives in classification, but also presenting the challenge of analysing high data volumes. While pixel-wise classification methods analyze each pixel independently, classification results can be significantly improved by including spatial information in a classifier.

In this work, we proposed a spectral-spatial method for hyperspectral image classification based on a graph cut [15]. The classification task is expressed as an energy minimization problem on the spatio-temporal graph of image pixels, and is solved by using the graph-cut α-expansion approach. The energy to optimize is computed as a sum of data and interaction energy terms, respectively. The data energy term is computed using the outputs of the probabilistic support vector machines classification. The second energy term, which expresses the interaction between spatially adjacent pixels in the eight-neighborhood, is computed by using dissimilarity measures between spectral vectors, such as vector norms, spectral angle map, or spectral information divergence. The performance of the proposed method was validated on hyperspectral images captured by the ROSIS and the AVIRIS sensors. Figure 3 compares classification results obtained by applying support vector machines and the proposed approach for the ROSIS hyperspectral image acquired over the University of Pavia. The new method yields higher classification accuracies when compared to the recent state-of-the-art approaches.
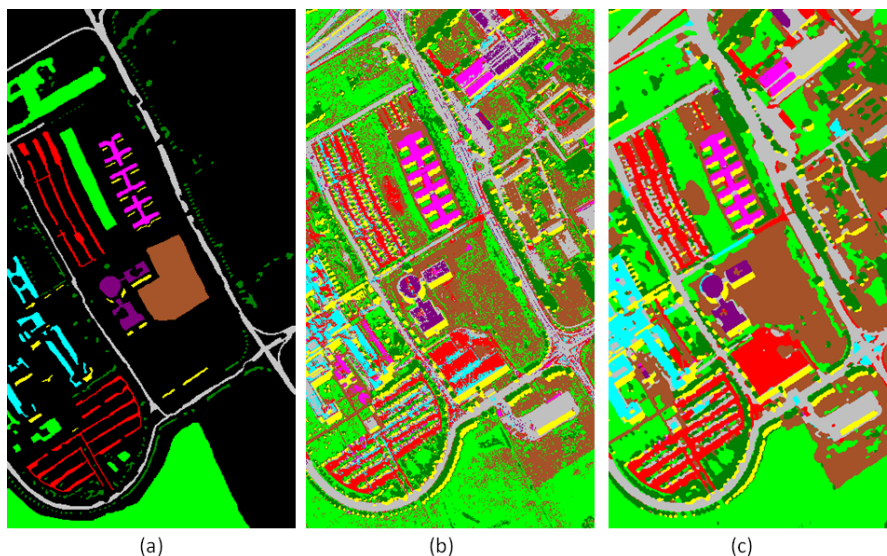

Figure 3. Hyperspectral image of the University of Pavia. (a) Ground-truth (b) Support vector machines classification map. (c) Graph-cut classification map.

5.3. Marked point processes

5.3.1. *Multiple target tracking using spatio-temporal marked point processes*
Participants: Paula Craciun, Josiane Zerubia [contact].

Multiple target tracking, stochastic geometry, point processes, remote sensing

Tracking can be defined as the problem of estimating the trajectories of objects in the image plane, as they move around the scene. Hence, a tracker assigns consistent labels to the objects in different frames of a sequence of images and can additionally provide information about the orientation, shape or size of the objects. Multi-target tracking has been historically achieved using sequential techniques, the major drawback of such methods residing in the impossibility to modify past results in the light of new data. However, applications such as offline video processing or information retrieval are not sequential in nature. Batch processing methods are preferred in this case since they do not suffer from the limitations of sequential methods. Nevertheless, these techniques remain poorly explored and highly underused.

We propose a novel approach based on spatio-temporal marked point processes to detect and track moving objects in a batch of high resolution images [17]. We develop a new, intuitive energy based model consisting of several terms that take into account both the image evidence and physical constraints such as target dynamics, track persistence and mutual exclusion. We construct a suitable optimization scheme that allows us to find strong local minima of the proposed highly non-convex energy [9]. The model has been validated on two types of data: remotely sensed satellite image sequences, characterized by high resolution, high signal to noise ratio and low temporal frequency; and biological image sequences, characterized by high resolution, low signal to noise ratio and high temporal frequency.

Tracking results are shown in Figure 4 , which shows the detection (dots) and tracking (lines) results of boats in a sequence of 14 high resolution remotely sensed images. The images are captured with a low temporal frequency at different acquisition angles.

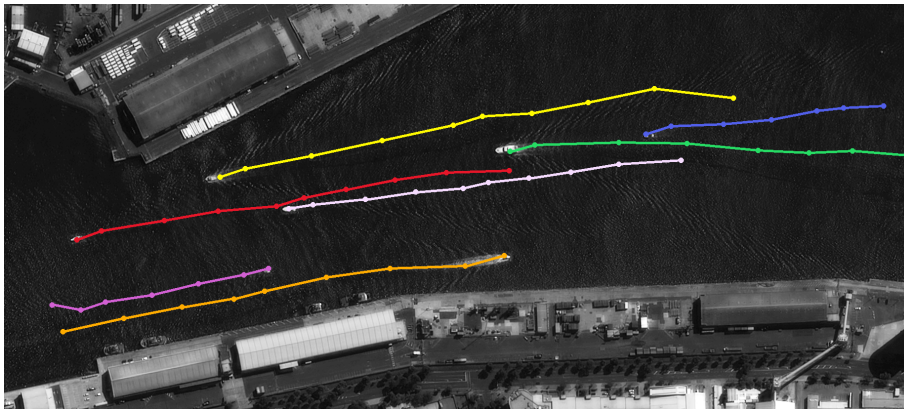*Figure 4. Detection and tracking results on a sequence of satellite images taken at different angles ©Inria/AYIN.*

5.3.2. Initialization and estimation of parameters for marked point processes applied to automatic object detection on satellite images

Participants: Aurélie Boisbunon, Josiane Zerubia [contact].

Sparse representations, large scale, stochastic algorithms, machine learning, image processing

Marked point processes (MPP) strongly rely on parameters, whose estimation affects both computation time and performances. In this work, we proposed two approaches: the first one consists in initializing MPPs with a first coarse solution obtained very quickly from sparse regularization methods [7], while the second one estimates the parameters by the Stochastic Approximation Expectation-Maximization (SAEM) algorithm [8]. We give details on both approaches below.

The first coarse solution is obtained from a deterministic sparse regularization method. This method is based on the representation of an image with objects as a sum of convolutions between atoms of a dictionary and matrices of positions of the objects in the image. Such a representation is displayed on Figure 5 . The atoms of the dictionary are fixed in advance and correspond to different instances of the objects (scales, angles, shapes, etc). This way, we transform the problem of object detection into the problem of estimating extremely sparse matrices. The algorithm we derived for solving the associated optimization problem is both parallelized and very efficient.

Up to recently, the parameters of MPPs were estimated by the Stochastic Expectation-Maximization (SEM) algorithm developed by Celeux & Diebolt (1985). This algorithm consists in alternatively estimating the expected pseudo-likelihood based on a random configuration and updating the parameter value by maximum of pseudo-likelihood. However, since it does not have a pointwise convergence, Ben Hadj et al. (2010) considered running a simulated annealing scheme after few iterations of SEM in order to reach convergence, at the cost of a higher computational time. Instead, we proposed to adapt the Stochastic Approximation Expectation-Maximization (SAEM) algorithm, developed by Delyon et al. (1999), to MPPs. Indeed, it both offers pointwise convergence and a similar computational time as SEM by efficiently taking into account past configurations in the update of the expected pseudo-likelihood.

Using both approaches resulted in the division of the computational time of the estimation of MPPs parameters by 2 and in an increase in performance of detection.
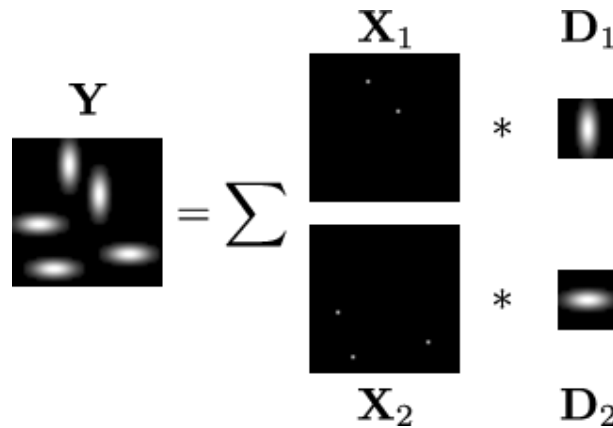

Figure 5. Representation of an image as a sum of convolutions between atoms of a dictionary and matrices of positions.

5.3.3. *Generic curvilinear structure modeling via marked point process theory*
Participants: Seong-Gyun Jeong, Yuliya Tarabalka, Josiane Zerubia [contact].

Curvilinear structure extraction, object detection, marked point process, stochastic inference

We proposed a marked point process model to analyze underlying curvilinear structure for wide ranges of input data, for instance, wrinkles, DNA filaments, road cracks, and blood vessels [12], [13]. It is based on sampling technique so that the model represents an arbitrary shape of the line network with a set of small line segments. The line segments should be fit into the given image data, and be harmonic with those of neighborhoods. To take these issues into consideration, we formulate a maximum a posteriori (MAP) estimation as an energy minimization problem. The energy function for given line configuration **s** can be decomposed into data likelihood term E_{data} and prior term E_{prior}:

$$E(\mathbf{s}) = \sum_i^{\#(\mathbf{s})} E_{\text{data}}(s_i) + \lambda \sum_{i \sim j} E_{\text{prior}}(s_i, s_j),$$

where $\#(\mathbf{s})$ denotes the total number of line segments in the current configuration, $i \sim j$ represents symmetric neighborhood system, and λ controls the relative importance of two terms. For the data term, we exploit oriented gradient information and homogeneity of the pixel intensities corresponding to line segment on the image site. The prior energy defines topology of the line configuration in that penalizes overlapping and attracts smooth connections. Another contribution of the work is to reduce parameter dependencies of the marked point process model using aggregation approach. We repeated to perform Markov chain Monte Carlo (MCMC) sampling with different parameter vectors to obtain multiple line hypotheses. Then, we combine line hypotheses to maximize the consensus among detection results.
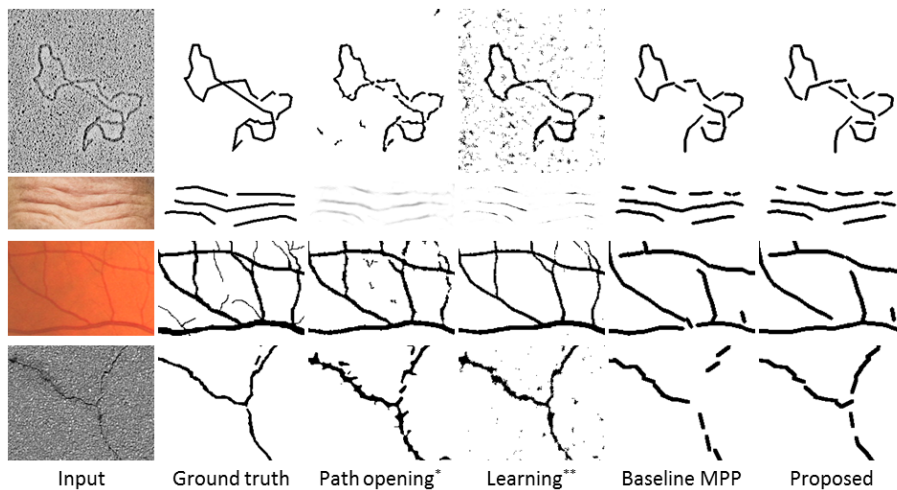
Figure 6. Comparison of the line detection results on DNA filaments, wrinkles, retina, and road cracks (top to bottom).

In figure 6 , we have compared line detection results of manually labeled image, morphological filtering (path opening), supervised feature learning, an MPP model using single parameter vector, and the proposed algorithm. The proposed algorithm extracts the most salient line structures for all datasets without any parameter estimation procedure.

5.4. Shapes and contours

5.4.1. *Riemannian metrics on spaces of curves and surfaces*

Participant: Ian Jermyn [contact].

This work is being done in collaboration with Anuj Srivastava of Florida State University [https://www.fsu. edu/].

Shape, Riemannian, metric, elastic, curve, surface, functional data, alignment

Statistical shape modelling has many applications in image processing and beyond. One of the key problems in this area is to develop and understand measures of shape similarity. One approach uses Riemannian metrics induced on 'shape space' by Riemannian metrics on the space of embeddings. Current work is focused on generalizing to surfaces the elastic metric used for curve embeddings, and in finding surface representations that simplify computations in the same way that the square root velocity representation simplifies computations in the case of curves. The notion of a 'square-root normal field' (SRNF), which leads to a reduced version of the full elastic metric, is a promising possibility in this direction.

The most recent work [16] has focused on estimating the inverse of the SRNF map. If this can be done even approximately, a very efficient framework results: the surfaces, represented by their SRNFs, can be efficiently analyzed using standard Euclidean tools, and only the final results need to be mapped back to the surface space. In this work, we developed a procedure for inverting SRNF maps of star-shaped surfaces, a special case for which analytic results can be obtained. We tested our method via the classification of 34 cases of ADHD (Attention Deficit Hyperactivity Disorder), plus controls, in the Detroit Fetal Alcohol and Drug Exposure Cohort study. We obtained state-of-the-art results.

Ground Truth (f_o)   Analytic Inversion ($\widetilde{f}$)   Numerical Solution (f^*)

$E(f_o; q_o) = 0$ $E(\widetilde{f}; q_o) = 5.7\text{E-}4$ $E(f^*; q_o) = .9\text{E-}4$

Energy   Errors on Surface
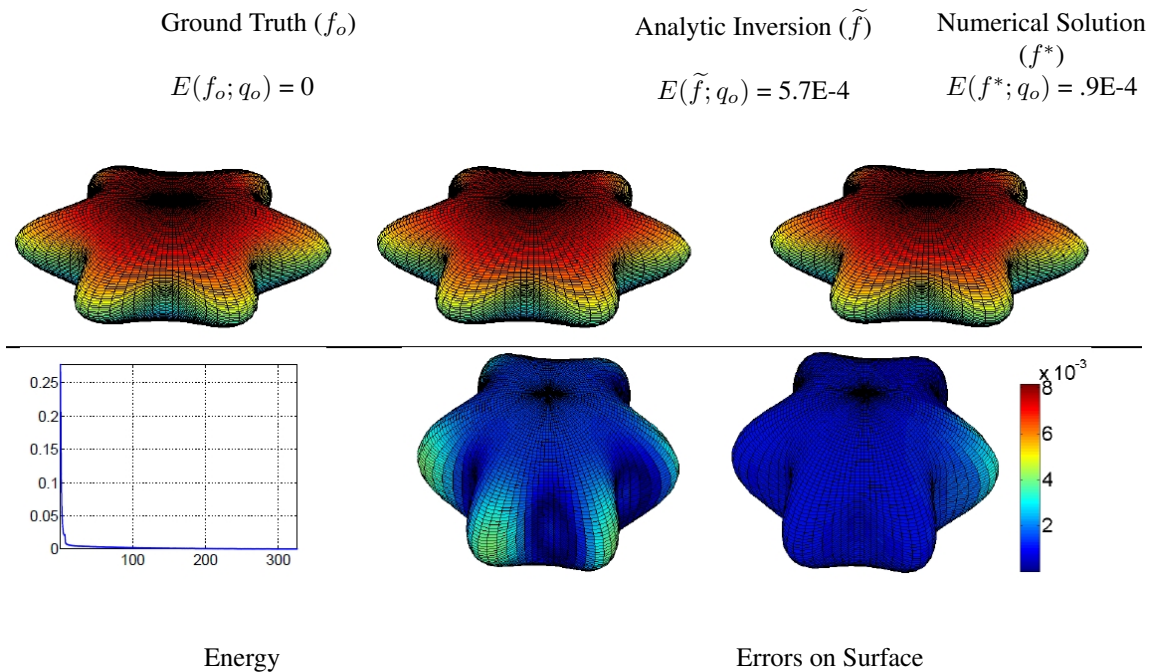
Figure 7. Reconstructing a surface from its SRNF. A target surface (f_o) is numerically reconstructed as f^* with initialization as the unit sphere. The energy plot shows the evolution of energy against iterations with initialization as a unit sphere. The analytically inverted surface $\widetilde{f}$ is shown for comparison. The corresponding energies $E(\widetilde{f}; q_o)$ and $E(f^*; q_o)$ are also shown. The errors between the reconstructed surfaces and the ground truth are shown on the ground truth surface with colours representing the magnitudes, i.e. $|f^*(s) - f_o(s)|$ for all $s \in S^2$.

5.4.2. Enforcing monotonous shape growth or shrinkage in video segmentation

Participant: Yuliya Tarabalka [contact].

This work has been done in collaboration with Dr. Guillaume Charpiat (STARS team, Inria-SAM), Dr. Bjoern Menze (Technische Universität München, Germany and Asclepios team, Inria-SAM), and Dr. Ludovic Brucker (NASA GSFC, USA) [http://www.nasa.gov].

Video segmentation, graph cut, shape analysis, shape growth

Automatic segmentation of objects from video data is a difficult task, especially when image sequences are subject to low signal-to-noise ratio or low contrast between the intensities of neighboring structures. Such challenging data are acquired routinely, for example, in medical imaging or satellite remote sensing. While individual frames can be analyzed independently, temporal coherence in image sequences provides a lot of information not available for a single image. In this work, we focused on segmenting shapes that grow or shrink monotonically in time, from sequences of extremely noisy images.

We proposed a new method for the joint segmentation of monotonically growing or shrinking shapes in a time sequence of images with low signal-to-noise ratio [3]. The task of segmenting the image time series is expressed as an optimization problem using the spatio-temporal graph of pixels, in which we are able to impose the constraint of shape growth or shrinkage by introducing unidirectional infinite-weight links connecting pixels at the same spatial locations in successive image frames. The globally-optimal solution is computed with a graph-cut algorithm. The performance of the proposed method was validated on three applications: segmentation of melting sea ice floes; of growing burned areas from time series of 2D satellite images; and of a growing brain tumor from sequences of 3D medical scans. In the latter application, we imposed an additional inter-sequences inclusion constraint by adding directed infinite-weight links between pixels of dependent image structures. Figure 8 shows a multi-year sea ice floe segmentation result. The proposed method proved to be robust to high noise and low contrast, and to cope well with missing data. Moreover, it showed linear complexity in practice.
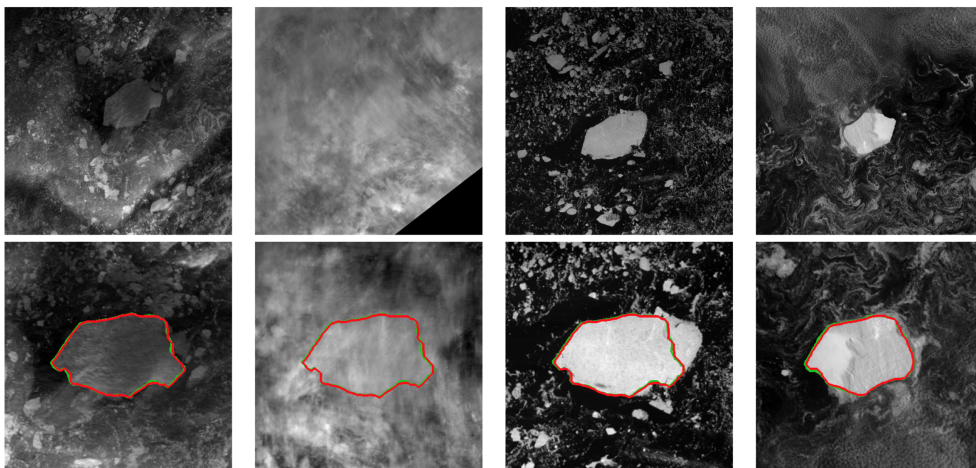
Figure 8. Top: MODIS images for four time moments (days 230, 233, 235 and 267 of 2008, respectively). Bottom: corresponding aligned images with segmentation contours (in red). Manual segmentation is shown in green.

5.4.3. *Multi-label image segmentation with partition trees and shape prior*

Participants: Emmanuel Maggiori, Yuliya Tarabalka [contact].

This work has been done in collaboration with Dr. Guillaume Charpiat (STARS team, Inria-SAM).

Partition trees, multi-class segmentation, shape priors, graph cut

The multi-label segmentation of images is one of the great challenges in computer vision. It consists in the simultaneous partitioning of an image into regions and the assignment of labels to each of the segments. The problem can be posed as the minimization of an energy with respect to a set of variables which can take one of

multiple labels. Throughout the years, several efforts have been done in the design of algorithms that minimize such energies.

We proposed a new framework for multi-label image segmentation with shape priors using a binary partition tree [19]. In the literature, such trees are used to represent hierarchical partitions of images, and are usually computed in a bottom-up manner based on color similarities, then processed to detect objects with a known shape prior. However, not considering shape priors during the construction phase induces mistakes in the later segmentation. This paper proposes a method which uses both color distribution and shape priors to optimize the trees for image segmentation. The method consists in pruning and regrafting tree branches in order to minimize the energy of the best segmentation that can be extracted from the tree. Theoretical guarantees help reducing the search space and make the optimization efficient (see Figure 9 (i)) and [19]. Our experiments (see Figure 9) show that the optimization approach succeeds in incorporating shape information into multi-label segmentation, outperforming the state-of-the-art.

(a) Color image. (b) Manual segment. (c) BPT optimization.

(d) Manual. (e) SVM. (f) GC. (g) BPT Opt.

	$\mathcal{A}$	$\mathcal{D}$	Time [s]	Time [s] (Props. 1-2)
SVM	0.764	0.50		
GC	0.799	0.49		
TC	0.804	0.47	214	84
Opt.	0.813	0.69		

(h) Performance. (i) Execution time.
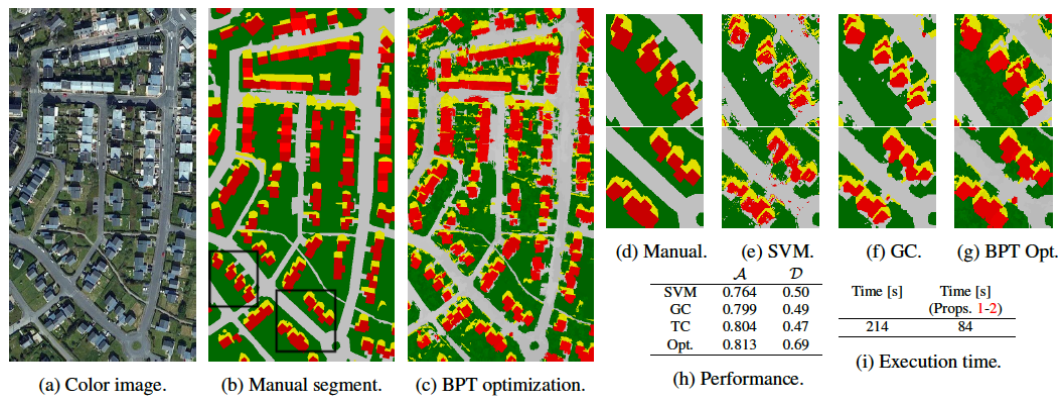
Figure 9. Classification results for the satellite image over Brest. $\mathcal{A}$ denotes overall classification accuracy, and $\mathcal{D}$ denotes average building's overlap. The performance of the proposed binary partition tree (BPT) optimization method is compared with the following methods: 1) support vector machines (SVM) classification; 2) graph cut (GC) with α-expansion; 3) cut on the BPT, regularized by the number of regions without using shape priors (TC).

5.5. Other detection approaches

5.5.1. *Image-based evaluation of treatment responses of facial wrinkles using LDDMM registration and Gabor features*

Participants: Nazre Batool, Josiane Zerubia [contact].

Face, skin texture, detection of wrinkles, LDDMM registration, response to treatment, Gabor filters, morphological processing

The goal of this work is to evaluate quantitatively the subtle variations in facial wrinkles for the same subject in response to treatment using image-based analysis. The novelty of this application is that a series of images of the same subject over a shorter time period of weeks are analyzed instead of more prevalent inter-person analysis of facial images. To overcome the challenges of detecting and evaluating such subtle changes, we propose a framework to compare image features in key wrinkle sites only while excluding the noise introduced by changes in surrounding skin texture. After initial registration using facial landmarks such as corners of eyes, nose, mouth, we propose a method based on Large Deformation Diffeomorphic Metric Mapping (LDDMM) to achieve finer registration. Fig. 10 (1a-1e) shows an example of registration using LDDMM for a pair of images.

Then we use N. Batool's previously proposed algorithm (Nazre & Chellappa (2015)) to detect key wrinkle sites. The algorithm is based on 'scaled' maximum Gabor filter responses and the incorporation of geometric constraints via morphological image processing. The binary output from the algorithm is used to create a unique wrinkle template for each subject. Fig. 10 (2a-2d) an example of obtaining a unique wrinkle template from an image using Gabor responses and wrinkle detection algorithm in (Nazre & Chellappa (2015)). Gabor responses in this template,in time series images are compared to detect subtle changes for a subject. We do not adopt the direct approach of comparing filter responses in the whole image instead of those in wrinkle template only because such an approach causes intermingling of skin texture variations in non-wrinkle sites with changes in wrinkle sites degrading the overall accuracy.
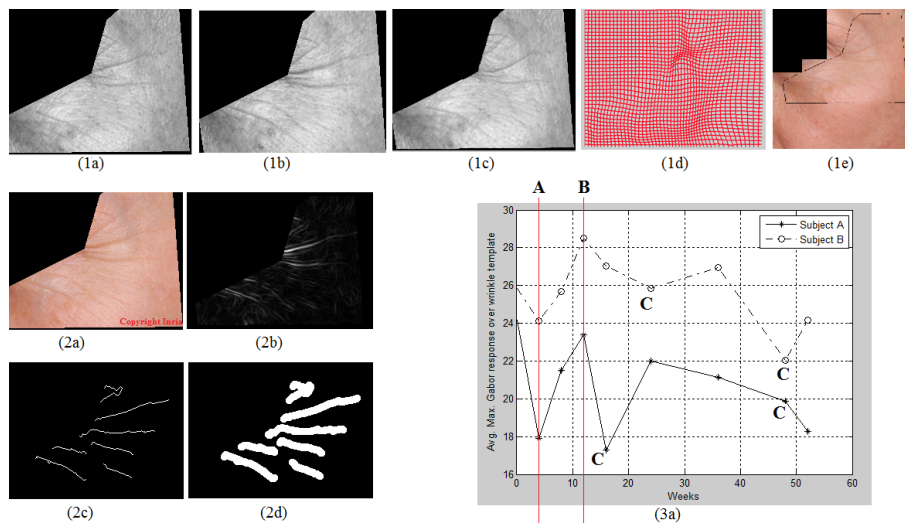
Figure 10. Overview of the evaluation framework. (1a) Week 4 image. (1b) Baseline image. (1c) Week 4 image registered using LDDMM to baseline image. (1d) Deformation of the underlying 2D space. (1e) Deformed week 4 images aligned in the original face image. (2a) Baseline image. (2b) Gabor maximum amplitude response. (2c) Detected wrinkles. (2d) The template for key wrinkle sites. (3) Plot of results for two subjects.

Fig. 10 (3a) shows a plot of results for two subjects where y-axis shows average maximum Gabor amplitude response in key wrinkle sites and x-axis corresponds to the number of weeks after the treatment. For both subjects a significant drop in the average response can been seen 4 weeks after the treatment (event 'A'). An increase in the Gabor response happened at week 12 (event 'B') which coincided with slight darkening/reddening of skin for both subjects. On the other hand, event 'C' represents co-occurrence of skin lightening with a decrease in Gabor response. These preliminary results indicate trends in wrinkle responses to treatment, skin darkening and lightening. In future, these trends will be validated by more rigorous experiments.

5.5.2. *SAR data classification using generalized Gamma mixture model*

Participant: Josiane Zerubia [contact].

This work has been performed in collaboration with Dr. Vladimir Krylov (University of Genoa, Italy), Prof. Heng-Chao Li, Prof. Ping-Zhi Fan (Southwest Jiaotong University, Chengdu, China) and Prof. William Emery (University of Colorado, Boulder, USA).

SAR images, statistical modeling, generalized Gamma mixture model

The accurate statistical modeling of synthetic aperture radar (SAR) images is a crucial problem in the context of effective SAR image processing, interpretation and application. In this work a semi-parametric approach is designed within the framework of finite mixture models based on the generalized Gamma distribution (GΓD) in view of its flexibility and compact analytical form. Specifically, we have developed a generalized Gamma mixture model (GΓMM) to implement an effective statistical analysis of high-resolution SAR images and proved the identifiability of such mixtures. A low-complexity unsupervised estimation method has been derived by combining the proposed histogram-based expectation-conditional maximization algorithm and the Figueiredo-Jain mixture estimation algorithm. This resulted in a numerical maximum likelihood (ML) estimator that can simultaneously determine the ML estimates of component parameters and the optimal number of mixture components. The state-of-the-art performance of the proposed method has been validated experimentally on a wide range of high-resolution SAR amplitude and intensity images.
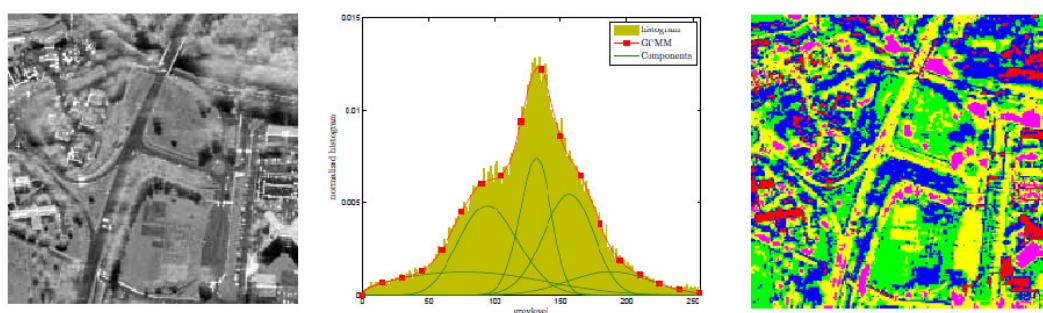
Figure 11. Statistical modeling of a RAMSES (©CNES, ONERA) image (left) by generalized Gamma mixture model (middle) and its visualization by maximum likelihood classification (right).

In Fig. 11 we demonstrate a typical result of the developed statistical modeling technique on a portion of a multilook airborne RAMSES (©CNES, ONERA) sensor acquisition over Toulouse suburbs (single polarization, downsampled to approximately 2m ground resolution). The unsupervised GΓMM estimate contains five components and reports a very accurate result that outperforms the considered benchmark statistical modeling methods. In order to visualize the estimated five statistical components we also report a maximum likelihood classification map.

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

6. New Results

6.1. Highlights of the Year

- *Michael Thomazo was awarded the AFIA Prize 2014 (French Association for Artificial Intelligence) for his PhD entitled "Conjunctive Query Answering Under Existential Rules - Decidability, Complexity, and Algorithms" defended in October 2013. He was also awarded the first accessit of Gilles Kahn Prize 2014 by the SIF (French Society for Computer Science) [14].*

- *Madalina Croitoru and Alain Gutierrez were awarded the Best Technical Paper of SGAI-2014 for "On Ontological Expressivity and Modelling Argumentation Schemes using COGUI", in collaboration with Wael Hamdan, Rady Khazem and Ghaida Rebdawi .*

- *Abdallah Arioua was awarded the Best Student Paper Award of SGAI-2014 for "Query Failure Explanation in Inconsistent Knowledge Bases: A Dialogical Approach" in collaboration with Nouredine Tamani, Madalina Croitoru and Patrice Buche .*

BEST PAPERS AWARDS :

[36] **AI'2014: Thirty-fourth SGAI International Conference on Artificial Intelligence**. W. HAMDAN, R. KHAZEM, G. REBDAWI, M. CROITORU, A. GUTIERREZ, P. BUCHE.

[28] **AI'2014: 34th SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence**. A. ARIOUA, N. TAMANI, M. CROITORU, P. BUCHE.

6.2. Ontology-Based Query Answering with Existential Rules

Participants: Jean-François Baget, Fabien Garreau, Mélanie König, Michel Leclère, Marie-Laure Mugnier, Swan Rocher, Federico Ulliana.

Ontolology-based query answering (and more generally *Ontology-Based Data Access*, *OBDA*) is a new paradigm in data management, which takes into account inferences enabled by an ontology when querying data. In other words, the notion of a database is replaced by that of a knowledge base, composed of data (also called facts) and of an ontology. In this context, *existential rules* (also called Datalog+) have been proposed to represent the ontological component [59], [58]. This expressive formalism generalizes both description logics used in OBDA (such as $\mathcal{EL}$ and DL-Lite), which form the cores of so-called tractable profiles of the Semantic Web ontological language OWL2) and Datalog, the language of deductive databases. Since about five years, we have been studying the theoretical foundations of this framework (mainly concerning decidability and complexity) and developing associated algorithmic techniques. We have started the development of a platform dedicated to OBDA with existential rules (see section 5.2).

Before presenting this year' results, we recall the two classical ways of processing rules, namely forward chaining and backward chaining, also known as "materialization" and "query rewriting" in the OBDA setting. In forward chaining, the rules are applied to enrich the initial data and query answering can then be solved solved by evaluating the query against the "saturate" database (as in a classical database system *i.e.*, with forgetting the rules). The backward chaining process can be divided into two steps: first, the initial query is rewritten using the rules into a first-order query (typically a union of conjunctive queries, UCQ); then the rewritten query is evaluated against the initial database (again, as in a classical database system). Since entailment is not decidable with general existential rules, both forward and backwards processes may not halt.

6.2.1. Improvement of Query Rewriting Algorithms

These last two years, we designed and implemented a query rewriting algorithm that takes as input a set of existential rules and a UCQ q and outputs a UCQ, which is a sound and complete rewriting of q, whenever such a rewriting exists [60], [61], [62]. This year's main improvement to this algorithm is the implementation of a unifier able to process rules without decomposing their head into single atoms. This improvement appeared to be have a very high impact on the efficiency of query rewriting (up to 274 quicker on an ontology where 32% of the rules have a head composed of two atoms instead of a single one). Beside, much effort has been devoted to experiments: to find appropriate benchmarks, to build a translator from the Semantic Web format OWL/OWL2 to our existential rule format dlgp (since most existing ontologies are available in OWL/OWL2 format), to select existing tools to compare with, run them, finally compare tools on several criteria.

- *Results partially published in the Semantic Web Journal [22].*

Query rewriting techniques have the interest of being independent from the data. However, a main bottleneck is that the size of the rewritten query can be exponential in the size of the original query, hence the produced rewriting maybe not usable in practice. A well-known source of combinatorial explosion are some very simple rules, which form the core of any ontology, typically expressing concept and relation hierarchies, concept properties and relation signatures. We have proposed a rewriting technique, which consists in compiling these rules into a preorder on atoms and embedding this preorder into the rewriting process. This allows to compute compact rewritings that can be considered as "pivotal" representations, in the sense that they can be easily translated into different kinds of queries that can be evaluated by different kinds of database systems. The provided algorithm computes a sound, complete and minimal UCQ rewriting, if one exists. Experiments show that this technique leads to substantial gains in the query rewriting process, in terms of size and runtime, and scales on very large ontologies (several ten thousands of rules).

- *Results not published yet. Reported in Mélanie König's PhD thesis [17].*

6.2.2. A Better Approximation of Chase Termination for Existential Rules and their Extension to Non-monotonic Negation

Forward chaining with existential rules is known as the *chase* in databases. Various acyclicity notions ensuring chase termination have been proposed in the knowledge representation and databases. Acyclicity conditions found in the literature can be classified into two main families: the first one constrains the way existential variables are propagated during the chase and the second one constrains dependencies between rules *i.e.*, the fact that a rule may lead to trigger another rule. These conditions are based on different graphs, but all of them can be seen as forbidding "dangerous" cycles in the considered graph. We defined a new family of graphs that allows to unify and strictly generalize these acyclicity notions without increasing worst-case complexity.

Second, we considered the extension to existential rules with nonmonotonic negation under stable model semantics and further extended acyclicity results obtained in the positive case by exploiting negative information.

- *This work is part of Fabien Garreau and Swan Rocher's PhD theses. Results published at the European Conference on Artificial Intelligence (ECAI 2014)[30](long version as an arXiv report) and at the Workshop on Non-monotonic Reasoning (NMR 2014) [31].*

6.2.3. Detailed Results and Complements on Query Answering under Greedy Bounded-Treewidth Sets of Existential Rules

The family of greedy bounded-treewidth sets of existential rules (gbts) is an expressive class of rules for which entailment is decidable. This decidability property relies on a structural property of the saturation by the chase (*i.e.*, the set of inferred facts): for any initial set of facts, the saturation of these facts has a bounded treewidth (where the treewidth is computed on a graph associated with a set of atoms). Moreover, a tree decomposition of bounded width can be incrementally built during the chase. This family generalizes the important family of guarded existential rules, which itself generalizes Horn description logics used in OBDA.

In papers published at IJCAI 2011 and KR 2012, we studied the complexity of entailment under gbts rules as well as under known subclasses of gbts (with respect to data, combined and query complexity) and provided a generic algorithm with optimal worst-case complexity. This year, we finally completed a long report (75 pages) containing the detailed proofs of the results, some of them being very technical; in this report, we also clarified and reformulated the description of the generic algorithm, according to Michael Thomazo's PhD thesis (defended in October 2013); finally, we complemented the landscape of gbts classes by studying the complexity of all subclasses obtained by combining the syntactic criteria which define already known classes.

- *Results available as an arXiv report [56]. Submitted to a major journal in Artificial Intelligence. In collaboration with Sebastian Rudolph (TU Dresden) and Michael Thomazo (now postdoctoral student in Sebastian Rudolph's group).*

6.2.4. *Extracting Bounded-level Modules from Deductive RDF Triplestores*

The Semantic Web is consolidating a legacy of well-established knowledge bases spanning from life sciences, to geographic data and encyclopedical repositories. Today, reusing knowledge and data available online is vital to ensure a coherent development of the Semantic Web, thereby capitalizing on the efforts made in the last years by many institutions and domain experts to publish quality information.

In this paper we studied how to extract modules from RDF knowledge bases equipped with Datalog inference rules, we called Deductive RDF Triplestores. A module is a Deductive RDF Triplestore entailed from the reference system, which is defined upon a restricted vocabulary (or signature). We proposed a new semantics for bounded-level modules allowing to control their size, and then presented extraction algorithms compliant with the novel semantics. This feature is helpful since many ontologies are extremely large, while users often need to reuse only a small part of resources in their work.

This work was partially carried out before the arrival of Federico Ulliana at GraphIK. For the future, we plan to study module extraction for knowledge bases equipped with existential rules, which extend the rules considered here.

- *Results published at the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI 15) [44]. In collaboration with Marie-Christine Rousset from LIG (University of Grenoble).*

6.2.5. *Axiomatisation of Consistent Query Answering via Belief Revision*

This work takes place in the OBQA setting where a query is being asked over a set of knowledge bases defined over a common ontology. When the union of knowledge bases along with the ontology is inconsistent, several semantics have been defined which are tolerant to inconsistency. These semantics all rely on computing repairs, *i.e.*, maximal (in terms of set inclusion) consistent subsets of the data set. They have been studied from a productivity point of view and a complexity point of view. We take a new point of view to define axiomatic characterisations of two such semantics, namely IAR (Intersection of All Repairs) and ICR ((Intersection of Closed Repairs). We argue that such characterisation can provide an alternative way of comparing the semantics and new insights into their properties. Furthermore such axiomatisation can be used when proposing a generalisation of inconsistency tolerant semantics. In order to provide the axiomatic characterisations we define belief revision operators that correspond to IAR and ICR.

- *Work published at [43]. In collaboration with Ricardo Rodriguez from University of Buenos Aires.*

6.3. Reasoning with Imperfect Information and Priorities

Participants: Abdallah Arioua, Patrice Buche, Madalina Croitoru, Jérôme Fortin, Souhila Kaci, Namrata Patel, Tjitze Rienstra, Nouredine Tamani, Rallou Thomopoulos.

This work focuses on two main notions, namely *argumentation systems*, which allow to represent and deal with conflicting information, and formalisms to represent *preferences*, which allow to compare possible outcomes in decision making and recommender systems.

6.3.1. Fundamental Aspects of Argumentation

A Dung-style argumentation framework aims at representing conflicts among elements called arguments. The basic ingredients of this framework is a set of arguments and a Boolean abstract (*i.e.*, its origin is not known) binary defeat relation on these arguments. This abstract framework can be instantiated in different ways, by representing arguments in a given knowledge representation formalism, which allows to take the semantics of arguments into account in the computation of the defeat relation.

Preference-based argumentation frameworks are instantiations of Dung's framework in which the defeat relation is derived from an attack relation and a preference relation over the arguments. Recently, Dung's framework has been extended in order to consider the strength of the defeat relation *i.e.*, to quantify the degree to which an argument defeats another argument. In this work, we instantiated this extended framework by a preference-based argumentation framework with a valued preference relation. As particular cases, the latter can be derived from a weight function over the arguments or a Boolean preference relation. We showed under some reasonable conditions that there are "less situations" in which a defense between arguments holds with a valued preference relation compared to a Boolean preference relation. Finally, we provided some conditions that the valued preference relation shall satisfy when it is derived from a weight function.

- *This is a joint work with Christophe Labreuche from Thales and published in [20]*

We also considered an extension to argumentative frameworks based on fuzzy set theory. The knowledge base is fuzzified to allow agents expressing their expertise (facts and rules) attached with grades of importance in the unit interval. Arguments are then attached with a strength score aggregating the importance expressed on their facts and rules. Extensions, corresponding to subsets of consistent arguments, are also attached with forces computed based on their strong arguments. The forces are used then to rank extensions from the strongest to the weakest one, upon which decisions can be made. We have also shown that the strength preference relation defined over arguments is reasonable according to classical rationality postulates and our fuzzy logical argumentation system can be seen as a computationally efficient instantiation of the generic model of structured argumentation framework. We furthered our theoretical research and demonstrate the added value of our approach in the practical setting of the European project EcoBioCap (see Sect.8.2).

- *Work published in IPMU 2014 [47] and Fuzz IEEE 2014 [46].*

One instantiation, among many others, of Dung's framework consists in constructing the arguments from a set of propositional logic formulas. Thus an argument is seen as a reason for or against the truth of a particular statement. Despite its advantages, the argumentation approach for inconsistency handling also has important shortcomings. More precisely, in some applications what one is interested in are not so much only the conclusions supported by the arguments but also the precise explanations of such conclusions. We showed that argumentation framework applied to classical logic formulas is not suitable to deal with this problem. On the other hand, intuitionistic logic appears to be a natural alternative candidate logic (instead of classical logic) to instantiate Dung's framework. We developed *constructive argumentation framework*. We showed that intuitionistic logic offers nice and desirable properties of the arguments. We also provided a characterization of the arguments in this setting in terms of minimal inconsistent subsets when intuitionistic logic is embedded in the modal logic S4.

- *This is a joint work with Yakoub Salhi from CRIL and published in [39]*

Lastly, we developed a model of abduction in abstract argumentation, where changes to an argumentation framework act as hypotheses to explain the support of an observation. We presented dialogical proof theories for the main decision problems (*i.e.*, finding hypotheses that explain skeptical/credulous support) and we showed that our model can be instantiated on the basis of abductive logic programs.

- *This work has been done in Tjitze Rienstra's thesis and published in [32].*

6.3.2. Use of Argumentation in Explanation, Querying and Decision Making

Besides work on the foundations of argumentation frameworks, we have studied the use of argumentation techniques in various tasks: explanation of query failure, reverse engineering, and decison making. These studies are mainly motivated by agri-food scenarii: bread conception, packaging conception, and durum wheat conception.

We have proposed an argumentation-based explanation for *query failure explanation* under the inconsistency tolerant semantics ICR in an Ontology-Based Data Access setting with existential rules. We used a rule-based language and we base our work on the equivalence between ICR-based query answering in inconsistent knowledge bases and sceptical acceptance of arguments. We proposed a multilevel explanation that exploits both the inference power of the logical language as well as arguments of dialectical nature. We also investigated an interactive argumentative approach where the process of explanation takes the form of a dialogue between the user and the reasoner.

- *Work published in COMMA 2014 [27] and SGAI 2014 [28] where it received the best student paper award.*

Within the framework of the European project EcoBioCap http://www.ecobiocap.eu about the design of next generation packagings using advanced composite structures based on constituents derived from the food industry, we have been developing a Decision Support System (DSS) for packaging material selection. [40], [49]. The DSS consists of two steps: (1) aggregating possibly conflicting needs expressed by several parties involved in the considered field and (2) querying a database of packagings with the resulting aggregation obtained at point (1). We instantiate for each need, called viewpoint or aspect, an argumentation system to reason about arguments solely expressed on it [45]. This will then be used to generate the query on the packaging database. To this aim we show how to instantiate ASPIC with the DLR-Lite logic modeling expert ontologies in this real world scenario [47].

- *Work published in AAMAS 2014 [45], IPMU 2014 [47], ICCS 2014 [40], and COMMA 2014 [49].*

Evaluating food quality is a complex process since it relies on numerous criteria historically grouped into four main types: nutritional, sensorial, practical and hygienic qualities. They may be completed by other emerging preoccupations such as the environmental impact, economic phenomena, etc. However, all these aspects of quality and their various components are not always compatible and their simultaneous improvement is a problem that sometimes has no obvious solution, which corresponds to a real issue for decision making. We propose a decision support method guided by the objectives defined for the end products of an agrifood chain. It is materialized by a backward chaining approach based on argumentation [47]. An extended version of this paper reporting on experimental results and expert evaluation has been published in Ecological Informatics [24].

- *Work published in IPMU 2014 [47], and Ecological Informatics 2014 [24].*

Knowledge elicitation, representation and reasoning explanation by / to non-computing experts has always been considered as a crafty task due to difficulty of expressing logical statements by non-logicians. We use the COGUI editor in order to elicit and represent argumentation schemes expressed using existantial rules within an inconsistent knowledge base. COGUI is a visual, graph based knowledge representation editor compatible with main Semantic Web languages (see Section 5.1). COGUI allows for default reasoning on top of ontologies. We investigate its use for modelling and reasoning using argumentation schemes and discuss the advantages of such representation. We show how this approach can be useful in the practical setting of EcoBioCap where the different argumentation schemes can be used to lead reasoning.

- *Work published in SGAI 2014 [36] where it received the best technical paper award. In collaboration with Wael Hamdan, Rady Khazem and Ghaisa Rebdawi from the Higher Institute of Applied Science and Technology (HIAST), Syria.*

6.3.3. Preferences

Qualitative and comparative preference statements of the form "prefer α to β" are useful components of many applications. This statement leads to the comparison of two sets of alternatives: the set of alternatives in which α is true and the set of alternatives in which β is true. Different ways are possible to compare two sets of objects leading to what is commonly known as preference semantics. The choice of the semantics to employ is important as they differently rank-order alternatives. Existing semantics are based on philosophical and non-monotonic reasoning grounds. In the meanwhile, they have been widely and mainly investigated by AI researchers from algorithmic point of view. We came to this problem from a new angle and completed existing

theoretical investigations of the semantics. In particular, we provided a comparison of the semantics on the basis of their psychological plausibility by evaluating their closeness to human behavior.

- *This is a joint work with Eric Raufaste from CLLE and published in [38]*

There has been a growing interest in the study of preferences for their utility in solving problems related with decision making. Most of the preference representation languages developed in the literature are based on comparative preference statements since they offer a simple and intuitive way for expressing preferences. They can be further interpreted following different semantics, imparting a greater flexibility on how outcomes can be compared. So far the main objective has been to rank-order the set of outcomes given a set of comparative preference statements and one or several semantics. Tackling this problem from a different angle, we looked into the behavioral aspects of the preference semantics and statements by attempting to formalise the intuition behind them using postulates studied in preference logics and non-monotonic reasoning. We selected the postulates w.r.t. three criteria: coherence, syntax independence and inference. Thus, our analysis provided a means to determine those properties that are satisfied for a given preference semantics.

- *This work has been done in Namrata Patel's thesis and published in [21]*

Intelligent 'services' are increasingly used on e-commerce platforms to provide assistance to customers. Numerous preference elicitation methods developed in the literature are now employed for this purpose. However, it is commonly known that there is a real bottleneck in preference handling as concerns the elicitation of preferences because it does not cater to the wide range of preference representation languages available. Thus, as a first step in developing a decision-support tool using an AI based on such languages, this paper describes a preliminary study conducted to address this issue. We proposed a method of eliciting real-time user preferences expressed in natural language (NL) which can be formally represented using comparative preference statements complying with different semantics, and provided a proof of concept to demonstrate its feasibility. Since we developed NL resources to detect preference semantics, we also made a comparative study with existing resources to underline the peculiarities of our model.

- *This work has been done in Namrata Patel's thesis and published in [37]*

6.4. 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. 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 . 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.

We investigate this problem in the specific context of bibliographic databases. Indeed, people working in bibliographical information systems have a lasting tradition of using norms and have integrated, along collections of documents notices (e.g. bibliographic records), collections of authority notices that categorize the different named entities used to describe documents (people, organizations, places, ...). In current databases, documents notices do not use directly the names of named entities to fill a particular field (author, editor, ...), but the unique identifier of the authority notice representing that named entity.

A few years ago, we began a collaboration with ABES (National Bibliographic Agency for Universities) to develop a method and a prototype to perform entity resolution between on one hand the authors of a new bibliographic record, and, on the other the authority references of an authority catalog (and namely the Sudoc catalog from the ABES agency). A problem with this approach is that it relies upon pre-established links between bibliographic records and authority notices. However, our experimentation and evaluation have shown that many existing links were erroneous, and thus led to the propagation of new linkage errors. We have thus began to work on methods and tools to repair linkage errors in bibliographical databases. The first step of our approach was to build a knowledge-base over an ontology (based on the international standards FRBR and CIDOC-CRM) aiming at representing bibliographic data (an RDFS base) as well as librarian knowledge.

From that, we developed a methodological framework allowing to design rules concluding on the coreference or the difference between entities of the bibliographic knowledge base. This framework was implemented in Cogui.

6.4.1. *An Original Methodology to Compute Coreference and Difference Links*

Our methodology can be briefly summarized as follows. The first step consists in computing "sure" links. In the second step, authority notices are enriched by information that comes from bibliographic notices to which they are linked by sure links. In the third step, Datalog rules that conclude on coreference or difference are triggered. The results are used to compute new sure links. These steps are iterated until stability *i.e.*, no new sure link is discovered. More specifically, the Datalog rules are the following form. The body of a rule is a conjunction of similarity criteria on attributes and its head states the coreference or the difference of two individual entities with a specific confidence level (represented as a symbolic value). We are currently instantiating this methodology for the Sudoc catalog, jointly with the ABES librarians, which will allow them to evaluate it.

6.4.2. *Partioning Semantics for Link Discovery in Bibliographic Knowledge Bases*

With the aim of evaluating and improving the quality of links in bibliographical knowledge bases, we have developed a decision support system based on partitioning semantics. The novelty of our approach consists in using symbolic values criteria for partitioning and suitable partitioning semantics. We have investigated the limits of those partitioning semantics: how the characteristics of the input (objects and criteria) influences characteristics of the result, namely correctness of the result and execution time. We have also evaluated and compared the above mentioned semantics on a real qualitative sample. This sample is issued from the catalogue of French university libraries (SUDOC) maintained by ABES.

- *This work is part of Lea Guizol's PhD thesis [16]. Work published in Fuzz IEEE 2014 [46].*

6.4.3. *Key Discovery on the Semantic Web*

Many techniques were recently proposed to automate the linkage of RDF datasets. Predicate selection is the step of the linkage process that consists in selecting the smallest set of relevant predicates needed to enable instance comparison. We call keys this set of predicates that is analogous to the notion of keys in relational databases. We have formally explained the different assumptions behind two existing key semantics (IC), and have evaluated experimentally these keys semantics by studying how discovered keys could help dataset interlinking or cleaning.

- *Work published in IC 2014 [50] and ICCS 2014 [29] in collaboration with Manuel Atencia and Jerome David from LIG, and Nathalie Pernelle, Fatiha Sais and Danai Symeonidou from LRI. See also the reconciliation-based approach in[23].*

6.4.4. *Fusion of Linked Data*

The problem of data fusion starts from reconciled datasets, whose objects are linked with semantic sameAs relations, as described above. We attempt to merge the often conflicting information of these reconciled objects in order to obtain unified representations that only contain the best quality information. We are studying an approach to determine the most appropriate value(s). Our method combines different quality criteria based on the value and its data source, and exploits, whenever possible, the ontology semantics, constraints and relations. Moreover we create a mechanism to provide explanations about the quality of each value, as estimated by our system. To achieve this, we generate annotations used for traceability and explanation purposes.

- *Work described in the Qualinca deliverable 4.2 research report, and accepted for publication in EGC'2015 : "Linked Data Annotation and Fusion driven by Data Quality Evaluation" (authors: Ioanna Giannopoulou, Fatiha Saïs from LRI, and Rallou Thomopoulos)*

.

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

6. New Results

6.1. Highlights of the Year

Yves Papegay received a "Wolfram Innovator Award" in December 2014

6.2. Robotics

6.2.1. Cable-driven parallel robots (CDPR)

6.2.1.1. Analysis of Cable-driven parallel robots

Participants: Alessandro Berti, Laurent Blanchet, Houssein Lamine, Jean-Pierre Merlet [correspondant], Yves Papegay, Rémy Ramadour.

We have continued the analysis of suspended CDPRs for control and design purposes. For control it is essential to determine the current pose of the robot for given leg lengths. This forward kinematic problem (FK) is usually very complex and admits several solutions. For parallel robot with rigid legs we have established the important property (P) that the FK may be solved in real-time i.e. being given the leg lengths ρ and platform pose $\mathbf{X}$ it is possible to determine the single pose $\mathbf{X_1}$ that can be reached from $\mathbf{X}$ if the leg lengths has been changed to $\rho + \Delta\rho$ provided that $\Delta\rho$ satisfies some properties. For CDPR with sagging cables determining all the FK solutions is more complex but we have proposed the first algorithm to solve it the for a full scale model of sagging cables [24]. For CDPR with non sagging cables the problem is also very complex because we cannot make any assumption on the number of cables under tension i.e for a CDPR with m cables we have to solve all the FK problems for all possible set of cables under tension from 1 to m and as soon as this number is lower than 6 the system of equations is much larger than for classical robots. We have however been able to propose an interval analysis based algorithm that allow one to get all the solutions [18]. But we have also shown that for non sagging cables the property (P) does not hold. Indeed it requires that the system of equations that governs the FK remains the same at all time. But for CDPR this system depend on the set of cables under tension (which is called the *cable configuration CC*) and it may change when the cable lengths change from 'ρ to $\rho + \Delta\rho$, even for redundant CDPR [23]. If the CC changes at some point the pose solution of the FK together with the cable tensions will differ from the one that is obtained when assuming no change in the CC. This has a drastic effect on control as we have now a system whose state equations may change over time but also on design as in the new CC the cable tensions may be quite different from the expected one. Hence property (P) will hold if and only if we are able to show that there will not be any change in the CC during the change of the cable lengths and therefore it is crucial to detect CC changes. But this require to fully simulate the discrete-time control laws together with the behavior of the coiling system. We have been able to implement a simulation tool that tracks a trajectory for the robot for arbitrary control laws and coiling system model [22], [25]. The principle of the algorithm is to determine if on a time interval $[t, t + \delta t]$ the solution of the FK with the current CC satisfies (P) by using Kantorovitch theorem. If this is not the case Δt is divided by 2 and the process is repeated. We then check if there is a time t_1 in $[t, t + \delta t]$ for which the tension of a cable in the CC may become 0 If there is no such t_1 for any cable in the current CC, then it will be the CC at time $t + \delta t$ and we may compute the pose and cable tensions at any time in $[t, t + \delta t]$. If there is a least one such t_1 (and there may be several $t_1^i, t_1^j, ...$ as we consider each cable in the CC) we order these times by increasing values and check sequentially if a cable tension become negative with the current CC at time $(t_1^i + t_1^{i+1})/2$. If yes we determine what can the CC at this time by looking at all possible CC. As soon as the new CC at time t_1^i has been determinated the simulation can go on. Implementing this algorithm has been difficult mainly for numerical reasons: the accuracy of the calculation may sometime exceed the floating point accuracy and we have to resort to symbolic computation and extended arithmetics. Our tests have shown that indeed CC changes may occur on trajectories: on a typical trajectory up to 10 different CC will appear with 5 or 6 cables under tension. These results have been confirmed experimentally on a prototype at LIRMM.

6.2.1.2. Tool for Agencement Analysis and Synthesis of CDPRs
Participants: Laurent Blanchet, Jean-Pierre Merlet [correspondant].

HEPHAISTOS has been working on tools to design the layout and geometry of CDPRs, while accounting for numerical errors as well as practical errors – actual position of the winches, of the attachments on the platform, errors of the controllers, of the cables, etc. Within this work, collision analysis plays an important role. Indeed the concept of cable robot aims to increase the workspace that is restricted for robots having rigid legs but interferences may reduce this workspace. Two types of interference analysis approaches exist: intersection of numerically-mapped boundaries (InB) and distance between features (DbF). The two sets of interference types that can be analysed using these approaches are distinct but overlapping. The first approach greatly benefits from Inria's computational geometry research and particularly from the AABB tree algorithms implemented in CGAL. Algorithms and implementation based on those were developed, along with several new algorithm and implementation to extend the scope of intersection types, and thus, of interference types. Algorithms to improve efficiency of given intersection types were also developed. We have already used the second approach, DbF, to develop algorithms for leg interference of parallel robots that are very efficient for non deformable cables but now well adapted for sagging cables. An interference detection algorithm has been developed and implemented for a restricted scope of applications [10], and research is on-going for a more generic case.

6.2.1.3. Visual-servoing of a parallel cable-driven robot
Participants: Rémy Ramadour, Jean-Pierre Merlet [correspondant], François Chaumette [correspondant].

The last two years, we studied how visual servoing could improve accuracy, controllability and performance of cable-driven parallel robots [13]. Previous works on this domain showed very interesting results but some issues remained to be investigated , such as :

- ratio accuracy/workspace : cable-driven parallel robots are known to allow a large reachable workspace, but also to have complex geometric and dynamic models which affect the accuracy. Using visual-servoing in a closed-loop scheme, we were able to enhance the accuracy by a factor of ten, allowing to manipulate daily-life objects in a whole living room.

- image-based joint-space control : in order to reach a desired pose, the usual method involves several computing and evaluations of both the Jacobian matrix of the manipulator and the interaction matrix linking visual features to the displacements of the end-effector. We designed a control scheme, based on an iterative updating using the Broyden update law, in order to link the visual features directly to the joint coordinates. This scheme is less sensitive to model uncertainties and require much less computing.

- stability of the command law : classical control laws ignore cable configuration effects that change the pose of the platform. We have proposed a counter-intuitive strategy: the robot MARIONET-ASSIST we are using has a specific geometry that allow to predict which cables set may be under tension for a given trajectory i.e. we are able to split the trajectory in parts for which we know all possible cables configurations. Among them we select the one that optimize an accuracy criteria and we enforce it by forcing the cables not part of the configuration to be slack by adding a sufficient amount of length to their nominal values. It allowed to enhance both the stability and the accuracy of a vision-based control scheme [26].

We also used interval analysis in order to guaranty every step of the process, in order to provide safety and reliability of our methods, as the robots that we use were initially deployed in the context of assistive technologies.

Finally, simulations and experiments on prototypes were conducted and presented in order to validate the mentioned results. However, the prototype that we used presents a very particular configuration (all wires are connected to the same point on the end-effector, allowing only translational movements), further works may be required in order to test our methods for a wider variety of cable-driven parallel robots.

6.2.1.4. Cable-Driven Parallel Robots for additive manufacturing in architecture
Participant: Yves Papegay.

Easy to deploy and to reconfigure, dynamically efficient in large workspaces even with payloads, cable-driven parallel robots are very attractive for solving deplacement and positioning problems in architectural building at scale 1 and seems to be a good alternative to crane and industrial manipulators in this area.

In a collaboration with CNAM and Ecole Nationale Supérieure d'Architecture Paris-Malaquais, we aim to design and realize a CDPR of large size as a proof of concept in additive manufacturing of building based on ultra-high performance concrete.

Challenges are modeling and control to get enough accuracy.

6.2.2. Assistance robotics

This is now the core of our activity and our work on CDPR is deeply connected to this field as they are an efficient solution for mobility assistance, a high priority for the elderly, helpers and medical community. We have presented our vision of assistance robotics in several occasions [20], [21], [19].

6.2.2.1. Assessment of elderly frailty
Participants: Karim Bakal, Jean-Pierre Merlet.

The assessment of elderly frailty is a difficult concept because it involves the physical capacities of a person and its environment (health-care services, families, funds...). We consider the assessment of upper limb capabilities by looking at the joint torques τ of the arm and the maximal force F that can be exerted by the hand, which are related by the equation

$$\tau = \mathbf{J^T} F \tag{3}$$

where $\mathbf{J}$ is a matrix which depends only upon the configuration of the arm. This equations constitutes an underconstrained linear system. In biomechanics the torque τ is measured together with the configuration of the arm and the force F is evaluated by using the method of Chiacchio, that involves the pseudo-inverse of $\mathbf{J^T}$ to calculate F. But there are several uncertainties that are neglected when using this method: the measurement errors on τ and on the configuration of the arm together with uncertainties on the physical parameters of the arm (such as the length of the bones). The method of Chiacchio provides one of the possible solutions of equation (2) and not necessary the one corresponding to the force at the hand. We use another approach based on interval analysis. We assume that all uncertainties may be bounded (τ is an interval vector τ_m, $\mathbf{J^T}$ is an interval matrix) so that equation (2) become an interval linear system. Interval analysis then allows one to determine an approximation as accurate as wanted of the set F_s of all forces F that satisfy the equation and therefore this set includes the real force at the hand. Now assume that with the same arm configuration we measure the force at the hand, here again with some bounded uncertainties (i.e. F is an interval vector F_m). Here again we may use interval analysis applied on equation (2) in order to determine an interval vector τ_v for the τ that is guaranteed to include the real τ. Furthermore τ must be included in the intersection τ_i of τ_v and τ_m while F must be included in the intersection F_i of F_m and F_s. If τ_i is strictly included in τ_m, then we may compute a better approximation of F_s. Reciprocally if F_i is strictly included in F_m we will get a better τ_v. If one of these situation occurs we repeat the process until no significant improvement of F_s or τ_v is obtained. In a second step we consider that the uncertainties that lead to uncertainties in the matrix $\mathbf{J^T}$ are constrained as we have to satisfy $\tau_v = \mathbf{J^T} F_s$. Here again we use interval analysis to determine if this constraint does not allow to reduce the size of the interval on the physical parameters in which case we may obtain a new $\mathbf{J^T}$ that is included in the initial one. In turn this may allow to obtain better τ_v and F_s. The process stops when no improvement has been obtained for F_s, τ_v and the physical parameters.

To test this approach the right upper limb joint torque of 10 males and the force capacity at the right hand was measured by a dynamometer (Biodex III, Biodex Medical Systems) and respectively by a 6-axis load sensor during an experiment performed at HandiBio laboratory. The configuration of the upper limb was measured with a motion capture system (Qualisys, Sweden). The approach is currently being evaluated.

6.2.2.2. Walking analysis
Participants: Claire Maillard, Ting Wang, Jean-Pierre Merlet [correspondant].

The walkers of the `ANG` family allow one to determine accurately the trajectory of the walker and therefore to analyse the walking of the user. We have used this property for performing until mid 2013 a large scale experiment: 23 young adults and 25 elderly people (> 69 years) were asked to walk along with two reference trajectories with the help of the walker. The objective of this research is to develop walking quality index and examine if the walker may be used to monitor the health state of elderly people at home. We compared and statistically analyzed the walking patterns of the two groups of people. The results show that it is possible to obtain new indicators by using the walker measurements [9],[14]. Next step will be to perform a similar analysis for a sit-to-stand (STS) exercice and to test our approach in two rehabilitation centers, MATIA in Spain (in the framework of the RAPP project) and Centre Héliomarin de Vallauris.

A start-up plan was proposed in November 2014 to transfer the walking analysis technology of Hephaistos with the `ANG` walker. In order to study the feasibility of our plan, we have interviewed Patrick Nenert (Kiné, Centre Hélio-Marin), Françoise Dubourgeois (DR, EPHAD) and Sophie Morgenstern (Métropole NCA, Living Lab Paillon 2020) about their impression of the walker and the possibility of the future collaboration with them. Several contact with local actors of the silver economy sector have already been established : Livinglab Paillon2020 (Nice), CIU-santé, as well as with research lab for collaboration on future projects (Lapcos, I3M, Gredeg).

6.2.2.3. Design and evaluation of assistive devices, ethics
Participants: Marc Beninati, Bernard Senach [correspondant], Jean-Pierre Merlet.

Providing appropriate support, services and information to the elderly, to their caregivers and to the medical profession, through a fleet of communicating devices must rely on a structured processes. A generic design and evaluation framework is being elaborated and will be validated through field experiments.

Assistance robotics raises many ethical questions. We started reflection about conducting experiments with frail and old people. A listing of questions to be addressed at each step of an experiment has been written (internal document). We have also hired a joint PhD student with University Bologna about the legal aspects of assistance robotics and we plan to organize a national forum on this topic with Nathalie Nevejans from University of Douai.

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.

Technology demonstrated by our prototype has been transferred to our industrial partner in 2012 when final version of our modeling and simulation environment has been delivered to Airbus in November 2012. Developer level know-how has been transferred in 2013 to a software company in charge of industrialization and maintenance of the modeling and simulation environment.

In 2014, we have worked again on several enhancements and extension of functionalities, namely to enhance the performances and the numerical quality of the generated C simulation code, and ease the integration of our environment into the Airbus toolbox.

LAGADIC Project-Team

6. New Results

6.1. Visual servoing

6.1.1. Photometric moment-based visual servoing

Participants: Manikandan Bakthavatchalam, François Chaumette.

The goal of this work is to determine an adequate set of visual features to control the six degrees of freedom of a dynamic system. Thanks to a collaboration with Omar Tahri from Le2I in Le Creusot, we have been able to improve the results obtained previously with shifted moments for increasing the stability domain of visual servoing [24].

6.1.2. Histogram-based visual servoing

Participants: Quentin Bateux, Eric Marchand.

Classically visual servoing considered the regulation in the image of a set of visual features (usually geometric features). Recently direct visual servoing schemes, such as photometric visual servoing, have been introduced in order to consider the image as a whole and thus avoid the extraction and the tracking of such geometric features. In this preliminary work, we propose a method to extend direct visual servoing approaches by using a global descriptor, namely intensity histograms, on the whole or multiple sub-sets of the images in order to achieve the control of a 6 degrees of freedom (DoF) robot.

6.1.3. Predictive visual servoing

Participants: Nicolas Cazy, Paolo Robuffo Giordano, François Chaumette.

This study is devoted to the application of predictive control to visual servoing. In a first step, we have developed and compared several predictive models that can be useful when some visual features leave the camera field of view or are lost because of occlusions [25].

6.1.4. Visual servoing of cable-driven parallel robot

Participant: François Chaumette.

This study is realized in collaboration with Rémy Ramadour and Jean-Pierre Merlet from Coprin group at Inria Sophia Antipolis. Its goal is to adapt visual servoing techniques for cable-driven parallel robot in order to achieve accurate manipulation tasks [46]. This study is in the scope of the Inria large-scale initiative action PAL (see Section 8.2.6).

6.1.5. Nanomanipulation

Participants: Le Cui, Eric Marchand.

We began a work, within the ANR P2N Nanorobust project (see Section 8.2.1), 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. This year, we focused on the characterisation of the projection model of a SEM along with the approach required for its calibration [26]. We then address the problem of 6 dof control using photometric feature under an optical microscope [27]. Finally, we focused on the definition of control law able to control the motion along the Z axes with a SEM microscope. Indeed, considering that a SEM is subject to parallel projection model, motion along this axis is not observable. We then address this problem using defocus information. An autofocus process has also been studied.

6.1.6. Audio-based control

Participants: Aly Magassouba, François Chaumette.

This study is not concerned with visual servoing, but to the application of the same principle of closed loop control schemes to audio sensors. It is made in collaboration with Nancy Bertin from Panama group at Inria Rennes-Bretagne Atlantique. In a first step, we have determined the analytical form of the interaction matrix of audio features based on the time difference of arrival on two microphones. From this modeling step, we have determined the different virtual linkages that can be realized in function of the number and configuration of sources.

6.2. Visual navigation of mobile robots

6.2.1. Automous navigation of wheelchairs

Participants: Vishnu Karakkat Narayanan, François Pasteau, Marie Babel, François Chaumette.

The goal of this work is to design an autonomous navigation framework of a wheelchair by means of a single camera and visual servoing. We first focused on a corridor following task where no prior knowledge of the environment is required. The servoing process matches the non-holonomic constraints of the wheelchair and relies on two visual features, namely the vanishing point location and the orientation of the median line formed by the straight lines related to the bottom of the walls. This overcomes the initialization issue typically raised in the literature [9]. The control scheme has been implemented onto a robotized wheelchair and results show that it can follow a corridor with an accuracy of ±3 cm [16].

We then focused on a door passing task [33]. This doorway passing (and corridor turning) task employs the position of a single doorpost in the image as an input to a Lyapunov-based control scheme which allows the wheelchair to take up a desired trajectory about the doorpost. This trajectory avoids collision with the wall and guarantees that the wheelchair positions itself in front of the doorway regardless of its initial position. Results in simulation demonstrate the convergence and robustness of both control schemes. Experiments conducted on a wheelchair indicate the validity of applying the proposed low-level control system [16].

This study is in the scope of the Inria large-scale initiative action PAL (see Section 8.2.6) as well as of the Apash project (see Section 8.1.1).

6.2.2. Semi-autonomous control of a wheelchair for navigation assistance along corridors

Participants: Vishnu Karakkat Narayanan, Marie Babel, François Pasteau, Alexandre Krupa.

This study concerns a semi-autonomous control approach that we designed for safe wheelchair navigation. The control relies on the combination of primary tasks of wall avoidance as well as door passing performed by a dedicated visual servoing framework and a manual steering task. A smooth transition from manual driving to assisted navigation is obtained thanks to a gradual visual servoing activation method that guarantees the continuity of the control law. The visual servoing task is then progressively activated, when the wheelchair gets closer to the walls or doorposts, in order to avoid collisions [43]. Experimental results clearly show the ability of the approach to provide an efficient solution for wall avoiding and doorway passing purposes [58]. This study is in the scope of the Inria large-scale initiative action PAL (see Section 8.2.6) as well as of the Apash project (see Section 8.1.1). Tests with disabled patients in the rehabilitation center Pôle Saint Hélier (Rennes) are under progress and first results prove the ability of our system to smoothly correct the trajectory of the wheelchair in case of hazardous situations.

Current research works are oriented towards multimodal sensor-based servoing, as well as haptic feedback that leads to an intuitive assistive wheelchair navigation. This work is realized in collaboration with Maud Marchal (Hybrid team). In addition, we are currently working with e-Motion team to design a vision-based human-aware semi-autonomous navigation system.

6.2.3. Social Spacing and human-robot interaction

Participants: Panagiotis Papadakis, Patrick Rives.

A novel probabilistic framework was introduced capable of instantiating diverse models of social spacing and accounting for distinctive dimensions in human-robot interaction, namely, perception capacity and certainty [42]. We have concretely shown how our method allows smooth adaptation in the situation awareness of a robot within common human-robot interaction examples and further showed its utility at the level of path planning by adapting trajectories to social sensitivity levels.

This approach is currently extended to take into account human activity cues in order to enhance robot mapping and navigation and in particular in filtering noisy human detections, detecting passages such as doors and staircases, inferring space occupancy and allowing navigation within unexplored areas.

6.2.4. Target tracking

Participants: Ivan Markovic, François Chaumette.

This study was realized in the scope of the FP7 Regpot Across project (see Section 8.3.1.2) during the three-month visit of Ivan Markovic, Ph.D. student at the University of Zagreb. It consisted in developing a pedestrian visual tracking from an omni-directional fish-eye camera and a visual servoing control scheme so that a mobile robot is able to follow the pedestrian [37]. This study has been validated on our Pioneer robot (see Section 5.4.2).

6.2.5. Obstacle avoidance

Participants: Suman Raj Bista, 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 [4]. 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 [11]. A new study devoted to indoors navigation from segments has started recently.

6.3. Visual tracking and state estimation

6.3.1. 3D model-based tracking

Participant: Eric Marchand.

This study focused on the issue of estimating the complete 3D pose of the camera with respect to a potentially textureless object, through model-based tracking. We proposed to robustly combine complementary geometrical and color edge-based features in the minimization process, and to integrate a multiple-hypotheses framework in the geometrical edge-based registration phase [45]. This method will be tested in the scope of the FP7 RemoveDebris project [36].

6.3.2. Pose estimation through plane tracking

Participants: Aurélien Yol, Eric Marchand.

We proposed a method for localizing an Unmanned Aerial Vehicle (UAV) using georeferenced aerial images. Here we provide a multiple usage localization algorithm based on vision only. To ensure robustness, we choose to use the Mutual Information (MI) within a dense tracking process. MI proved to be very robust toward local and global scene variations. However, dense approaches are often related to drift disadvantages. We solve this problem by using georeferenced images. The localization algorithm has been demonstrated through the localization of a hexarotor UAV fitted with a downward looking camera during real flight tests [53].

6.3.3. 3D tracking of deformable objects

Participants: Bertrand Delabarre, Eric Marchand.

We consider the problem of dense non-rigid visual tracking robust towards global illumination perturbations of the observed scene. The similarity function is based on the sum of conditional variance (SCV). With respect to most approaches that minimize the sum of squared differences, which is poorly robust towards illumination variations in the scene, the choice of SCV as our registration function allows the approach to be naturally robust towards global perturbations. Moreover, a thin-plate spline warping function is considered in order to take into account deformations of the observed template [28].

6.3.4. *Structure from motion*

Participants: Riccardo Spica, Paolo Robuffo Giordano, François Chaumette.

Structure from motion (SfM) is a classical and well-studied problem in computer and robot vision, and many solutions have been proposed to treat it as a recursive filtering/estimation task. However, the issue of *actively* optimizing the transient response of the SfM estimation error has not received a comparable attention. In the work [18], we showed how to design an online active SfM scheme characterized by an error transient response equivalent to that of a reference linear second-order system with desired poles. Indeed, in a nonlinear context, the observability properties of the states under consideration are not (in general) time-invariant but may depend on the current state and on the current inputs applied to the system. It is then possible to simultaneously act on the estimation gains and system inputs (i.e., the camera velocity for SfM) in order to optimize the observation process and impose a desired transient response to the estimation error. The theory has a general validity and can be applied to many different contexts such as when dealing with point features [18], solid objects like spheres or cylinders [51], or planar regions [47]. Furthermore, the active SfM scheme can also be embedded within a classical visual servoing law exploiting the redundancy of the camera motion w.r.t. the considered visual task [52].

6.3.5. *Robust visual odometry*

Participants: Tawsif Gokhool, Patrick Rives, Renato José Martins.

Our aim is concentrated around building ego-centric topometric maps represented as a graph of salient keyframe nodes [14]. Additionally, visual odometry from frame to keyframe alignment helps significantly in drift reduction. On the other hand, the sparsity in this kind of graphical representation leads to reduced overlapping between keyframes which can degrade localisation robustness. Our chosen spherical 360^0 field of view (FOV) configuration alleviates the overlapping issue by providing an enriched model of the environment with photometric and geometric information content. Following a multitude of advantages with information fusion, merging of frames in a single representation deals with the problem of data redundancy and sensor noise suppression.

Therefore, the second fold of this work consisted in improving the identified conceptual loopholes above by first proposing a generic uncertainty propagation model as applied to our spherical RGB-D database. Secondly, a probabilistic framework was derived which led to a Mahalanobis inconsistency test incorporating both geometric and photometric uncertainty models [32]. Our framework was further improved by adding up a probabilistic model to filter out dynamic points temporally. Finally, the entire probabilistic framework was applied in order to track the most stable points over time.

6.4. 3D Scene Mapping

6.4.1. *New RGB-D sensor design for indoor 3D mapping*

Participants: Eduardo Fernandez Moral, Patrick Rives.

A multi-sensor device has been developed for omnidirectional RGB-D (color+depth) image acquisition (see Fig. 3 .a). This device allows to acquire such omnidirectional images at high frame rates (30 Hz). This approach has advantages over other alternatives used today in terms of accuracy and real-time spherical image construction for indoor environments, which are specially interesting for mobile robotics. This device has important prospective applications as fast 3D-reconstruction or Slam.

A calibration method for such device was developed [31], which takes into account the bias of each sensor independently. The proposed calibration method does not require any specific calibration pattern, taking into account the planar structure from the scene to cope with the fact that there is no overlapping between sensors.

In a first instance, this sensor has been exploited for localization and mapping research with mobile robots. For that, the sensor is mounted on a mobile platform together with an standard computer (see Fig. 3 .a). A method to perform image registration and visual odometry has been developed. This method relies in the matching of planar primitives that can be efficiently obtained from the depth images. This technique performs considerably faster than previous registration approaches like ICP, or dense photoconsistency alignment. These last achieve however a better accuracy than our method, what suggests that our method can be used as an initial step to speed-up those.

Slam is also addressed with this device. A solution to this problem using our omnidirectional RGB-D sensor is being researched. The ongoing experiments have shown some initial results for metric-topological pose-graph Slam, where the map consists of a set of spherical keyframes, which are located in a topological arrangement according to their shared observations.

6.4.2. *Compact 3D scene representation*

Participants: Renato José Martins, Patrick Rives, Tawsif Gokhool.

This work follows in the direction of precise and compact scene representation of large scale environments. The aim is to build a complete geometric and photometric "minimal" model, which is stored within a sparse set of augmented spherical images to asset photo-geometry consistence of the scene from multiple points-of-views. In this direction, an uncertainty model from the full structure combined with those of poses was proposed for point-to-point egocentric fusion. This model allows to reduce sensor noise in a particular keyframe sphere when performing a multi-frame fusion scheme of coherent near information. This first fusion scheme is then improved by exploiting the rigidity/influence of neighboring points representing the surface. For that, an intermediary higher level abstraction of the point cloud is generated by partitioning the input domain into elementary cells, then reducing the number of degrees of freedom and enforcing constraints over the points segmented as being part of the same surface.

The adopted solution is a "weaker" representation of a 3D boundary mesh, based on discontinuous convex planar patches, with the segmentation being done considering the geometry (region growing) or photometry (SLIC superpixels). This synthetic scene built with the planar geometric police proved to well represent the original scene (for both indoor and outdoor real data) with a significant small amount of patches and it is exploited to build robust useful "dynamic" 4D world model, which in turn can be used for assisted/autonomous navigation or virtual reality applications.

6.4.3. *Semantic mapping*

Participants: Romain Drouilly, Patrick Rives, Panagiotis Papadakis.

Autonomous navigation is one of the most challenging problems to address to allow robots to evolve in our everyday environments. Map-based navigation has been studied for a long time and researches have produced a great variety of approaches to model the world. However, semantic information has only recently been taken into account in those models to improve robot efficiency [56]. The goal of this work is to study how semantics can be used to improve all the steps of navigation process. In a first time, we have developed a new navigation-oriented hybrid metric-topological-semantic model of the world. It captures high-level information and uses it to build extremely compact description of large environments. Then we have used it to design an efficient localization algorithm, able to find a given map content faster than classical methods and allowing human-understandable queries [30]. In a second time, we have studied how semantics can be used to discover unobserved things in the scene. Particularly, we have shown that both statics and dynamic entities, identified by a robot, can inform about the structure of the environment in unobserved areas [29] . We have used this to do "map extrapolation", that is extending a map beyond robot's perceptual limits by reasoning on semantics. This approach has been shown to be of great interest in everyday-life environment. Finally, we have proposed a new scheme for trajectory planing, taking into account not only geometric constraints

but also high-level understanding of the world. We have shown the usefulness of this approach to navigate complex environments with highly dynamic areas on both simulated and real-world datasets, well-suited for large outdoor environment navigation.

6.4.4. *Augmented reality*
Participant: Eric Marchand.

Using Slam methods becomes more and more common in Augmented Reality (AR). To achieve real-time requirement and to cope with scale factor and the lack of absolute positioning issue, we proposed to decouple the localization and the mapping step. This approach has been validated on an Android Smartphone through a collaboration with Orange Labs [38][39]

6.5. Medical robotics

6.5.1. *Non-rigid target tracking in ultrasound images based on hierarchical grid interpolation*
Participants: Lucas Royer, Jason Chevrie, Marie Babel, Alexandre Krupa.

In order to track the motion of a tumour or cyst during needle insertion, we developed a first approach to track a deformable target within a sequence of 2D ultrasound images. It is based on a dedicated hierarchical grid interpolation algorithm (HGI) that is typically used for real-time video compression purposes. This approach provides a continuous motion representation of the target by using a grid of control points that models both their global displacement and local deformations. The motion of each control point is estimated by a hierarchical and multi-resolution local search method in order to minimize the sum of squared difference of the target pixel intensity between successive images. This approach was validated from 2D ultrasound images of real human tissues undergoing rigid and non-rigid deformations [48] and was recently adapted for tracking 3D deformations.

6.5.2. *Non-rigid target tracking in ultrasound images based on physically-based model*
Participants: Lucas Royer, Alexandre Krupa.

A second approach for automatically tracking deformable target within 2D ultrasound images has been developed [50]. It combines dense information with a physically-based model and has therefore the advantage of not using any fiducial marker. The physical model is represented by a mass-spring damper system driven by external and internal forces. The external forces are obtained by maximizing an image similarity metric between a reference target and the deformed target along the time. The internal forces of the mass-spring damper system constrain the deformation to be physically plausible and therefore efficiently reduce the sensitivity to the speckle noise. This second approach was validated on simulated and real data, both for rigid and non-rigid motions of soft tissues [49]. It was recently extended for deformable target tracking in 3D ultrasound volumes.

6.5.3. *3D steering of flexible needle by visual servoing*
Participants: Alexandre Krupa, Pierre Chatelain.

The objective of this work is to provide robotic assistance during needle insertion procedures such as biopsy or ablation of localized tumor. A method has been developed for steering a beveled-tip flexible needle actuated by a robotic arm in such a way to control the needle curvature in 3D space [34]. It is based on the design of a new duty-cycling control strategy that makes possible to control both the 2 lateral angular velocities and the insertion velocity of the needle tip (3 DOF). An image-based visual servoing approach has then been developed to automatically position the needle tip on a 3D target indicated by the user. It is based on the use of geometrical visual features extracted from 2 images provided by 2 orthogonal cameras observing a translucent gelatin phantom where the needle is inserted. Preliminary results of this automatic targeting task demonstrate the feasibility of this new concept and its robustness to needle kinematic model errors [35]. We recently extended this approach to automatically steer the needle toward a target by an image-based visual servoing that uses geometrical features extracted from images provided by a 3D ultrasound probe.

6.5.4. *Optimization of ultrasound image quality by visual servoing*
Participants: Pierre Chatelain, Alexandre Krupa.

This study focuses on the automatic positioning of a 2D ultrasound probe in such a way to optimize the quality of the acquired ultrasound images. It is based on the recent framework of ultrasound confidence map, developed in the Chair for Computer Aided Medical Procedures and Augmented Reality of Prof. Nassir Navab, which aims at estimating the per-pixel quality of the ultrasound signal based on a model of sound propagation in soft tissues. In collaboration with Nassir Navab we considered this ultrasound confidence map as a new modality and recently designed a visual servoing control law for image quality optimization.

6.6. Control of single and multiple Unmanned Aerial Vehicles

6.6.1. *State estimation and flight control of quadrotor UAVs*
Participant: Paolo Robuffo Giordano.

Over the last years the robotics community witnessed an increasing interest in the Unmanned Aerial Vehicle (UAV) field. In particular quadrotor UAVs have become more and more widespread in the community as experimental platform for, e.g., testing novel 3D planning, control and estimation schemes in real-world indoor and outdoor conditions. Indeed, in addition to being able to take-off and land vertically, quadrotors can reach high angular accelerations thanks to the relatively long lever arm between opposing motors. This makes them more agile than most standard helicopters or similar rotorcraft UAVs, and thus very suitable to realize complex tasks such as aerial mapping, air pollution monitoring, traffic management, inspection of damaged buildings and dangerous sites, as well as agricultural applications such as pesticide spraying.

Despite these clear advantages, a clear shortcoming of the quadrotor design lies in its inherent underactuation (only 4 actuated propellers for the 6 dofs of the quadrotor pose). This underactuation limits the quadrotor flying ability in free or cluttered space and, furthermore, it also degrades the possibility of interacting with the environment by exerting desired forces in arbitrary directions. In [17], a novel design for a quadrotor UAV with tilting propellers which is able to overcome these limitations has been presented and experimentally validated. Indeed, the additional set of 4 control inputs actuating the propeller tilting angles can be shown to yield full actuation to the quadrotor position/orientation in space, thus allowing it to behave as a fully-actuated flying vehicle and to overcome the aforementioned underactuation problem.

This work has been realized in collaboration with the Max Planck Institute for Biological Cybernetics, Tübingen, Germany.

6.6.2. *Collective control of multiple UAVs*
Participant: Paolo Robuffo Giordano.

The challenge of coordinating the actions of multiple robots is inspired by the idea that proper coordination of many simple robots can lead to the fulfilment of arbitrarily complex tasks in a robust (to single robot failures) and highly flexible way. Teams of multi-robots can take advantage of their number to perform, for example, complex manipulation and assembly tasks, or to obtain rich spatial awareness by suitably distributing themselves in the environment. Within the scope of robotics, autonomous search and rescue, firefighting, exploration and intervention in dangerous or inaccessible areas are the most promising applications.

In the context of multi-robot (and multi-UAV) coordinated control, *connectivity* of the underlying graph is perhaps the most fundamental requirement in order to allow a group of robots accomplishing common goals by means of *decentralized* solutions. In fact, graph connectivity ensures the needed continuity in the data flow among all the robots in the group which, over time, makes it possible to share and distribute the needed information. However, connectivity alone is not sufficient to perform certain tasks when only *relative sensing* is used. For these systems, the concept of *rigidity* provides the correct framework for defining an appropriate sensing and communication topology architecture. Rigidity is a combinatorial theory for characterizing the "stiffness" or "flexibility" of structures formed by rigid bodies connected by flexible linkages or hinges. In a broader context, rigidity turns out to be an important architectural property of many multi-agent systems when

a common inertial reference frame is unavailable. Applications that rely on sensor fusion for localization, exploration, mapping and cooperative tracking of a target, all can benefit from notions in rigidity theory. The concept of rigidity, therefore, provides the theoretical foundation for approaching decentralized solutions to the aforementioned problems using distance measurement sensors, and thus establishing an appropriate framework for relating system level architectural requirements to the sensing and communication capabilities of the system.

In [22], a decentralized gradient-based rigidity maintenance action for a group of quadrotor UAVs has been proposed and tested in real experimental conditions. By starting in a rigid configuration, the group of UAVs is able to estimate their relative position from sole relative distance measurements, and then use these estimated relative positions in a control action able to preserve rigidity of the whole formation despite presence of sensor limitations (maximum range and line-of-sight occlusions), possible collisions with obstacles and inter-robot collisions. Furthermore, in [54] the rigidity theory has been extended to the case of *bearing measurements*, and directed graphs.

These works were realized in collaboration with the robotics group at the Max Planck Institute for Biological Cybernetics, Tübingen, Germany and with Technion, Israel.

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

6. New Results

6.1. Highlights of the Year

Our work on sketch-based modeling for product designers (Sec. 6.4.4) has received significant attention. It appeared on the news page of University of British Columbia http://news.ubc.ca/2014/08/13/powerful-math-creates-3-d-shapes-from-simple-sketches/ and our video has been watched more than 7000 times on Youtube http://youtu.be/tbUljHJv4Rg. We filed a patent on this technology and we have contacts with several companies about a potential transfer.

Our poster on *C-LOD: Context-aware Material Level-of-Detail applied to Mobile Graphics* [] received the 3rd place in the ACM's Graduate Student Research Competition at SIGGRAPH 2014. This work is a collaboration with George Alex Koulieris and Katerina Mania from the Technical University of Crete and Douglas Cunningham from the Technical University of Cottbus.
BEST PAPER AWARD :

[] **Computer Graphics Forum**. G. A. KOULIERIS, G. DRETTAKIS, D. W. CUNNINGHAM, K. MANIA.

6.2. Plausible and Realistic Image Rendering

6.2.1. *Multi-View Intrinsic Images for Outdoors Scenes with an Application to Relighting*

Participants: Sylvain Duchêne, Clement Riant, Gaurav Chaurasia, Stefan Popov, Adrien Bousseau, George Drettakis.

We introduce a method to compute intrinsic images for a multi-view set of outdoor photos with cast shadows, taken under the same lighting. We use an automatic 3D reconstruction from these photos and the sun direction as input and decompose each image into reflectance and shading layers, despite the inaccuracies and missing data of the 3D model. Our approach is based on two key ideas. First, we progressively improve the accuracy of the parameters of our image formation model by performing iterative estimation and combining 3D lighting simulation with 2D image optimization methods. Second we use the image formation model to express reflectance as a function of discrete visibility values for shadow and light, which allows us to introduce a robust visibility classifier for pairs of points in a scene. This classifier is used for shadow labeling, allowing us to compute high quality reflectance and shading layers. We then create shadow-caster geometry that preserves shadow silhouettes. Combined with the intrinsic layers, this approach allows multi-view relighting with moving cast shadows. We present results on several multi-view datasets, and show how it is now possible to perform image-based rendering with changing illumination conditions.

This work is part of an industrial partnership with Autodesk and is under revision for ACM Transactions On Graphics.

6.2.2. *Compiler and Tiling Strategies for IIR Filters*

Participants: Gaurav Chaurasia, George Drettakis.

We present a compiler for parallelizing IIR or recursive filters. IIR filters are frequently used for $O(1)$ convolutions, but they cannot exploit GPUs because they are very hard to parallelize and also exhibit poor memory locality which hinders performance on both CPUs and GPUs. We present algorithmic tiling strategies for IIR filters which overcome these limitations. Tiled IIR filters are notoriously hard to implement and hence largely ignored by programmers and hardware vendors. We present a compiler front-end that supports intuitive functional specification and tiling of IIR filters. We demonstrate that different tiling strategies may be optimal on different platforms and filter parameters; our compiler can express the exhaustive set of alternatives in just 10-20 lines of code. This enables programmers to easily explore a large variety of trade-offs at different levels of granularity, thereby making it easier and more likely to discover the optimal implementation, while also producing intuitive and maintainable code. Our initial results show that our compiler is as terse as vendor provided libraries, but it allows exploiting the algorithmic advantages of tiling which cannot be provided by any precompiled library.

For example, our compiler can compute a nearly 8 times faster summed area table (4096×4096 image) in 20 lines of code including a fully customized CUDA schedule, as compared to 10 lines in NVIDIA Thrust which does not allow tiling or customizing the CUDA schedule.

This ongoing work is a collaboration with Jonathan Ragan-Kelley (Stanford University), Sylvain Paris (Adobe) and Fredo Durand (MIT).

6.2.3. Video based rendering

Participants: Abdelaziz Djelouah, George Drettakis.

In this project our objective is to propose a new algorithm for novel view synthesis in the case of dynamic scene. The main difference compared to static image-based rendering is the limited number of viewpoints and the presence of the extra time dimension. In a configuration where the number of cameras is limited, segmentation becomes crucial to identify moving foreground regions. To facilitate the difficult task of multi-view segmentation, we currently target scenes captured with stereo cameras. Stereo pairs provide important information on the geometry of the scene while simplifying the segmentation problem.

This ongoing work is a collaboration with Gabriel Brostow from University College London in the context of the CR-PLAY EU project.

6.2.4. Temporally Coherent Video De-Anaglyph

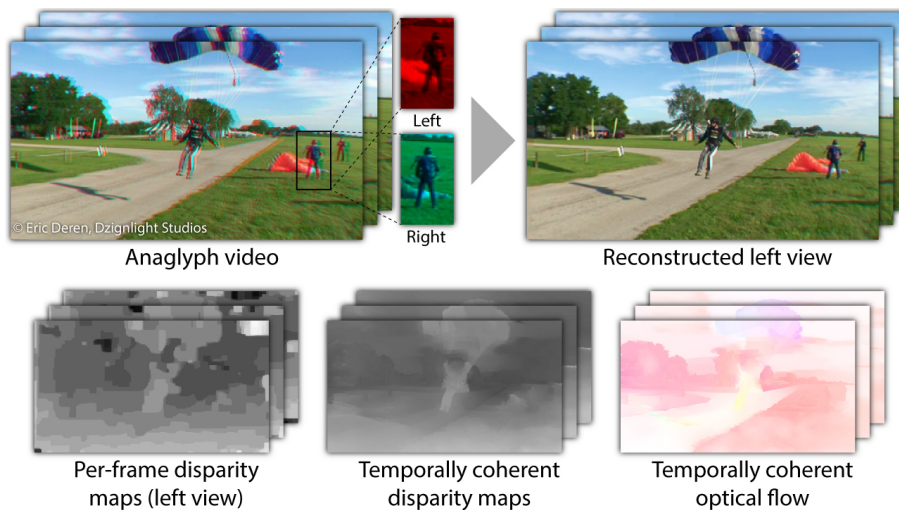**Participants:** Joan Sol Roo, Christian Richardt.

Figure 4. Top: We convert anaglyph videos (left) to temporally coherent full-color stereo videos (right). Bottom: Our approach starts with rough, per-frame disparity maps (left) and produces temporally coherent disparity maps and optical flow (center and right) that are used for reconstructing the stereo views.

This work investigates how to convert existing anaglyph videos to the full-color stereo format used by modern displays. Anaglyph videos only contain half the color information compared to the full-color videos, and the missing color channels need to be reconstructed from the existing ones in a plausible and temporally coherent fashion. In our approach, we put the temporal coherence of the stereo video results front and center (see Figure 4). As a result, our approach is both efficient and temporally coherent. In addition, it computes temporally coherent optical flow and disparity maps that can be used for various post-processing tasks. As a practical contribution, we also make the source code of our implementation available online under CeCILL-B license.

This work was carried out by Joan Sol Roo during his internship in the summer of 2013. The work was presented as a talk and poster at SIGGRAPH 2014 [19].

6.2.5. *Probabilistic Connection Path Tracing*

Participants: George Drettakis, Stefan Popov.

Bi-directional path tracing (BDPT) with Multiple Importance Sampling (MIS) is one of the most versatile unbiased rendering algorithms today. BDPT repeatedly generates sub-paths from the eye and the lights, which are connected for each pixel and then discarded. Unfortunately, many such bidirectional connections turn out to have low contribution to the solution. The key observation in this project, is that we can find better connections to an eye sub-path by considering multiple light sub-paths at once and creating connections probabilistically only with the most promising ones. We do this by storing light paths, and estimating probability density functions (PDF) of the discrete set of possible connections to all light paths. This has two key advantages: we efficiently create connections with high quality contributions by Monte Carlo sampling, and we reuse light paths across different eye paths. We also introduce a caching scheme for PDFs by deriving a low-dimensional approximation to sub-path contribution.

This ongoing work is a collaboration with Fredo Durand from MIT and Ravi Ramamoorthi from the University of California San Diego in the context of the CRISP associate team.

6.2.6. *Unified Color and Texture Transfer for By-Example Scene Editing*

Participants: Fumio Okura, Kenneth Vanhoey, Adrien Bousseau, George Drettakis.

Color and texture transfer methods are at the heart of by-example image editing techniques. Color transfer well represents the change of overall scene appearance; however it does not represent the change of texture and shape. On the other hand, by-example texture transfer expresses the texture change but it often destroys the target scene structure. We seek the best combination of by-example color and texture transfer to combine these transfer methods so as to selectively work where each method is suitable. Given the source and exemplar pair, the proposed algorithm learns local error metrics which describe if local change between the source and exemplar is best expressed by color or texture transfer. The metric provides us with a local prediction of where we need to synthesize textures using a texture transfer method. This work is a collaboration with Alexei Efros from UC Berkeley in the context of the associate team CRISP.

6.2.7. *Improved Image-Based Rendering*

Participants: Rodrigo Ortiz Cayon, Abdelaziz Djelouah, George Drettakis.

Image-based rendering algorithms based on warping present strong artifacts when rendering surfaces at grazing angle. We are working on a new IBR algorithm that overcomes this problem by rendering superpixel segments as piece-wise homography transformations. The input to our method is a set of images calibrated and a 3D point cloud generated from multi-view stereo reconstruction. In pre-processing we robustly fit planes to superpixel segments that contain reconstruction information and then propagate plausible depth and normal information for image-based rendering. Novel views are obtained by re-projecting superpixel segments as homography from different input views, then adaptively blending them according to distortion and confidence estimations.

6.2.8. *Structured Procedural Textures*

Participants: Kenneth Vanhoey, George Drettakis.

Textures form a popular tool to add visual detail to shapes, objects and scenes. Manual texture design is however a time-consuming process. An alternative is to generate textures from an input exemplar (*i.e.* an acquired photograph) automatically. The difficulty is to synthesize textures of arbitrary size from a single input, preferably with no repetition artifacts. State of the art synthesis techniques can be categorized in two: copy-based techniques and procedural noise-based ones. The first copy pixels using iterative algorithms. The latter deduce a continuous mathematical function from the exemplar, and evaluate it on the space to be textured. They have the advantage of continuity (no resolution-dependence, minimized memory storage, etc.) and fast

local evaluation suitable for parallel GPU implementation. They are however tedious to define and manipulate. Current state of the art methods are limited to reproducing Gaussian patterns, that is, textures with no or few structure.

We investigate how to go beyond this limit. Noise-based methods constrain the Fourier power spectrum of a texture-generating noise function to resemble the spectrum of the exemplar. By also constraining the phase of the Fourier spectrum to resemble the exemplar, an exact reproduction is obtained, thus lacking variety and showing maximal repetition. By randomizing the phases, an unstructured "same-looking" image is obtained. This is suitable for noise-like patterns (*e.g.*, marble, wood veins, sand) but not for structured ones (*e.g.*, brick wall, mountain rocks, woven yarn).

In this project, we proceed by investigating the phase spectrum of an image. It contains the structure but identifying how and where is difficult. To characterize structure, we will exploit the splatting process of local random-phase noise and exhibit possible correlations between local phases and spatial placement.

This ongoing work is a collaboration with Ian Jermyn from Durham University.

6.3. Perception for Plausible Rendering

6.3.1. *An Automated High Level Saliency Predictor for Smart Game Balancing*
Participant: George Drettakis.

Successfully predicting visual attention can significantly improve many aspects of computer graphics: scene design, interactivity and rendering. Most previous attention models are mainly based on low-level image features, and fail to take into account high level factors such as scene context, topology, or task. Low-level saliency has previously been combined with task maps, but only for predetermined tasks. Thus, the application of these methods to graphics (e.g., for selective rendering) has not achieved its full potential.

In this work, we present the first automated high-level saliency predictor incorporating two hypotheses from perception and cognitive science that can be adapted to different tasks. The first states that a scene is comprised of objects expected to be found in a specific context as well objects out of context which are salient (scene schemata) while the other claims that viewer's attention is captured by isolated objects (singletons). We proposed a new model of attention by extending Eckstein's Differential Weighting Model. We conducted a formal eye-tracking experiment which confirmed that object saliency guides attention to specific objects in a game scene and determined appropriate parameters for a model. We presented a GPU-based system architecture that estimates the probabilities of objects to be attended in real-time (Figure 5). We embedded this tool in a game level editor to automatically adjust game level difficulty based on object saliency, offering a novel way to facilitate game design. We perform a study confirming that game level completion time depends on object topology as predicted by our system.

This work is a collaboration with George Alex Koulieris and Katerina Mania from the Technical University of Crete and Douglas Cunningham from the Technical University of Cottbus. The work was published in the ACM Transactions on Applied Perception (TAP) Journal [15] and presented as a Talk at SIGGRAPH 2014 in Vancouver.

6.3.2. *C-LOD: Context-aware Material Level-of-Detail applied to Mobile Graphics*
Participant: George Drettakis.

Attention-based Level-Of-Detail (LOD) managers downgrade the quality of areas that are expected to go unnoticed by an observer to economize on computational resources. The perceptibility of lowered visual fidelity is determined by the accuracy of the attention model that assigns quality levels. Most previous attention based LOD managers do not take into account saliency provoked by context, failing to provide consistently accurate attention predictions.

Figure 5. A low level saliency algorithm indicates that the most salient area of the image is the dark area behind the chair. Our tool highlights the vase at a consistent/singleton location as the most salient object in the image.

In this work, we extended a recent high level saliency model with four additional components yielding more accurate predictions: an object-intrinsic factor accounting for canonical form of objects, an object-context factor for contextual isolation of objects, a feature uniqueness term that accounts for the number of salient features in an image, and a temporal context that generates recurring fixations for objects inconsistent with the context. We conducted a perceptual experiment to acquire the weighting factors to initialize our model. We then designed C-LOD, a LOD manager that maintains a constant frame rate on mobile devices by dynamically re-adjusting material quality on secondary visual features of non-attended objects. In a proof of concept study we established that by incorporating C-LOD, complex effects such as parallax occlusion mapping usually omitted in mobile devices can now be employed, without overloading GPU capability and, at the same time, conserving battery power. We validated our work via eye-tracking (Figure 6)
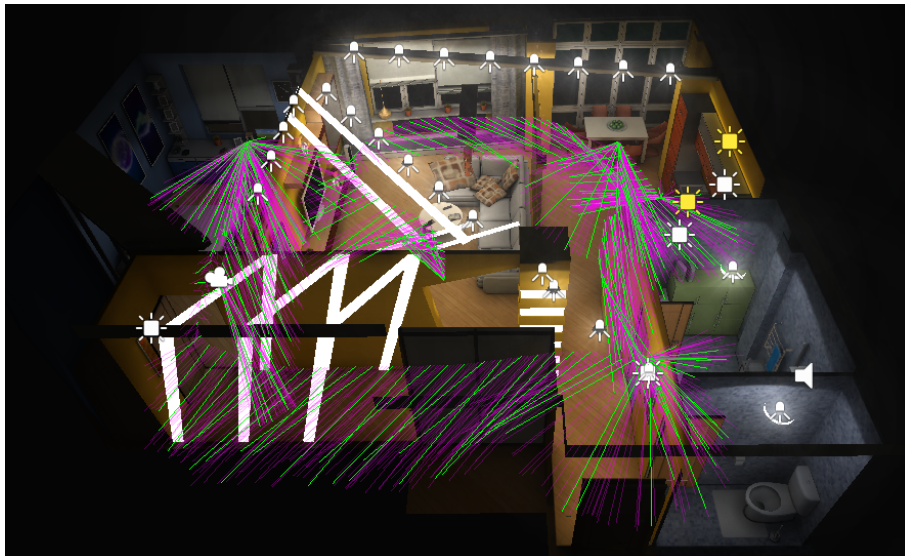

Figure 6. Our validation tool indicates the subject's gaze point with magenta colored beams. The green beams indicate predictions by our attention model.

This work is a collaboration with George Alex Koulieris and Katerina Mania from the Technical University of Crete and Douglas Cunningham from the Technical University of Cottbus. The work was published in a special issue of Computer Graphics Forum [] and was presented at the Eurographics Symposium on Rendering 2014 in Lyon. It was also presented as a poster at SIGGRAPH 2014 in Vancouver winning the 3rd place in the ACM's Graduate Student Research Competition.

6.4. Interaction and Design for Virtual Environments

6.4.1. *Evaluation of Direct Manipulation using Finger Tracking for Complex Tasks in an Immersive Cube*

Participants: Emmanuelle Chapoulie, George Drettakis.

We present a solution for interaction using finger tracking in a cubic immersive virtual reality system (or immersive cube). Rather than using a traditional flystick device, users can manipulate objects with fingers of both hands in a close-to-natural manner for moderately complex, general purpose tasks. Our solution couples finger tracking with a real-time physics engine, combined with a heuristic approach for hand manipulation, which is robust to tracker noise and simulation instabilities. We performed a first study to evaluate our interface with tasks involving complex manipulations, such as balancing objects while walking in the cube. The users finger-tracked manipulation was compared to manipulation with a 6 degree-of-freedom flystick, as well as with carrying out the same task in the real world. Users were also asked to perform a free task, allowing us to observe their perceived level of presence in the scene. Our results showed that our approach provides a feasible interface for immersive cube environments and is perceived by users as being closer to the real experience compared to the flystick. However, the flystick outperforms direct manipulation in terms of speed and precision.

This work is a collaboration with Maria Roussou and Evanthia Dimara from the University of Athens, Maud Marchal from Inria Rennes, and Jean-Christophe Lombardo from Inria Sophia Antipolis. The work has been published in the journal Virtual Reality [13].

Figure 7. A user balancing a tray with both hands [13].

We have also worked on a followup study in which we examine a much more controlled context, studying only very limited movements, in 1D, 2D and 3D. To do this we designed specific devices that can be instantiated both in the virtual world and as physical objects. We compared finger manipulation to wand and to real configurations; the study demonstrated the feasibility of such a controlled comparison for the study of finger-based interaction. This work is in collaboration with InSitu, specifically F. Tsandilas, W. Mackay and L. Oehlberg, and has been accepted for publication in 2015 at IEEE 3DUI.

6.4.2. *Reminiscence Therapy using Image-Based Rendering in VR*

Participants: Emmanuelle Chapoulie, George Drettakis, Rachid Guerchouche, Gaurav Chaurasia.

We present a novel VR solution for Reminiscence Therapy (RT), developed jointly by a group of memory clinicians and computer scientists. RT involves the discussion of past activities, events or experiences with others, often with the aid of tangible props which are familiar items from the past; it is a popular intervention in dementia care. We introduced an immersive VR system designed for RT, which allows easy presentation of familiar environments. In particular, our system supports highly-realistic Image-Based Rendering in an immersive setting. To evaluate the effectiveness and utility of our system for RT, we performed a study with healthy elderly participants to test if our VR system could help with the generation of autobiographical memories. We adapted a verbal Autobiographical Fluency protocol to our VR context, in which elderly participants were asked to generate memories based on images they were shown. We compared the use of our image-based system for an unknown and a familiar environment. The results of our study showed that the number of memories generated for a familiar environment is higher than the number of memories obtained for an unknown environment using our system. This indicates that IBR can convey familiarity of a given scene, which is an essential requirement for the use of VR in RT. Our results also showed that our system is as effective as traditional RT protocols, while acceptability and motivation scores demonstrated that our system is well tolerated by elderly participants.

This work is a collaboration with Pierre-David Petit and Philippe Robert from the CMRR in Nice. The work has been published in the Proceedings of IEEE Virtual Reality [18].

Figure 8. Left: our hardware setup. Right: new point of view reconstructed from input cameras.

6.4.3. *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. We conducted a thorough

study to evaluate different lightfield editing interfaces, tools and workflows from a user perspective. We additionally investigate the potential benefits of using depth information when editing, and the limitations imposed by imperfect depth reconstruction using current techniques. We perform two different experiments, collecting both objective and subjective data from a varied number of point-based editing tasks of increasing complexity: In the first experiment, we rely on perfect depth from synthetic lightfields, and focus on simple edits. This allows us to gain basic insight on lightfield editing, and to design a more advanced editing interface. This is then used in the second experiment, employing real lightfields with imperfect reconstructed depth, and covering more advanced editing tasks. Our study shows that users can edit lightfields with our tested interface and tools, even in the presence of imperfect depth. They follow different workflows depending on the task at hand, mostly relying on a combination of different depth cues. Last, we confirm our findings by asking a set of artists to freely edit both real and synthetic lightfields.

This work is a collaboration with Adrian Jarabo, Belen Masia and Diego Gutierrez from Universidad de Zaragoza and Fabio Pellacini from Sapienza Universita di Roma. This work was published at ACM Transactions on Graphics 2014 (Proc. SIGGRAPH) [14].
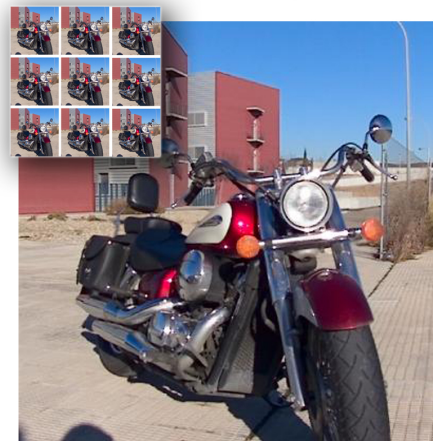

Figure 9. A lightfields represents multiple nearby views of a scene. We conducted a study to evaluate how people edit such data.

6.4.4. True2Form: 3D Curve Networks from 2D Sketches via Selective Regularization

Participant: Adrien Bousseau.

True2Form is a sketch-based modeling system that reconstructs 3D curves from typical design sketches. Our approach to infer 3D form from 2D drawings is a novel mathematical framework of insights derived from perception and design literature. We note that designers favor viewpoints that maximally reveal 3D shape information, and strategically sketch descriptive curves that convey intrinsic shape properties, such as curvature, symmetry, or parallelism. Studies indicate that viewers apply these properties selectively to envision a globally consistent 3D shape. We mimic this selective regularization algorithmically, by progressively detecting and enforcing applicable properties, accounting for their global impact on an evolving 3D curve network. Balancing regularity enforcement against sketch fidelity at each step allows us to correct for inaccuracy inherent in free-hand sketching. We perceptually validate our approach by showing agreement between our algorithm and viewers in selecting applicable regularities. We further evaluate our solution by: reconstructing a range of 3D models from diversely sourced sketches; comparisons to prior art; and visual comparison to both ground-truth and 3D reconstructions by designers.
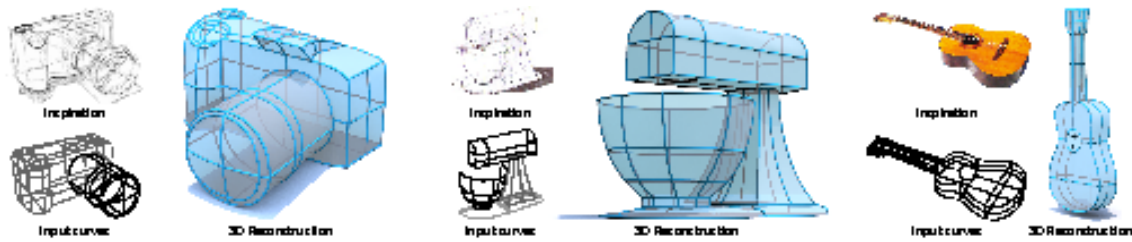
Figure 10. Our single-view modeling system allows us to reconstruct 3D models by tracing curves over existing sketches and photographs.

This work is a collaboration with James McCrae and Karan Singh from the University of Toronto and Xu Baoxuan, Will Chang and Alla Sheffer from the University of British Columbia. The paper was published at ACM Transactions on Graphics 2014 (Proc. SIGGRAPH) [17].

6.4.5. *BendFields: Regularized Curvature Fields from Rough Concept Sketches*

Participants: Adrien Bousseau, Emmanuel Iarussi.

Designers frequently draw curvature lines to convey bending of smooth surfaces in concept sketches. We present a method to extrapolate curvature lines in a rough concept sketch, recovering the intended 3D curvature field and surface normal at each pixel of the sketch. This 3D information allows us to enrich the sketch with 3D-looking shading and texturing. We first introduce the concept of *regularized curvature lines* that model the lines designers draw over curved surfaces, encompassing curvature lines and their extension as geodesics over flat or umbilical regions. We build on this concept to define the orthogonal cross field that assigns two regularized curvature lines to each point of a 3D surface. Our algorithm first estimates the projection of this cross field in the drawing, which is non-orthogonal due to foreshortening. We formulate this estimation as a scattered interpolation of the strokes drawn in the sketch, which makes our method robust to sketchy lines that are typical for design sketches. Our interpolation relies on a novel smoothness energy that we derive from our definition of regularized curvature lines. Optimizing this energy subject to the stroke constraints produces a dense non-orthogonal 2D cross field, which we then lift to 3D by imposing orthogonality. Thus, one central concept of our approach is the generalization of existing cross field algorithms to the non-orthogonal case. We demonstrate our algorithm on a variety of concept sketches with various levels of sketchiness. We also compare our approach with existing work that takes clean vector drawings as input.

This work is a collaboration with David Bommes from Titane project team, Inria Sophia-Antipolis. The manuscript has been accepted for publication with minor revisions at ACM Transactions on Graphics (TOG).

6.4.6. *Line Drawing Interpretation in a Multi-View Context*

Participant: Adrien Bousseau.

Many design tasks involve the creation of new objects in the context of an existing scene. Existing work in computer vision only provides partial support for such tasks. On the one hand, multi-view stereo algorithms allow the reconstruction of real-world scenes, while on the other hand algorithms for line-drawing interpretation do not take context into account. This work combines the strength of these two domains to interpret line drawings of imaginary objects drawn over photographs of an existing scene. The main challenge we face is to identify the existing 3D structure that correlates with the line drawing while also allowing the creation of new structure that is not present in the real world. We propose a labeling algorithm to tackle this problem, where some of the labels capture dominant orientations of the real scene while a free label allows the discovery of new orientations in the imaginary scene.

This work is a collaboration with Jean-Dominique Favreau and Florent Lafarge from Titane project team, Inria Sophia-Antipolis and is under submission for the CVPR conference.

6.4.7. *Wrap It! Computer-Assisted Design and Fabrication of Wire Wrapped Jewelry*
Participants: Adrien Bousseau, Emmanuel Iarussi.

We developed an interactive tool to assist the process of creating and crafting wire wrapped pieces of jewelry. In a first step, we guide the user in conceiving designs which are suitable to be fabricated with metal wire. In a second step, we assist fabrication by taking inspiration from jigs-based techniques, frequently used by craftsmen as a way to guide and support the wrapping process. Given a vector drawing composed of curves to be fabricated, it is crucial to first decompose it into segments that can be constructed with metal wire. Literature on jewelry-making provides a wide range of examples to perform this task, but they are hard to generalize to any input design. Based on the observation of these examples, we distill and generalize a set of design principles behind the finished pieces of jewelry. Relying on those principles, we propose an algorithm that generates a decomposition of the input where each piece is a single component of wire, that can be wrapped and gathered with the others. In addition, we also automate the design of custom physical jigs for fabrication of the jewelry piece. A jig consists of a board with holes on it, arranged in a regular grid structure. By placing a set of pins (of different radius) on the jig, the craftman builds a support structure that guides the wrapping process. The wire is bended and twisted around those pins to create the shape. Given the input design curves and the available jig parameters (size, number and radius of the pins), we propose an algorithm to automatically generate an arrangement of pins in order to better approximate the input curve with wire. Finally, users can follow automatically-generated step-by-step instructions to place the pins in the jig board and fabricate the end piece of jewelry.

This ongoing work is a collaboration with Wilmot Li from Adobe, San Francisco. The project was initiated by a 3-months visit of Emmanuel Iarussi at Adobe.

6.4.8. *Studying how novice designers communicate with sketches and prototypes*
Participant: Adrien Bousseau.

We performed a user study to better understand how novice designers communicate a concept during the different phases of its development. Our study was conducted as a one-day design contest where participants had to propose a concept, present it to a jury, describe it to an engineer and finally fabricate a prototype with the help of another participant. We collected sketches and videos for all steps of this exercise in order to evaluate how the concept evolves and how it is described to different audiences. We hope that our findings will inform the development of better computer-assisted design tools for novices.

This is an ongoing work in collaboration with Wendy McKay, Theophanis Tsandilas and Lora Oehlberg from the InSitu project team - Inria Saclay, in the context of the ANR DRAO project.

6.4.9. *Vectorising Bitmaps into Semi-Transparent Gradient Layers*
Participants: Christian Richardt, Adrien Bousseau, George Drettakis.

Vector artists create complex artworks by stacking simple layers. We demonstrate the benefit of this strategy for image vectorisation, and present an interactive approach for decompositing bitmap drawings and studio photographs into opaque and semi-transparent vector layers. Semi-transparent layers are especially challenging to extract, since they require the inversion of the non-linear compositing equation. We make this problem tractable by exploiting the parametric nature of vector gradients, jointly separating and vectorising semi-transparent regions. Specifically, we constrain the foreground colours to vary according to linear or radial parametric gradients, restricting the number of unknowns and allowing our system to efficiently solve for an editable semi-transparent foreground.

(a) Input photograph *(b) Multi-layer vector representation* *(c) Editing result*

Figure 11. Our interactive vectorisation technique lets users vectorise an input bitmap (a) into a stack of opaque and semi-transparent vector layers composed of linear or radial colour gradients (b). Users can manipulate the resulting layers using standard tools to quickly produce new looks (c). Semi-transparent layers are outlined for visualisation; these edges are not part of our result.

We propose a progressive workflow, where the user successively selects a semi-transparent or opaque region in the bitmap, which our algorithm separates as a foreground vector gradient and a background bitmap layer. The user can choose to decompose the background further or vectorise it as an opaque layer. The resulting layered vector representation allows a variety of edits, as illustrated in Figure 11 , such as modifying the shape of highlights, adding texture to an object or changing its diffuse colour. Our approach facilitates the creation of such layered vector graphics from bitmaps, and we thus see our method as a valuable tool for professional artists and novice users alike.

This work is a collaboration with Jorge Lopez-Moreno, now a postdoc at the University of Madrid, and Maneesh Agrawala from the University of California, Berkeley in the context of the CRISP Associated Team. The paper was presented at the Eurographics Symposium on Rendering (EGSR) 2014, and is published in a special issue of the journal Computer Graphics Forum [16].

<div align="center">

STARS Project-Team

</div>

6. New Results

6.1. Highlights of the Year

NeoSensys, a spin off of the Stars team which aims at commercializing video surveillance solutions for the retail domain, has been created in September 2014.

6.2. 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.

6.2.1. *Perception for Activity Recognition*

Participants: Julien Badie, Slawomir Bak, Piotr Bilinski, François Brémond, Bernard Boulay, Guillaume Charpiat, Duc Phu Chau, Etienne Corvée, Carolina Garate, Michal Koperski, Ratnesh Kumar, Filipe Martins, Malik Souded, Anh Tuan Nghiem, Sofia Zaidenberg, Monique Thonnat.

For perception, the main achievements are:

- Our new covariance descriptor has led to many publications and applications already. The work on this topic is now more about the precise use of the descriptor in varied applications than the design of new descriptors.

- The new action descriptors have led to finer gesture classification. As our target application is the detection of the Alzheimer syndrome from gesture analysis, which requires still finer descriptors, we will continue the work on this topic.

- The different shape priors developed (for shape growth enforcement, shape matching, articulated motion) have been formulated and designed so that efficient optimization tools could be used, leading to global optimality guarantees. These particular problems can thus be considered as solved, but there is still much work to be done on shape and related optimization, in particular to obtain shape statistics for human action recognition.

- The success obtained in the control of trackers is a proof of concept, but this work still needs to be pursued to get more practical and to be applied on more real world videos.

More precisely, the new results for perception for activity recognition are:

- People Detection for Crowded Scenes (6.3),
- Walking Speed Detection on a Treadmill using an RGB-D camera : experimentations and results (6.4),
- Head detection using RGB-D camera (6.5),
- Video Segmentation and Multiple Object Tracking (6.6),
- Enforcing Monotonous Shape Growth or Shrinkage in Video Segmentation (6.7),
- Multi-label Image Segmentation with Partition Trees and Shape Prior (6.8),
- Automatic Tracker Selection and Parameter Tuning for Multi-object Tracking (6.9),
- An Approach to Improve Multi-object Tracker Quality using Discriminative Appearances and Motion Model Descriptor (6.10),
- Person re-identification by pose priors(6.11),
- Global tracker : an online evaluation framework to improve tracking quality (6.12),
- Human action recognition in videos (6.13),
- Action Recognition using 3D Trajectories with Hierarchical Classifier (6.14),
- Action Recognition using Video Brownian Covariance Descriptor for Human (6.15),
- Towards Unsupervised Sudden Group Movement Discovery for Video Surveillance (6.16).

6.2.2. Semantic Activity Recognition

Participants: Vania Bogorny, Luis Campos Alvares, Vasanth Bathrinarayanan, Guillaume Charpiat, Duc Phu Chau, Serhan Cosar, Carlos F. Crispim Junior, Giuseppe Donatielo, Baptiste Fosty, Carolina Garate, Alvaro Gomez Uria Covella, Alexandra Konig, Farhood Negin, Anh-Tuan Nghiem, Philippe Robert, Carola Strumia.

For activity recognition, the main advances on challenging topics are:

- The utilization by clinicians for their everyday work of a first monitoring system able to recognize complex activities, to evaluate in real-time older people performance in an ecological room at Nice Hospital.

- The successful processing of over 80 older people videos and matching their performance for autonomy at home (e.g. walking efficiency) and cognitive disorders (e.g. realisations of executive tasks) with gold standard scales (e.g. NPI, MMSE). This research work contributes to the early detection of deteriorated health status and the early diagnosis of illness.

- The fusion of events coming from camera networks and heterogeneous sensors (e.g. RGB videos, Depth maps, audio, accelerometers).

- The management of the uncertainty of primitive events.

- The generation of event models in an unsupervised manner.

For this research axis, the contributions are :

- Autonomous Monitoring for Securing European Ports (6.17),

- Video Understanding for Group Behavior Analysis (6.18),

- Evaluation of an event detection framework for older people monitoring: from minute to hour-scale monitoring and Patients autonomy and dementia assessment (6.19),

- Uncertainty Modeling Framework for Constraint-based Event Detection in Vision Systems (6.20),

- Assisted Serious Game for older people (6.21),

- Enhancing Pre-defined Event Models using Unsupervised Learning (6.22),

- Using Dense Trajectories to Enhance Unsupervised Action Discovery (6.23),

- Abnormal Event Detection in Videos and Group Behavior Analysis (6.24).

6.2.3. Software Engineering for Activity Recognition

Participants: François Brémond, Daniel Gaffé, Sabine Moisan, Annie Ressouche, Jean-Paul Rigault, Omar Abdalla, Mohamed Bouatira, Ines Sarray, Luis-Emiliano Sanchez.

For the software engineering part, the main achievements are the Software Engineering methods and tools applied to video analysis. We have demonstrated that these approaches are appropriate and useful for video analysis systems:

- Run time adaptation using MDE is a promising approach. Our current prototype resorts to tools and technologies which were readily available. This made possible a proof of concepts.

- Introducing metrics in feature models was valuable to reduce the huge set of valid configurations after a dynamic context change and to provide a real time selection of an appropriate running configuration.

- The synchronous approach is well suited to describe reactive systems in a generic way, it has a well-established formal foundation allowing for automatic proofs, and it interfaces nicely with most model-checkers.

The contributions for this research axis are:

- Model-Driven Engineering for Activity Recognition Systems(6.25),

- Scenario Analysis Module (6.26),

- The Clem Workflow (6.27),

- Multiple Services for Device Adaptive Platform for Scenario Recognition (6.28).

6.3. People Detection for Crowded Scenes

Participants: Malik Souded, François Brémond.

keywords: people detection, crowded scenes, features, boosting.

This works aims at proposing an efficient people detection algorithm which can deal with crowded scenes.

6.3.1. Early Work

We have previously proposed an approach which optimizes state-of-the-art methods [Tuzel 2007, Yao 2008], based on training cascade of classifiers using LogitBoost algorithm on region covariance descriptors. This approach performs in real time and provides good detection performances in low to medium density scenes (see some examples in figure 10). However, this approach shows its limits on crowded scenes. Both detection accuracy and detection time are highly impacted in this case. The detection time increases dramatically due to the number of people in images, which forces the evaluation of many cascade levels, while the numerous partial occlusions highly decrease the detection rate (the considered detector is a full-body detector). To deal with these issues, we are working on a new approach.

6.3.2. Current Work

Our approach is based on training a cascade of classifiers using Boosting algorithms too, but on large sets of various features with several parameters for each of them (LBP, Haar-Like, HOG, Region Covariance Descriptor, etc.). The variety of features is motivated by three main reasons:

- Using fast features like LBP and Haar-like in the first levels of the cascade allows a fast rejection of a high part of negatives. The remaining ones will be rejected by a more sophisticated feature like Covariance Descriptor. This will highly decrease the detection time.

- Covariance Descriptor are not discriminative enough for very small regions. Our aim is to train the new detector on specific body parts, especially the upper one (shoulders and heads) to increase detection rate in highly crowded scenes (with a high rate of partial occlusions). Using a large set of various features allows the training system to select the ones which provide the best discriminative power for these regions.

- The possibility to combine several features to describe the same region, even by a simple concatenation, providing more discriminative power than using single features.

Another part of this approach consists in the optimization of the detector at two levels:

- Optimizing the training process by first clustering both positive and negative training samples. This clustering allows to focus on the hard samples which are too close to the other class from a classification point of view, providing more accurate detectors.

- Iterative training of several detectors on randomly selected samples, and weighting of the training samples according to their classification confidence, which allows to improve the clustering process.

The evaluation of this approach is still in progress.

6.4. Walking Speed Detection on a Treadmill using an RGB-D camera : experimentations and results

Participants: Baptiste Fosty, François Brémond.

keywords: RGB-D camera analysis, walking speed, serious games

Within the context of the developement of serious games for people suffering from Alzheimer disease (Az@Game project), we have developed an algorithm to compute the walking speed of a person on a treadmill. The goal is to use this speed inside the game to control the displacement of the avatar, and then for the patient to perform some physical as well as cognitive tasks. For the evaluation of the accuracy of the algorithm, we collected a video data set of healthy people walking on a motorized treadmill.

Figure 10. Some examples of detection using the previously proposed approach (see section Early Work).

Protocol. With the help of a specialist in the domain of physical activities, a protocol has been set up to cover the spectrum of the possible walking speeds and to prove the reproducibility of the results. This protocol consists in performing three times ten minutes of walking on the motorized treadmill, each attempt being itself divided in five times two minutes at the following speeds : 1.5 km/h, 2.5 km/h, 3.5 km/h, 4.5 km/h and 5.5 km/h. Participants, mostly people from the age of 18 to 60 without any physical disorder that could influence the gait, were asked to keep a natural gait and to follow the rotation of the treadmill.

Ground truth. The quantitative performances of the walking speed computation are evaluated by comparison with the speed of the walking person. The speed references are twofold :

- a theoretical value : the speed displayed by the treadmill, set up by the participant but imposed by the protocol (see Figure 12 , red graph),

- a practical value : white marks have been painted on the treadmill to recompute the real speed of the rotation and so the walking speed (see Figure 12 , green and blue graphs).

Results. The results presented herein are based on the videos of 36 participants who performed the protocol described above, with 17 males and 19 females, with an average age of 32.1 ± 7.7 years, an average height of 171.1 ± 9.1 cm and an average weight of 67.4 ± 13.6 kg.

Speed (km/h)	Number of observations	Average (km/h)	Standard deviation (km/h)	Mean absolute error (km/h)
1.5 (WM)	11823 (3170)	1.59 (1.508)	0.20 (0.014)	0.15 (0.011)
2.5 (WM)	14549 (5113)	2.56 (2.478)	0.20 (0.026)	0.16 (0.027)
3.5 (WM)	17399 (7133)	3.53 (3.453)	0.20 (0.040)	0.16 (0.051)
4.5 (WM)	19800 (9129)	4.47 (4.427)	0.21 (0.059)	0.17 (0.079)
5.5 (WM)	22163 (11180)	5.36 (5.397)	0.26 (0.116)	0.21 (0.116)
Total (WM)	**85734 (35725)**	**3.5 (3.453)**	**0.22 (0.074)**	**0.17 (0.071)**

Figure 11. Walking speed results. The number of observations corresponds to the number of time the speed has been computed. The WM lines refer to the results of the detection of the white mark.

The table in figure 11 shows the statistical evaluation of the performances of the system. The average column shows that the accuracy of the system is better for the median speeds (around 4.5km/h). When the person is walking slower, the system overestimate the speed due to the wrongly detected steps whereas when faster, there is an underestimation because of missing the exact time when the distance between feet is maximum (framerate too low).

A paper reporting this work is actually under writing process.

6.5. Head Detection Using RGB-D Camera

Participants: Marine Chabran, François Brémond.

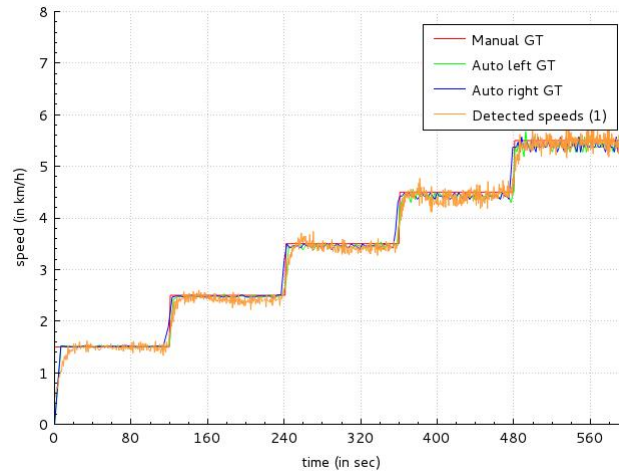**keywords:** RGB-D camera analysis, head detection, serious games

Figure 12. Graphical results of one attempt of protocal issue. The red graph corresponds to the theoretical speed, green and blue graphs correspond respectively to the detected speed of the left and right white marks on the treadmill, and the orange graph is the detected walking speed of the person.

The goal of this work is to improve a head detection algorithm using RGB-D sensor (like a Kinect camera) for action recognition as part of a study of autism. The psychologists want to compare the learning process of children with autism syndrome depending on games (digital or physical toys).

The algorithm described in [79] represents a head by its center position. It takes three steps to determine this point :

- Determine possible head center positions using a head model : inner circle radius=6 cm, outer circle radius=20 cm (Figure 13).
  A good inner point is a point on the inner circle verifying :

$$depthHeadCenter + 30cm > depthInnerPoint > depthHeadCenter - 30cm.$$

  A good outer point is a point on the outer circle verifying :

$$depthHeadCenter < depthOuterPoint + 15cm.$$

- Merge close head centers separated by less than 4 pixels.
- Select final head center according to its score (calculated according to the number of good inner and outer points).

For now, it works well within video where people are close to the camera (about 1 meter) and without any background just behind them (Figure 14).

The problem is when the person is sitting and the head is ahead of the body (Figure 15) or close to a wall, the difference between head depth and outer circle depth becomes not sufficient (about 10 cm).

We have evaluated the performance of this algorithm with two data sets (Table 1). For Lenval Hospital data set, we have evaluated 2 series of 200 frames, for the Smart Home data set, we have evaluated 3 series of 300 frames (a total of 1300 heads).

Figure 13. Each circle is divided in n parts (n=8). The points on the inner circle must have a similar depth with the center point, the points on the outer circle must be further than the center point compared to the camera

Table 1. Performance of head detection and people detection on two different data sets.

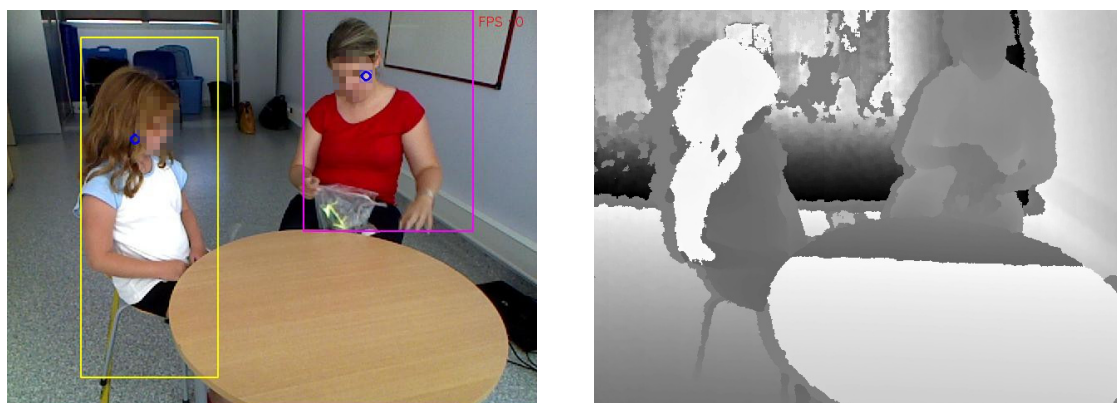| Videos	Head Detection (%)	People detection (%)
Lenval Hospital dataset (Figure 14)	89.7	96.9
Rest home dataset (Figure 15)	62.8	85.3

Figure 14. Result of head detection - good detection. The bounding box represents the person, the small blue circle represents the head center.
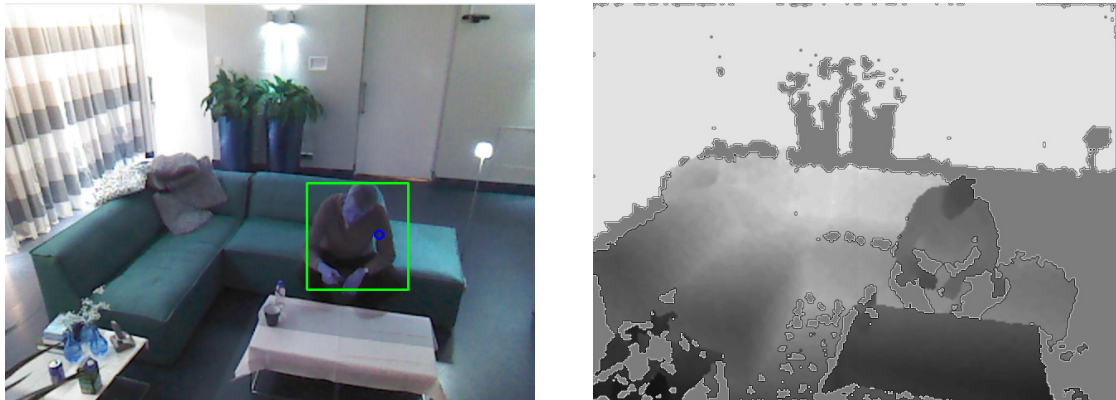
Figure 15. Result of head detection - wrong detection. The green bounding box represents the person, the small blue circle represents the head center.

6.6. Video Segmentation and Multiple Object Tracking

Participants: Ratnesh Kumar, Guillaume Charpiat, Monique Thonnat.

keywords: Fibers, Graph Partitioning, Message Passing, Iterative Conditional Modes, Video Segmentation, Video Inpainting

This year we focussed on multiple object tracking, and writing of the thesis manuscript of Ratnesh (defense on December 2014).

The first contribution of this thesis is in the domain of video segmentation wherein the objective is to obtain a dense and coherent spatio-temporal segmentation. We propose joining both spatial and temporal aspects of a video into a single notion *Fiber*. A Fiber is a set of trajectories which are spatially connected by a mesh. Fibers are built by jointly assessing spatial and temporal aspects of the video. Compared to the state-of-the-art, a fiber based video segmentation presents advantages such as a natural spatio-temporal neighborhood accessor by a mesh, and temporal correspondences for most pixels in the video. Furthermore, this fiber-based segmentation is of quasi-linear complexity w.r.t. the number of pixels. The second contribution is in the realm of multiple object tracking. We proposed a tracking approach which utilizes cues from point tracks, kinematics of moving objects and global appearance of detections. Unification of all these cues is performed on a Conditional Random Field. Subsequently this model is optimized by a combination of message passing and an Iterated Conditional Modes (ICM) variant to infer object-trajectories. A third, minor, contribution relates to the development of suitable feature descriptor for appearance matching of persons. All of our proposed approaches achieve competitive and better results (both qualitatively and quantitatively) than state-of-the-art open source datasets.

This first part of the thesis was published at IEEE WACV at the beginning of this year [43], and the work on *multiple object tracking* was recently presented at Asian Conference on Computer Vision [44]

Sample visual results from our recent publication [44] can be seen in Figure 16 .

6.7. Enforcing Monotonous Shape Growth or Shrinkage in Video Segmentation

Participant: Guillaume Charpiat [contact].

This work has been done in collaboration with Yuliya Tarabalka (Ayin team, Inria-SAM), Bjoern Menze (Technische Universität München, Germany), and Ludovic Brucker (NASA GSFC, USA) [http://www.nasa.gov].
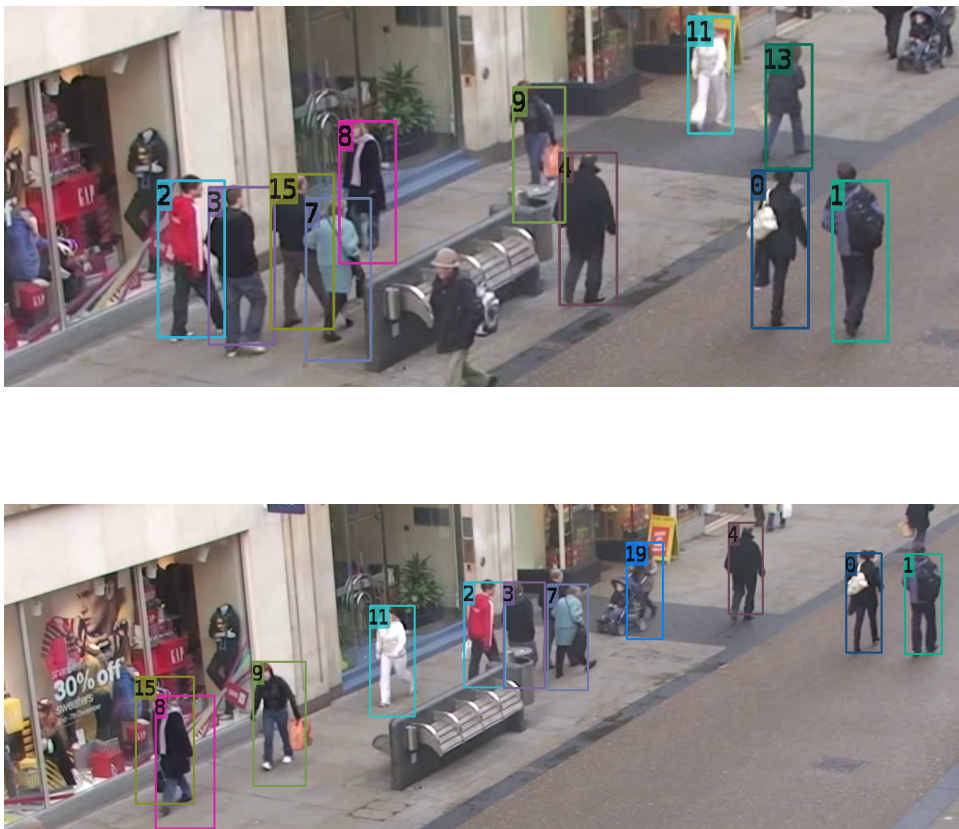
Figure 16. **Consistent people crossing in dense scenarios.** *The two images are 121 frames apart.*

keywords: Video segmentation, graph cut, shape analysis, shape growth

The automatic segmentation of objects from video data is a difficult task, especially when image sequences are subject to low signal-to-noise ratio or low contrast between the intensities of neighboring structures. Such challenging data are acquired routinely, for example, in medical imaging or satellite remote sensing. While individual frames can be analyzed independently, temporal coherence in image sequences provides a lot of information not available for a single image. In this work, we focused on segmenting shapes that grow or shrink monotonically in time, from sequences of extremely noisy images.

We proposed a new method for the joint segmentation of monotonically growing or shrinking shapes in a time sequence of images with low signal-to-noise ratio [32]. The task of segmenting the image time series is expressed as an optimization problem using the spatio-temporal graph of pixels, in which we are able to impose the constraint of shape growth or shrinkage by introducing unidirectional infinite-weight links connecting pixels at the same spatial locations in successive image frames. The globally-optimal solution is computed with graph-cuts. The performance of the proposed method was validated on three applications: segmentation of melting sea ice floes; of growing burned areas from time series of 2D satellite images; and of a growing brain tumor from sequences of 3D medical scans. In the latter application, we imposed an additional inter-sequences inclusion constraint by adding directed infinite-weight links between pixels of dependent image structures. Figure 17 shows a multi-year sea ice floe segmentation result. The proposed method proved to be robust to high noise and low contrast, and to cope well with missing data. Moreover, in practice, its complexity was linear in the number of images.
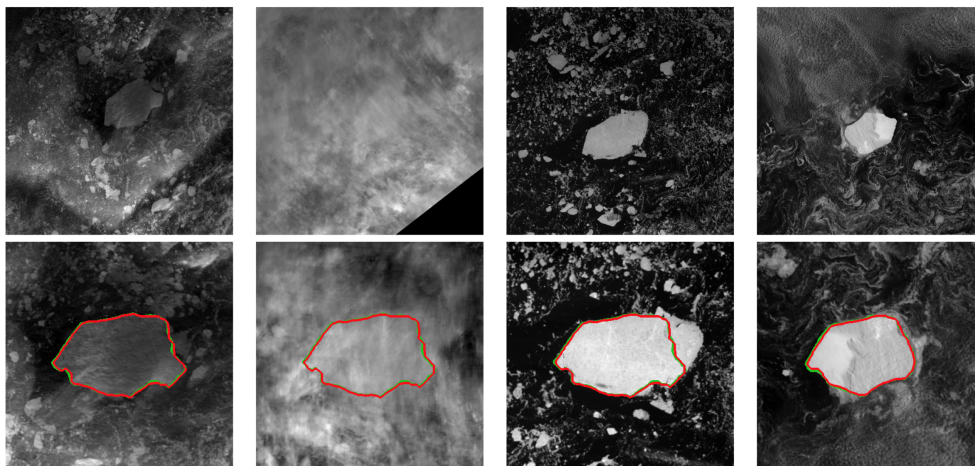
Figure 17. Top: MODIS images at four time moments (days 230, 233, 235 and 267 of 2008, respectively). Bottom: corresponding aligned images with segmentation contours (in red). Manual segmentation (ground truth) is shown in green.

6.8. Multi-label Image Segmentation with Partition Trees and Shape Prior

Participant: Guillaume Charpiat [contact].

This work has been done in collaboration with Emmanuel Maggiori and Yuliya Tarabalka (Ayin team, Inria-SAM).
keywords: partition trees, multi-class segmentation, shape priors, graph cut

The multi-label segmentation of images is one of the great challenges in computer vision. It consists in the simultaneous partitioning of an image into regions and the assignment of labels to each of the segments. The problem can be posed as the minimization of an energy with respect to a set of variables which can take one of multiple labels. Throughout the years, several efforts have been done in the design of algorithms that minimize such energies.

We propose a new framework for multi-label image segmentation with shape priors using a binary partition tree [50]. In the literature, such trees are used to represent hierarchical partitions of images, and are usually computed in a bottom-up manner based on color similarities, then processed to detect objects with a known shape prior. However, not considering shape priors during the construction phase induces mistakes in the later segmentation. This study proposes a method which uses both color distribution and shape priors to optimize the trees for image segmentation. The method consists in pruning and regrafting tree branches in order to minimize the energy of the best segmentation that can be extracted from the tree. Theoretical guarantees help reducing the search space and make the optimization efficient. Our experiments (see Figure 18) show that the optimization approach succeeds in incorporating shape information into multi-label segmentation, outperforming the state-of-the-art.

(a) Color image.   (b) Manual segment.   (c) BPT optimization.

(d) Manual.   (e) SVM.   (f) GC.   (g) BPT Opt.

	$\mathcal{A}$	$\mathcal{D}$
SVM	0.764	0.50
GC	0.799	0.49
TC	0.804	0.47
Opt.	0.813	0.69

(h) Performance.

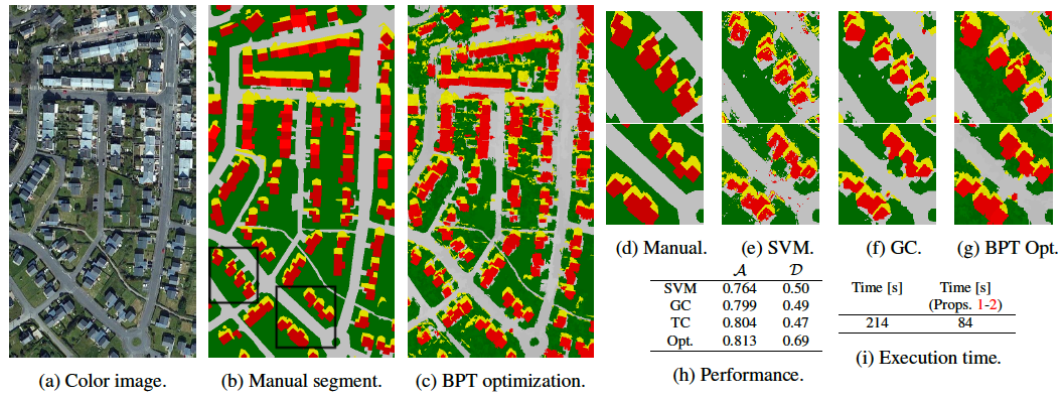| Time [s]	Time [s] (Props. 1-2)
214	84

(i) Execution time.

Figure 18. Classification results for the satellite image over Brest. $\mathcal{A}$ denotes overall classification accuracy, and $\mathcal{D}$ denotes average building's overlap. The performance of the proposed binary partition tree (BPT) optimization method is compared with the following methods: 1) support vector machines (SVM) classification; 2) graph cut (GC) with α-expansion; 3) cut on the BPT, regularized by the number of regions without using shape priors (TC).

6.9. Automatic Tracker Selection and Parameter Tuning for Multi-object Tracking

Participants: Duc Phu Chau, Slawomir Bak, François Brémond, Monique Thonnat.

Keywords: object tracking, machine learning, tracker selection, parameter tuning

Many approaches have been proposed to track mobile objects in a scene [87], [45]. However the quality of tracking algorithms always depends on video content such as the crowded level or lighting condition. The selection of a tracking algorithm for an unknown scene becomes a hard task. Even when the tracker has already been determined, there are still some issues (e.g. the determination of the best parameter values or the online estimation of the tracking reliability) for adapting online this tracker to the video content variation. In order to overcome these limitations, we propose the two following approaches.

The main idea of the first approach is to learn offline how to tune the tracker parameters to cope with the tracking context variations. The tracking 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. In an offline phase, training video sequences are classified by clustering their contextual features. Each context cluster is then associated to satisfactory tracking parameters using tracking annotation associated to training videos. In the online control phase, once a context change is detected, the tracking parameters are tuned using the learned parameter values. This work has been published in [30].

A limitation of the first approach is the need of annotated data for training. Therefore we have proposed a second approach without training data. In this approach, the proposed strategy combines an appearance tracker and a KLT tracker for each mobile object to obtain the best tracking performance (see figure 19). This helps to better adapt the tracking process to the spatial distribution of objects. Also, while the appearance-based tracker considers the object appearance, the KLT tracker takes into account the optical flow of pixels and their spatial neighbours. Therefore these two trackers can improve alternately the tracking performance.
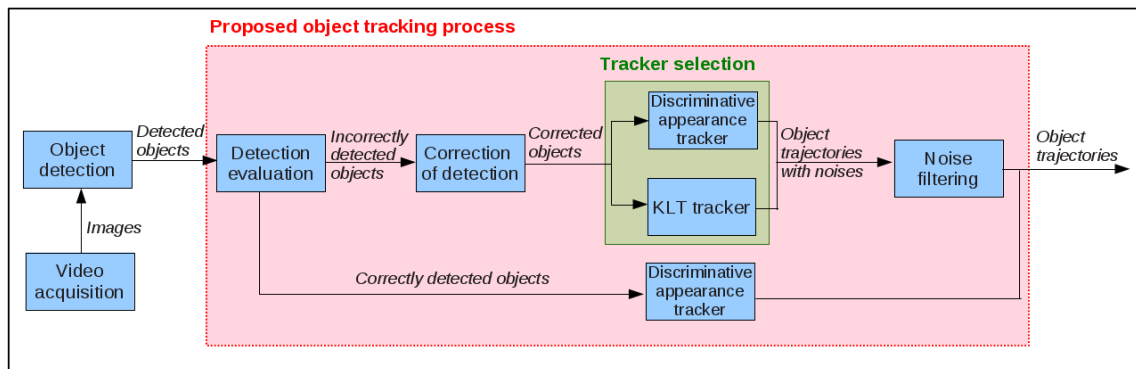

Figure 19. The scheme of the second approach

The second approach has been experimented on three public video datasets. Figure 20 presents correct tracking results of this approach even with strong object occlusion in PETS 2009 dataset. Table 2 presents the evaluation results of the proposed approach, the KLT tracker, the appearance tracker and different trackers from the state of the art. While using separately the KLT tracker or the appearance tracker, the performance is lower than other approaches from the state of the art. The proposed approach by combining these two trackers improves significantly the tracking performance and obtains the best values for both metrics. This work has been published in [39].

Table 2. Tracking results on the PETS sequence S2.L1, camera view 1, sequence time 12.34. The best values are printed in bold.

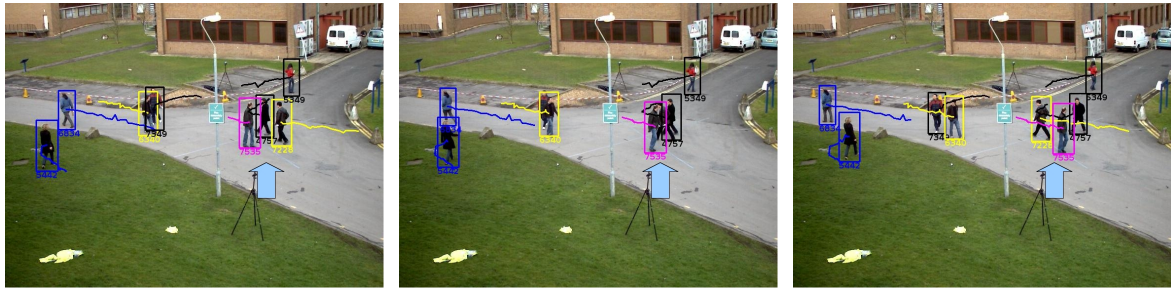| Method	MOTA	MOTP
Berclaz et al. [60]	0.80	0.58
Shitrit et al. [86]	0.81	0.58
KLT tracker	0.41	**0.76**
Appearance tracker	0.62	0.63
Proposed approach	**0.86**	0.72

Figure 20. Tracker result: Three persons of Ids 7535, 7228 and 4757 (marked by the cyan arrow) are occluded each other but their identities are kept correctly after occlusion.

6.10. An Approach to Improve Multi-object Tracker Quality Using Discriminative Appearances and Motion Model Descriptor

Participants: Thi Lan Anh Nguyen, Duc Phu Chau, François Brémond.

Keywords: Tracklet fusion, Multi-object tracking

Many recent approaches have been proposed to track multi-objects in a video. However, the quality of trackers is remarkably effected by video content. In the state of the art, several algorithms are proposed to handle this issue. The approaches in [39] and   [64] propose methods which compute online or learn descriptor weights during tracking process. These algorithms adapt the tracking to the scene variations but are less effective when mis-detection occurs in a long period of time. Inversely, the algorithms in [59] and   [58] can recover a long-term mis-detection by fusing tracklets. However, the descriptor weights in these tracklet fusion algorithms are fixed in the whole video. Furthermore, above algorithms track objects based on object appearance which is not reliable enough when objects look similar to each other.

In order to overcome mentioned issues, the proposed approach brings three contributions: (1) appearance descriptors and motion model combination, (2) online discriminative descriptor weight computation and (3) discriminative descriptors based tracklet fusion. In particular, the appearance of one object can be discriminative with other objects in this scene but can be similar with other objects in another scene. Therefore, tracking objects based on only object appearance is less effective. In order to improve tracker quality, assuming that objects move with constant velocity, this approach firstly combines a constant velocity model from [70] and other appearance descriptors. Continuously, discriminative descriptor weights are computed online to adapt the tracking to each video scene. The more a descriptor discriminates one tracklet over other tracklets, the higher its weight value is. Next, based on these descriptor weights, the similarity score between the target tracklet with its candidate is computed. In the last step, tracklets are fused to a long trajectory by Hungarian algorithm with the optimization of global similarity scores.

The proposed approach gets results of tracker in   [63] as input and is tested on challenge datasets. This approach achieves comparable results with other trackers from the state of the art. Figure 1 shows that the tracklet keeps its ID even when occlusion occurs. Table 1 shows the better performance of this approach compared to other trackers from the state of the art.

6.11. Person Re-identification by Pose Priors

Participants: Slawomir Bak, Sofia Zaidenberg, Bernard Boulay, Filipe Martins, Francois Brémond.

keywords: re-identification, pose estimation, metric learning
Human appearance registration, alignment and pose estimation
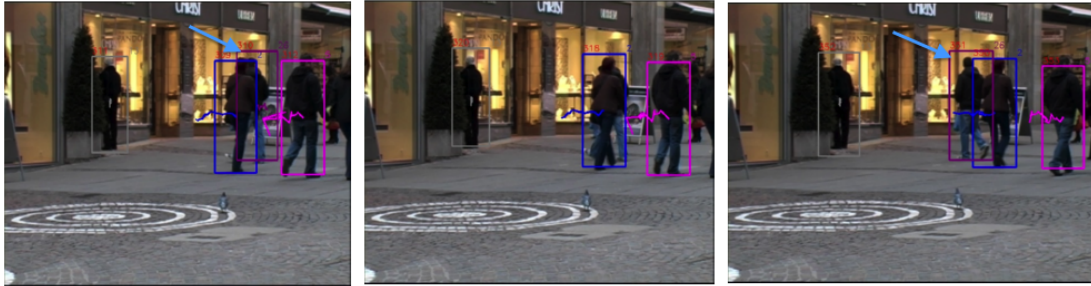
Figure 21. The proposed approach with TUD dataset: Object ID_{26} (presented by pink bounding box) keeps its ID correctly after 11 frames of mis-detection.

Table 3. Tracking results on datasets: TUD-Stadtmitte and TUD-crossing. The best values are printed in bold

Dataset	Method	MT(%)	PT(%)	ML(%)
TUD-Stadtmitte	[57]	60.0	30.0	10.0
TUD-Stadtmitte	[30]	**70.0**	10.0	20.0
TUD-Stadtmitte	[71]	**70.0**	30.0	**0.0**
TUD-Stadtmitte	[95]	**70.0**	30.0	**0.0**
TUD-Stadtmitte	Ours	**70.0**	30.0	**0.0**
TUD-Crossing	[89]	**53.8**	38.4	7.8
TUD-Crossing	Ours	**53.8**	46.2	**0.0**
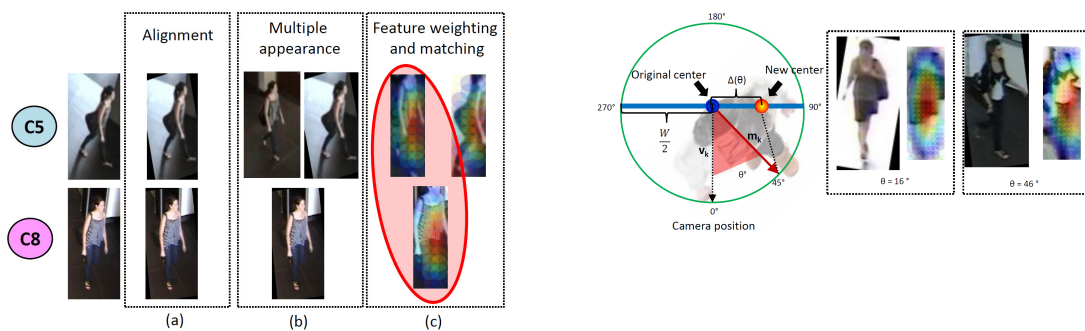
(a) left   (b) right

Figure 22. Improvements on re-identification using viewpoint cues: (a) target alignment; (b) multiple target appearance based on clustering; (c) pose orientation-driven weighting. The left illustration shows an example of the same person viewed from two different cameras. The right image presents pose estimation algorithm.

Re-identifying people in a network of cameras requires an invariant human representation. State of the art algorithms are likely to fail in real-world scenarios due to serious perspective changes. Most of existing approaches focus on invariant and discriminative features, while ignoring the body alignment issue. In this work we proposed 3 methods for improving the performance of person re-identification. We focus on eliminating perspective distortions by using 3D scene information. Perspective changes are minimized by affine transformations of cropped images containing the target (1). Further we estimate the human pose for (2) clustering data from a video stream and (3) weighting image features. The pose is estimated using 3D scene information and motion of the target. Pose orientation is computed by dot product between viewpoint vector and motion of the target (see figure 22). We validated our approach on a publicly available dataset with a network of 8 cameras. The results demonstrated significant increase in the re-identification performance over the state of the art [36].

Matching employing pose priors
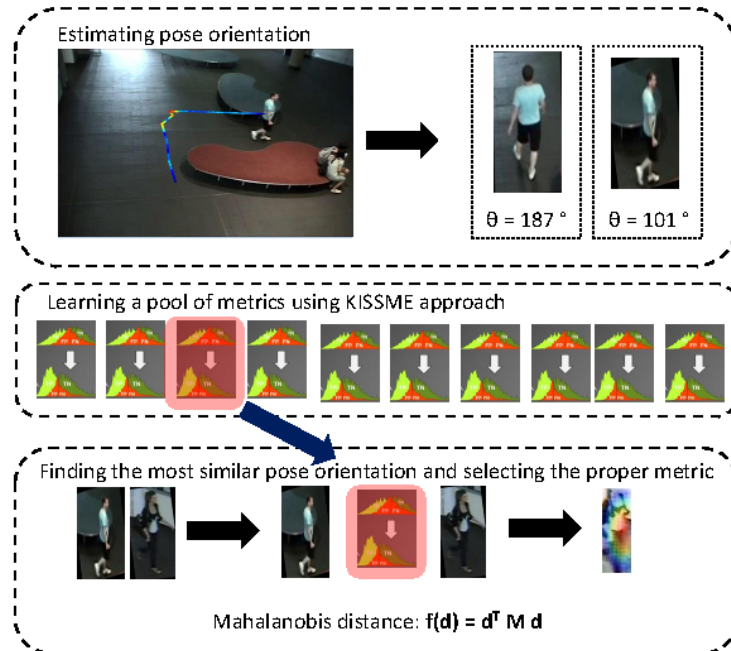
Figure 23. The proper metric is selected from the pool of previously learned metric for matching different poses.

Currently we are working on learning the matching strategy of appearance extracted from different poses. We employ well known metric learning tools for matching given poses. Let us assume that pose can be described by the angle between the motion vector of the target and the viewpoint vector of the camera (see figure 22). Thus for each target appearance we can express the pose as the angle in the range of [0,360). We decide to divide this range into n bins. Given n bins of estimated poses, we learn how to match different poses corresponding to different bins. In the result, we learn $n * (n + 1)/2$ metrics. While learning metrics, we follow a well known scheme based on image pairs, containing two different poses of the same target as positives and pairs of different poses containing different targets as negatives. The learned metrics stand for the metric pool. This metric pool is learned offline and does not depend on camera pair. In the result, once metric pool is learned, it can be used for any camera pair.

Given two images from different (or the same) camera, we first estimate the poses for each image. Having two poses, we select a corresponding metric from the metric pool. The selected metric provides the strategy to compute similarity between two images (see figure 23).

6.12. Global Tracker : An Online Evaluation Framework to Improve Tracking Quality

Participants: Julien Badie, Slawomir Bak, Duc Phu Chau, François Brémond, Monique Thonnat.

keywords: online quality estimation, re-identification, tracking results improvements

This work addresses the problem of estimating the reliability of a tracking algorithm during runtime and correcting the anomalies found. Evaluating and tuning a tracking algorithm generally requires multiple runs and ground truth. The proposed framework called global tracker overcomes these limitations by combining an online evaluation algorithm and a recovering post-process.

Designing an evaluation framework that does not require ground truth has many different applications. One of them is to provide feedback to the tracking algorithm that can tune its own parameters to improve the results on the next frame. Another convenient application is to filter the reliable information from the tracking algorithm that can be used by the next processing step such as event recognition or re-identification.
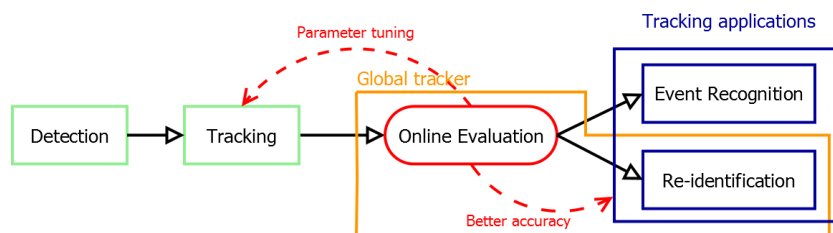
Figure 24. Integration of the global tracker, combining online evaluation and re-identification

The proposed online evaluation framework is based on control features. It means that several representative values or characteristics (the features) are chosen to monitor what is happening. Control features are the features that the online evaluation framework uses to monitor the status of the tracked objects. The framework is divided into two steps :

- computing the control features related to each tracked object of the scene on the current frame
- detecting the possible anomalies and then classifying them into two categories : anomalies due to tracking errors or benign anomalies (when a person leaves the scene or cross an obstacle for example).

This approach has been tested on two datasets (PETS 2009 and Caviar) with two different tracking algorithms (a multi-feature tracker and a tracker based on graph partitioning). The results show that the global tracker, even associated with a tracking algorithm that does not have good results, can perform as well as the state of the art.

This approach has been published in AVSS 2014 [33] which details the differences between real errors and benign anomalies.

6.13. Human Action Recognition in Videos

Participants: Piotr Bilinski, François Brémond.

keywords: Action Recognition; Human Action Recognition

Table 4. Tracking results on sequence S2.L1.View1 of the PETS2009 dataset using CLEAR metrics

Methods	MOTA	MOTP
Heili et al. [72]	0.89	0.66
Zamir et al. [96]	**0.90**	0.69
Milan et al. [77]	**0.90**	**0.74**
Tracker 1	0.62	0.63
Tracker 1 + global tracker	0.85	0.71
Tracker 2	0.85	**0.74**
Tracker 2 + global tracker	**0.90**	**0.74**

Table 5. Tracking results on the Caviar dataset using Mostly Tracked (MT), Partially Tracked (PT) and Mostly Lost (ML) metrics

Method	MT (%)	PT (%)	ML (%)
Li et al. [76]	84.6	14.0	1.4
Kuo et al. [74]	84.6	14.7	**0.7**
Tracker 1	78.3	16.0	5.7
Tracker 1 + global tracker	**86.4**	**8.3**	5.3

This Ph.D. thesis targets the automatic recognition of human actions in videos. Human action recognition is defined as a requirement to determine what human actions occur in videos. This problem is particularly hard due to enormous variations in visual and motion appearance of people and actions, camera viewpoint changes, moving background, occlusions, noise, and enormous amount of video data.

Firstly, we review, evaluate, and compare the most popular and the most prominent state-of-the-art techniques, and we propose our action recognition framework based on local features, which we use throughout this thesis work embedding the novel algorithms. Moreover, we introduce a new dataset (CHU Nice Hospital) with daily self care actions of elder patients in a hospital.

Then, we propose two local spatio-temporal descriptors for action recognition in videos. The first descriptor is based on a covariance matrix representation, and it models linear relations between low-level features. The second descriptor is based on a Brownian covariance, and it models all kinds of possible relations between low-level features.

Then, we propose three higher-level feature representations to go beyond the limitations of the local feature encoding techniques.

The first representation is based on the idea of relative dense trajectories. We propose an object-centric local feature representation of motion trajectories, which allows to use the spatial information by a local feature encoding technique.

The second representation encodes relations among local features as pairwise features. The main idea is to capture the appearance relations among features (both visual and motion), and use geometric information to describe how these appearance relations are mutually arranged in the spatio-temporal space.

The third representation captures statistics of pairwise co-occurring visual words within multi-scale feature-centric neighbourhoods. The proposed contextual features based representation encodes information about local density of features, local pairwise relations among the features, and spatio-temporal order among features.

Finally, we show that the proposed techniques obtain better or similar performance in comparison to the state-of-the-art on various, real, and challenging human action recognition datasets (Weizmann, KTH, URADL, MSR Daily Activity 3D, HMDB51, and CHU Nice Hospital).
The Ph.D. thesis was defended on December 5, 2014.

6.14. Action Recognition Using 3D Trajectories with Hierarchical Classifier

Participants: Michal Koperski, Piotr Bilinski, François Brémond.

keywords: action recognition, computer vision, machine learning, 3D sensors

The goal of our work is to extend recently published approaches ([61], [93]) for Human Action Recognition to take advantage of the depth information from 3D sensors.

We propose to add depth information to trajectory based algorithms ([61], [93]). Currently mentioned algorithms compute trajectories by sampling video frames and then tracking points of interest - creating the trajectory. Our contribution is to create even more discriminative features by adding depth information to previously detected trajectories. In our work we propose methods to deal with noise and missing measurements in depth map.

The second contribution is a technique to deal with actions which do not contain enough motion to compute discriminative trajectory descriptors. Actions like sitting, standing, laptop use do not contain large amount of motion, or motion is occluded by the object. For such cases we proposed LDP (Local Depth Pattern) descriptor which does not require motion to be computed.

Proposed descriptors are further processed using a Bag of Words method and SVM classifier. We use hierarchical approach where at first level we train classifier to recognize if given example contains high or low amount of motion. Then at second layer we train SVM classifier to recognize action labels.

Figure 25. Visualization of MSR Dailiy Activty 3D data set (left) - video input frame, (center) - frame with detected trajectories (red - static points, green detected trajectories, (right) - corresponding depth map

The evaluation of our method was conducted on "Microsoft Daily Activity3D" data set [94] which consists of 16 actions (drink, eat, read book, call cellphone, write on a paper, use laptop etc.) performed by 10 subjects. We achieve superior performance among techniques which do not require skeleton detection. This work was published in proceedings of the 21st IEEE International Conference on Image Processing, ICIP 2014 [42]

6.15. Action Recognition using Video Brownian Covariance Descriptor for Human

Participants: Piotr Bilinski, Michal Koperski, Slawomir Bak, François Brémond.

keywords: action recognition, computer vision, machine learning

This work addresses a problem of recognizing human actions in video sequences. Recent studies have shown that methods which use bag-of-features and space-time features achieve high recognition accuracy [61], [93], [42]. Such methods extract both appearance-based and motion-based features. In image processing, a novel trend has emerged that ignores explicit values of given features, focusing instead on their pairwise relations. The most known example of such an approach is covariance descriptor [92]. Inspired by Brownian motion statistics [88] and application in people Re-identification [35]; we propose to model relationships between different pixel-level appearance features such as intensity and gradient using Brownian covariance, which is a natural extension of classical covariance measure. While classical covariance can model only linear relationships, Brownian covariance models all kinds of possible relationships. We propose a method to compute Brownian covariance on space-time volume of a video sequence. We show that proposed Video Brownian Covariance (VBC) descriptor carries complementary information to the Histogram of Oriented Gradients (HOG) descriptor. The fusion of these two descriptors gives a significant improvement in performance on three challenging action recognition datasets. The result of this work was published in proceedings of the 11th IEEE International Conference on Advanced Video and Signal-Based Surveillance, AVSS 2014 [38].
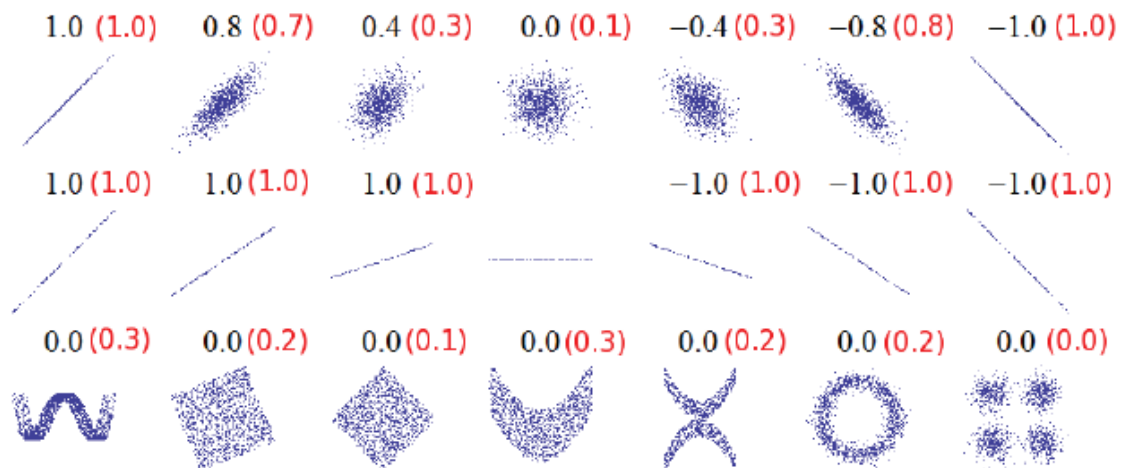
Figure 26. Comparison between Covariance and Brownian distance correlation. Covariance values in black, Brownian values in red.

6.16. Towards Unsupervised Sudden Group Movement Discovery for Video Surveillance

Participants: Sofia Zaidenberg, Piotr Bilinski, François Brémond.

keywords: Event detection; Motion estimation; Anomaly estimation; Situation awareness; Scene Understanding; Group Activity Recognition; Stream Selection

We present a novel and unsupervised approach for discovering "sudden" movements in surveillance videos. The proposed approach automatically detects quick motions in a video, corresponding to any action. A set of possible actions is not required and the proposed method successfully detects potentially alarm-raising actions without training or camera calibration. Moreover, the system uses a group detection and event recognition framework to relate detected sudden movements and groups of people, and to provide a semantical interpretation of the scene. We have tested our approach on a dataset of nearly 8 hours of videos recorded from

two cameras in the Parisian subway for a European Project. For evaluation, we annotated 1 hour of sequences containing 50 sudden movements.

This work has been published in [47].

6.17. Autonomous Monitoring for Securing European Ports

Participants: Vasanth Bathrinarayanan, François Brémond.

Keywords: Event Recognition, Port Surveillance

This work is done for the European research project SUPPORT (Security UPgrade for PORTs). This project addresses potential threats on passenger life and the potential for crippling economic damage arising from intentional unlawful attacks on port facilities, by engaging representative stakeholders to guide the development of next generation solutions for upgraded preventive and remedial security capabilities in European ports. The overall benefit is securing and efficient operation of European ports enabling uninterrupted flows of cargo and passengers while suppressing attacks on high value port facilities, illegal immigration and trafficking of drugs, weapons and illicit substances.

Scene understanding platform was tested on this new dataset, which has archived footage from past incidents and some acted scenarios. The processing pipeline of algorithms contains camera calibration, background subtraction using GMM (Gaussian Mixture Model), people detection using DPM (Deformable Parts Model), Tracking (Frame to Frame), Event recognition.

We collected several hours of videos which contained security related events like Intrusion to port by different methods (sea, gates, fences), Spying activities from outside the port, robbery or theft, ticketless travelling, restricted zone access, abondon luggage and some abnormal behaviors. The system was modelled and validated for all the above events to be detected and also a live real time demo was done for the completion of the project. All the events from our systems are later sent to project partners for fusion of the data with other sensors data and information from police, internal and external reports, etc., to detect complex security threats (see figure 27).

6.18. Video Understanding for Group Behavior Analysis

Participants: Carolina Garate, François Brémond.

keywords: Computer vision, group tracking, scene understanding, group behavior recognition, video surveillance, event detection.

The main work in this PhD thesis concerns the recognition of the behaviors of a group of people (2-5 persons) involved in a scene depicted by a video sequence.

Our goal focuses on the automatic recognition of behavior patterns in video sequence for groups of people (2-5 persons). We want to build a real time system able to recognize various group scenarios.

The approach includes different tasks to achieve the final recognition. The first one consists in tracking groups of moving regions detected in the video sequence acquired by the cameras. The second task attempts to classify these moving regions into people classes. Finally, the last task recognizes group scenarios using a priori knowledge containing scenario models predefined by experts and also 3D geometric and semantic information of the observed environment.

Our approach considers a chain process consisting of 5 consecutive steps for video processing. The steps are : 1) segmentation, 2) physical object detection, 3) physical objects tracking, 4) group tracking and 5) group behavior recognition. Our research focuses on the last two phases.

First, group scenarios have been defined (and then recognized) using the general scenario description language. Second, the likelihood of the group scenario recognition has been quantified. Third, machine learning techniques have been investigated to learn and recognize these scenarios.
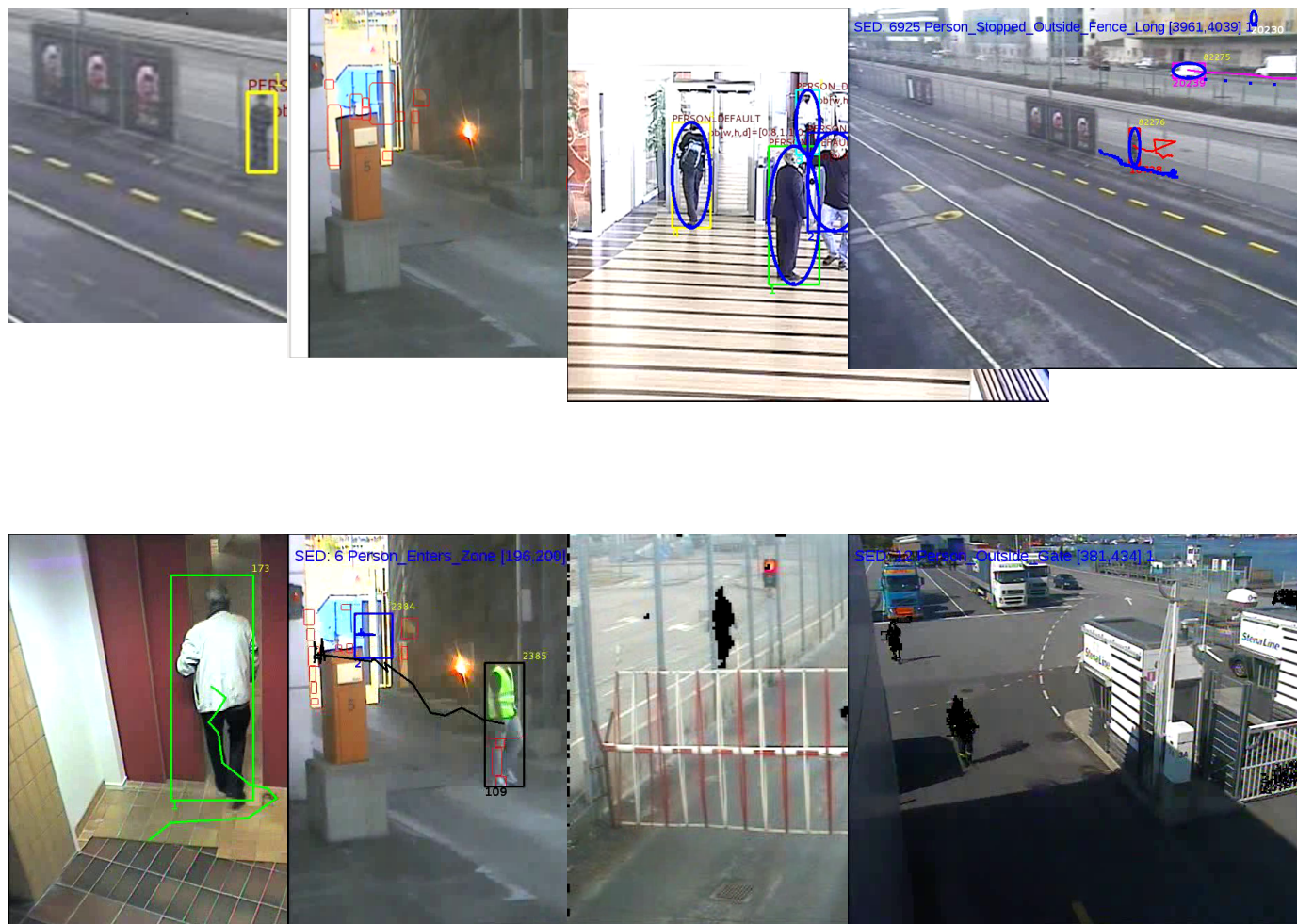
Figure 27. Detection of complex security threats from the tested dataset for Port surveillance

We have processed the data set from 1 month video surveillance camera in the Torino subway and the Minds eye data set. Recognizing several and different events such as: walking groups, standing still groups, running groups, calm groups (i.e. having a bounding box with stable size), active groups (i.e. with bounding box's size variations, meaning that group members move a lot).

6.19. Evaluation of an Event Detection Framework for Older People Monitoring: from Minute to Hour-scale Monitoring and Patients Autonomy and Dementia Assessment

Participants: Carlos F. Crispim-Junior, Alvaro Gomez Uria Covella, Carola Strumia, Baptiste Fosty, Duc Phu Chau, Anh-Tuan Nghiem, Alexandra Konig, Auriane Gros, Philippe Robert, François Brémond.

keywords: RGBD cameras, description-based activity recognition, older people,

Two main works are reported here: the continuous evaluation and extention of our event detection framework for older people monitoring, and the proposal of a behavioral classification model for the assessment of autonomy and cognitive health level of older people using automatically detected events. The evaluation of our event monitoring framework was extended from 29 to 49 recordings of senior participants undertaking physical tasks (7 min per participant, total : 5.71 hours) and instrumental activities of daily living (IADL, 15 minutes per participant, total: 12.25 hours). The recordings have taken place in a ecological observation room set in the Memory Center of Nice hospital. In the extended evaluation we employed a RGBD sensor as input instead of a standard RGB camera due to its advantages like invariance to illumination changes and real-time measurements of 3D information which foster better performance of the underlying algorithms for people detection and tracking. Table 6 presents the event monitoring performance of the present system for 49 participants. Event detection performance on physical task generalized to the larger dataset with a small performance increase of 1.4% (average F-Score). Concerning IADL detection although the global performance value (F-Score, 80.7 %) are the same, the new approach have made a trade-off between recall and precision to obtain more reliable detection of activities and their parameter estimations. Low precision values on preparing drink (e.g., making coffee) and watering plant are due to these activities being performed in very close - if not overlapping - locations (contextual zones). Low precision values in reading are due to the preferred reading location be close to image edges where most parts of person body are frequently outside the camera field of view.

Using the event monitoring system as input we have devised a behavioral classification model for the automatic assessment of participant cognitive health and autonomy level. Besides to event data the model also uses fine-grained data about person gait attributes (*e.g.*, stride-length, cadence, etc), obtained by a RGBD-based algorithm for gait analysis also developed in STARS team. Briefly, the event monitoring system supports the doctor by automatically annotating the patient daily living activities and assessing his/her gait parameters in a quantitative way, and the behavioral model performs the classification of participant's dementia and autonomy levels as a complement for standard psychometric scales for autonomy. We achieved an average accuracy of 83.67 % at the prediction of patient autonomy (poor, mediocre, good), and of 73.46 % for cognitive level class (healhy, memory cognitive impairment - MCI, alzheimer's disease), all models using a Naíve Bayes classifier. The results suggest that the behavioral classification model using automatically detected events outperforms the same model using events manually annotated by domain experts (81 %). On the contrary, the model using annotated data still outperforms the automated detection at dementia classification (79.46 %). Results indicate it is easier to predict the autonomy level than the Dementia, since the latter may be seen as the cause /source and the first its consequences. Deciding whether a decay on cognitive abilities relates to normal aging or early MCI or a given mild cognitive decay is an early symptom of Alzheimer's diasease or a severe case of MCI is also a open-problem for medical community. Future work will focus on investigating whether the remaining performance to achieve is related to the performance failures of the underling event monitoring system, to important behavioral aspects still not covered by the behavioral model, or even to the inherently ambiguous nature of the dementia classes.

Table 6. Event Monitoring Performance

Physical Tasks	Recall	Precision
Single Task	100%	88%
Dual Task	100%	98%
IADLs	Recall	Precision
Preparing drug box	87%	93%
Watering plant	80%	63%
Reading	60%	88%
Prepare drink	90%	68%
Talk on phone	89%	89%

We have also started the evaluation of the event monitoring system in Nursing home scenario passing from a minute time-scale to hours. A first participant was monitored with two RGBD sensors, one for bed-related events (sleep, bed exits) and one for living room and daily living activity events for 14 days. Preliminary results are 80 % for entering in bed and 100 % for bed exit in set of 6 events of each class in 13 hours monitoring (6 pm - 7am). Figure 28 illustrates the detection of restroom usage during the night. The automatic monitoring of participant activities during night is an important contribution to medical/nursing staff as wandering behavior at night is a common cause of accident in older people population. For instance, detecting whether a bed-exit during the night will be followed by a restroom visit or a bedroom exit plays a significant role at predicting a possibly dangerous situation.

Figure 28. Long-term monitoring of Older People: Sleep-related events

Two papers are envisaged to report the results of this year research to scientific community, one describing the new version of the event monitoring system, and a second one for the developed behavioral classification model. As a publication of this year we hightlight the paper in partnership with Alexandra Konig and Philippe Robert - entitled Validation of an automatic video monitoring system for the detection of instrumental activities of daily living in dementia patients - in the Journal of Alzheimer disease where we summarize the results of the validation of our event monitoring system for the recognition of activities of daily living of participants of Alzheimer's disease study.

6.20. Uncertainty Modeling Framework for Constraint-based Event Detection in Vision Systems

Participants: Carlos F. Crispim-Junior, François Brémond.

keywords: description-based activity recognition, uncertainty modeling, vision system, older people

Event detection has advanced significantly in the past decades relying on pixel- and feature-level representations of video-clips. Although effective, those representations have difficulty on incorporating scene semantics. Alternatively, ontology and description-based approaches for event modeling can explicitly embed scene semantics, but the deterministic nature of such languages is susceptible to noise from underlying components of vision systems. We have developed a probabilistic framework to handle uncertainty on our constraint-based ontology framework for event detection. This task spans from elementary scenarios uncertainty handling (from low-level data and event intra-class variance) to complex scenario semantic modeling, where time ordering in between event sub-components and the effect of missing components (for instance, due to miss-detection) plays a significant role.

Preliminary results of this work have been published in [40], where the presented formalism for elementary event (scenario) uncertainty handling is evaluated on the detection of activities of daily living of participants of the Alzheimer's disease study of Nice hospital using the newest version of our vision system using a RGB-D sensor (Kinect®, Microsoft©) as input. Two evaluations have been carried out: the first one, (a 3-fold cross-validation) focuses on elementary scenario constraint modeling and recognition, and the second one was devoted to complex scenario recognition following a semi-probabilistic approach (n:45).

Table 7 presents the performance of the uncertainty modeling framework on elementary scenario (primitive state) detection for $N : 10\ participants; 15\ min.\ each; Total : 150\ min$. The 3-fold cross-validation scheme (n:10 participants) is employed for constraint probabilistic distribution learning and event detection evaluation on 10 RGB-D recordings of participants of the Nice hospital clinical protocol for Alzheimer's disease study. "Crisp" term stands for our deterministic constraint-based ontology language for event modeling. Results are reported as the average performance on the crisp and uncertainty frameworks on the validation sets. Results confirm that the uncertainty modeling improves the detection of elementary scenarios in recall (*e.g.,* In zone phone: 84 to 100 %) and precision indices (*e.g.,* In zone Reading: 54.5 to 85.7%).

Table 7. Framework Performance on Elementary Scenario Detection on a 3-fold-cross-validation scheme

	Crisp		Uncertainty	
IADL	Rec.	Prec.	Rec.	Prec.
In zone Pharmacy	100.0	71.4	100	**83.3**
In zone Phone	84.0	95.45	**100.0**	**100.0**
In zone Plant	100.0	81.8	100.0	81.8
In zone Tea	93.3	**77.7**	93.3	73.7
In zone Read	75.0	54.5	75.0	**85.7**

Table 8 presents the performance of the proposed framework on Composite Event Detection for $N : 45\ participants; 15\ min.\ each; Total : 675min$. Here a hybrid strategy is adopted where the uncertainty modeling is used for elementary scenarios and the crisp constraint-based framework is used for composite event modeling. Results show improvement on recall index of event detection performance, but the uncertainty framework performance on precision index is still worse than the crisp approach. The latter performance may be attributed to the crisp constraints that did not have their uncertainty addressed yet.

Future work will focus on modeling complex scenario constraints such as time ordering and missing components, and on extending the set of low-level uncertainties which are addressed. Moreover, we have been conducting a joint work with partners of Dem@care project to evaluate the uncertainty framework for multiple sensor fusion at decision level. Currently, processed data from different visual modalities (standard RGB, RGBD, and wearable cameras) have been gathered for 17 participants of Nice hospital pilot@lab, and preliminary results are expected for the first semester of 2015.

Table 8. Framework Performance on Composite Event Detection Level

IADL	Crisp		Uncertainty	
	Rec.	Prec.	Rec.	Prec.
Talk on Phone	88.76	89.77	**88.76**	85.86
Preparing Tea/Coffee	81.42	40.36	**92.85**	**55.08**
Using Pharmacy Basket	87.75	95.65	**89.79**	**97.77**
Watering plant	78.57	**84.61**	**100.0**	28.86

6.21. Assisted Serious Game for Older People

Participants: Minh Khue Phan Tran, François Brémond, Philippe Robert.

keywords: interactive system, elderly people, serious game

A system able to interact with older people has been recently devised. The system consists of two parts: Recognition and Interaction. Recognition part, requiring an Asus Xtion Pro Live Camera, consists in observing the scene to decide when is the best moment to interact with users. Afterwards, the Interactive system tries to engage the patient via an interface and through Microsoft Kinect Camera, the patient can interact with the interface using voice or gesture. The interface is designed with Unity 3D game engine (see figure 29).
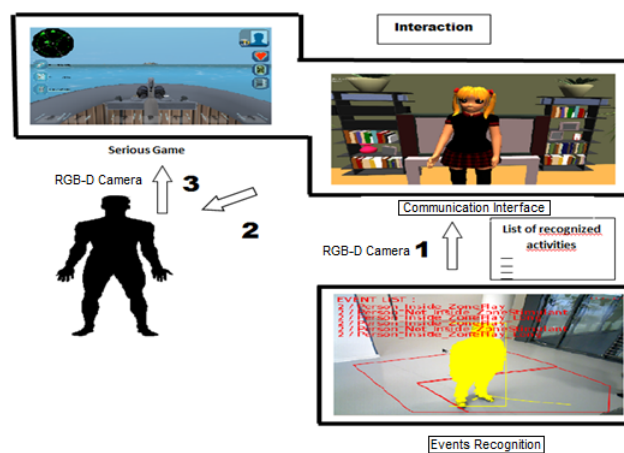

Figure 29. Functional diagram of our interactive system

An experiment was conducted in a memory center for older people, Institut Claude Pompidou in Nice, in order to test different functionalities of the system. Here, participants can experiment the system in a private room (see figure 30) equipped with a large screen and can start the game without having to use devices (mouse, keyboard). The "best moments" to interact with participants are defined when they stay more than 5 seconds in front of the screen. Once these moments are recognized, the interface of Interactive part is called. The avatar indicates the place to be for playing and starts the game.

19 older people have participated to the experiment. 16 succeeded to follow the indications of avatar up to the start of the game. Most of them have appreciated the interaction with the avatar. Even 13 of them have continued to play the second game after suggestion of the avatar.

Future work aims at looking at other indicators (behavior, gaze) that the system can rely on to improve user's interaction.
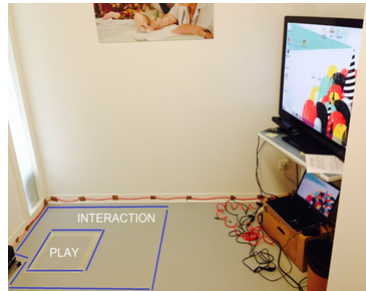
Figure 30. Experimental private room

6.22. Enhancing Pre-defined Event Models Using Unsupervised Learning

Participants: Serhan Coşar, François Brémond.

keywords: Pre-defined activity models, unsupervised learning, tailoring activity models

In this work, we have developed a new approach to recognize human activities from videos, given models that are learned in an unsupervised way and that can take advantage of *a priori* knowledge provided by an expert of the application domain. The description-based methods use pre-defined models and rules to recognize concrete events. But, if the data has unstructured nature, such as daily activities of people, the models cannot handle the variability in data (e.g., the way of preparing meal is person dependent).

In order to overcome this drawback, we have combined the description-based method in [66] with an unsupervised activity learning framework, as presented in Figure 31 . We have created a mutual knowledge loop system, in which both frameworks are combined in a way to compensate their individual limitations. In [66], scene regions are pre-defined and the activity models are created via defining an expected duration value (e.g., 2 seconds) and a posture type (e.g., standing) by hand. Thus, these hand-crafted models fail to cover the variability in data and require an update by experts whenever the scene or person changes. To automatically define these parameters, we utilize the unsupervised activity recognition framework. The unsupervised approach first learns scene regions (zones) in the scene using trajectory information and then, it learns the duration and posture distribution for each zone. By matching the pre-defined zones with learned zones, we connect the learned parameter distributions with hand-crafted models.

The knowledge is passed in a loopy way from one framework to another one. By knowledge we mean: (i) the geometric information and scene semantics of the description-based system are used to label the zones that are learned in an unsupervised way, (ii) the activity models that are learned in an unsupervised way are used to tune the parameters (i.e. tailoring) in the activity models of the description-based framework. It is assumed that the person detection and tracking are already performed and we have the trajectory information of people in the scene beforehand.

We have tested the performance of the knowledge-loop based framework on two datasets: i) Hospital-RGB, ii) Hospital-RGBD. Each dataset contains one person performing everyday activities in a hospital room. The activities considered in the datasets are "watching TV", "preparing tea", "answering phone", "reading newspaper/magazine", "watering plant", "organizing the prescribed drugs", "writing a check at the office desk" and "checking bus routes in a bus map". Each person is recorded using RGB and RGBD cameras of 640×480 pixels of resolution. RGB dataset consists of 41 videos and RGBD dataset contains 27 videos. For each person, video lasts approximately 15 minutes.
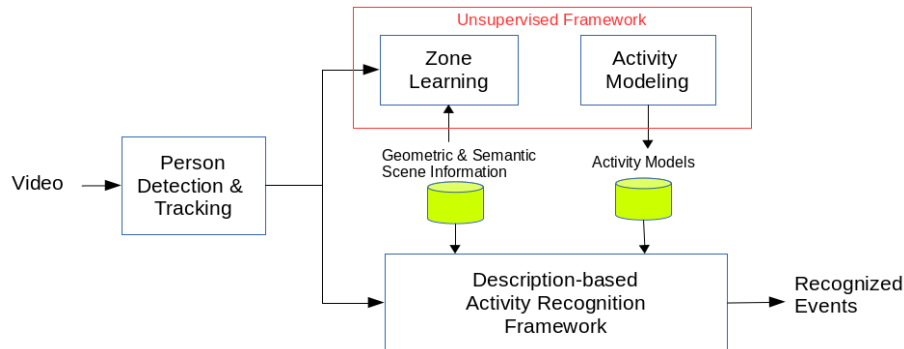
Figure 31. The flow diagram of the framework that combines pre-defined activity models with unsupervised models.

The performance of the approach in [66] with hand-crafted models and our approach with learned models for Hospital-RGB and Hospital-RGBD datasets are presented in Table 9 and in Table 10 . The results have been partially presented in Ellomiietcv2014 (**waiting hal acceptation**). It can be clearly seen that updating the constraints in activity models using data learned by the unsupervised approach enables us to detect activities missed by the pre-defined models. For "watching TV" and "using pharmacy basket" activities in RGB dataset and "answering phone" and "preparing tea" activities in RGBD dataset, there is increase in false positive rates. The reason is that, for some activities, the duration and posture distributions learned by the unsupervised approach can be inaccurate because of other actions occurring inside a zone (e.g., a person standing inside tea zone and reading). For this reason, the constraints updated in activity models are too wide and other activities that occur inside the zone are also detected. Despite the small increase of false positives in some activities, we have increased the true positive rates and obtained sensitivity rates around 90% and 87% on average in RGB and RGBD datasets, respectively, and precision rates around 81% on average in RGBD dataset. Thanks to the distributions learned for time duration and posture, we can enhance the activity models that are manually defined in the description-based, and thereby detect missed events.

Table 9. The activity recognition results of the description-based approach and our knowledge-loop based approach for the Hospital-RGB. The bold values represent the best result for each activity class.

ADLs	Hand-crafted Models		Unsupervised Models	
	Sensitivity (%)	Precision (%)	Sensitivity (%)	Precision (%)
Answering Phone	70	82.35	**95**	**90.47**
Watching TV	84.61	78.57	**100**	54.16
Using Office Desk	91.67	47.82	91.67	**52.38**
Preparing Tea	**80.95**	70.83	76.19	**80**
Using Phar. Basket	100	**90.90**	100	76.92
Watering Plant	100	81.81	88.89	**88.89**
Reading	45.46	83.34	**81.82**	**90**
TOTAL	81.81	**76.52**	**90.5**	76.11

6.23. Using Dense Trajectories to Enhance Unsupervised Action Discovery

Participants: Farhood Negin, Serhan Coşar, François Brémond.

keywords: zone learning, action descriptors, dense trajectories, supervised action recognition, unsupervised activity recognition

Table 10. The activity recognition results of the description-based approach and our knowledge-loop based approach for the Hospital-RGBD datasets. The bold values represent the best result for each activity class.

ADLs	Hand-crafted Models		Unsupervised Models	
	Sensitivity (%)	Precision (%)	Sensitivity (%)	Precision (%)
Answering Phone	80	**100**	**84.21**	88.89
Watching TV	55.56	45.46	**77.78**	**58.34**
Preparing Tea	100	**73.68**	92.85	65
Using Phar. Basket	100	90	100	**100**
Watering Plant	40	66.67	**83.34**	**71.42**
Reading	**100**	66.67	71.42	**83.34**
Using Bus Map	50	71.42	**100**	**100**
TOTAL	75.07	73.41	**87.08**	**80.99**

The main purpose in this work is to monitor older people in an unstructured scene (e.g., home) and to recognize the types of activities they perform. We have extended the work in Ellomiietcv2014 that was basically an unsupervised method to learn behavioral patterns of individuals without restraining subjects to follow a predefined activity model. The main concern in previous work is to find different zones in the scene where activities take place (scene topology) by employing trajectory information provided by tracking algorithm. The previous work in Ellomiietcv2014 (**waiting hal acceptation**) proposes a Hierarchical Activity learning Model (HAM) to learn activities based on previously identified topologies. The current work examines the same potential while first, incorporating image descriptors [93] in a bag-of-word representation to differentiate actions in a supervised manner and second, combining the two approaches (supervised and unsupervised) to provide clues about actions inside each zone by classifying retrieved descriptors using a classifier.

Recently, dense trajectories are widely used for action recognition and have been shown state-of-the-art performance [93]. For the purpose of the current work, we use HOG and HOF descriptors for supervised action recognition. Figure 32 shows a general description of the supervised framework. For the learning phase, the dense trajectories are extracted from input images coming from RGBD camera. Following Ellomiietcv2014, three-level topology of the scene is constructed by trajectory information coming from tracking algorithm [62]. The topology is used to split input video stream into chunks by checking where the person is with respect to the learned zones. Then, for every video chunk, dense descriptors are extracted and stored. A codebook representation is obtained by applying a k-means clustering algorithm on the whole set of extracted features. Next, the action histograms are calculated by employing the codebook. A SVM classifier is trained and stored to use in test phase via calculated histograms.
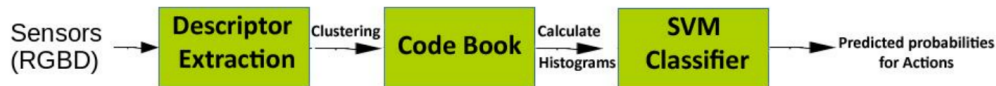
Figure 32. Flow diagram for supervised action recognition.

In recognition phase, we similarly split the test videos by comparing each trajectory point with learned topologies, extract the descriptor for each split, and the histograms are calculated via k-NN using the codebook generated in learning phase. Then, the histograms are classified using the trained SVM classifier and resulting labels are evaluated by comparing with the ground truth.

We have assessed the performance of the supervised activity recognition framework using 183 video splits of 26 subjects. We divided the video dataset to training and testing groups. Training set includes 93 videos of 15 subjects and the test set includes 90 videos of 11 subjects. Notice that the number of videos is counted after splitting process has been done on input data. We used the videos recorded from CHU Nice hospital while real patients are visiting their doctors and are asked to perform several activities in specified locations of the room. The activities we considered in our tests include: "preparing tea", "watching TV", "using phone", "reading on chair", "using pharmacy", and "using bus map". For RGB-D camera, we have used the person detection algorithm in [79] and tracking algorithm in [62]. The classification results for using HOG and HOF descriptors and corresponding confusion matrices are depicted in Table 11 and in Table 12 . For SVM classifier, we used RBF kernel.

Table 11. Confusion matrix for recognition results for HOG descriptor

Activity Names	1	2	3	4	5	6
1 Watching TV	11	0	0	0	0	0
2 Preparing Tea	0	18	0	0	0	0
3 Reading in Chair	1	0	10	0	0	0
4 Using Bus Map	0	0	0	14	0	0
5 Using Pharmacy Basket	0	0	0	0	10	0
6 Using Phone	0	0	0	0	0	25
Total	98.89%					

Table 12. Confusion matrix for recognition results for HOF descriptor

Activity Names	1	2	3	4	5	6
1 Watching TV	4	1	1	4	0	1
2 Preparing Tea	0	5	0	5	0	8
3 Reading in Chair	1	0	2	4	0	4
4 Using Bus Map	0	0	0	13	0	1
5 Using Pharmacy Basket	1	0	0	0	9	0
6 Using Phone	0	1	1	5	0	19
Total	57.78%					

As a future work, we are going to benefit from the action descriptors to discriminate different activities occurring in the same zone.

6.24. Abnormal Event Detection in Videos and Group Behavior Analysis

Participants: Giuseppe Donatielo, Vania Bogorny, Serhan Cosar, Luis Campos Alvares, Carolina Garate, François Brémond.

keywords: activity recognition, abnormal events, group behavior analysis, trajectory clustering

This work addresses two different issues: (i) abnormal event detection and (ii) group behavior analysis in videos.

6.24.1. Abnormal Event Detection

For abnormal event detection we are proposing a fused approach that combines trajectory-based and pixel-based analysis. In this work we first discover the activity zones based on object trajectories, and we investigate abnormal events considering objects that move in wrong direction and/or with abnormal speed. Second, inside each zone we extract dense tracklets and using the clustering technique we discover different types of actions, and are able to distinguish between normal and abnormal actions inside each zone.

While existing approaches for abnormal behavior detection do either use trajectory based or pixel based methods, we propose a fused solution which can detect simple abnormal behavior based on speed and direction, as well as more complex behavior as abnormal activities. In a first step we automatically learn the zones of the scene where most activities occur, by taking as input the trajectories of detected mobiles, analyzing then statistical information of each mobile in each zone (speed and direction), through the use of a scale-resolution analysis. This approach implies a considerable complexity decrease of having huge data set and then an extensive impact of the algorithm speed, without losing useful information. Figure 33 shows an example of this first part.

Figure 33. Example of Trajectories (left), Trajectories over the grid (center), that represents a given scale resolution, Zones discovering (right)

The next step concerns a pixel based analysis inside each zone. This step takes as input each zone computed in the previous step and the bounding box of the object trajectories, and extracts action descriptors inside the bounding box of each object trajectory inside the zone. With this step we obtain the different body movements of each detected mobile inside a zone. By clustering the body motions and using Bags of Words, we detect different types of abnormal activities inside each zone. Figure 34 shows an example of what just mentioned.

The last step of our approach is a clustering operation of all information gathered in the previous two steps, that is for each mobile, speed, direction, and body movements-actions in each zone are applied to discriminate between different types of abnormal behavior in the scene. A flow diagram of our approach is presented in Figure 35 .

We have tested our approach on several real videos recorded. We show with experiments on two open datasets that our approach is able to detect several types of abnormal behavior.

6.24.2. Group Behavior Analysis

Group behavior analysis is focused on the extraction of groups based on object trajectories and the analysis is performed over the dense tracklets, computed for the groups bounding boxes. From the analysis of the dense tracklets we detect different levels of agitation. These works are ongoing and have not yet been published.
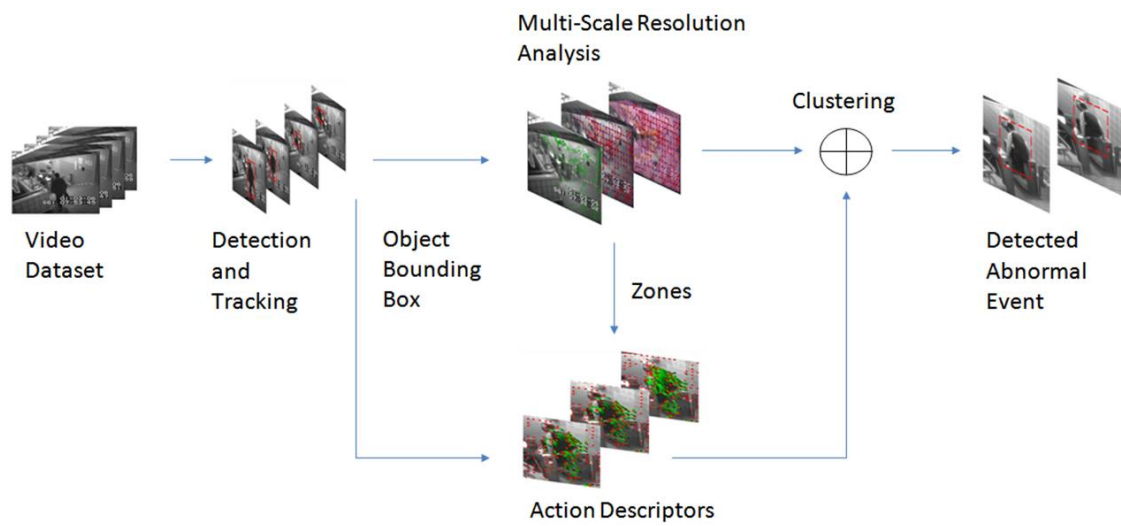
Figure 34. Tracked object

Figure 35. Flow Diagram of our approach

6.25. Model-Driven Engineering for Activity Recognition Systems

Participants: Sabine Moisan, Jean-Paul Rigault, Luis Emiliano Sanchez.

We continue to explore the applicability of model driven engineering (MDE) to activity recognition systems. Of course, setting up a complete methodology is a long term objective.

6.25.1. Feature Models

Features models are convenient representations of system variability but the drawback is a risk of combinatorial explosion of the number of possible configurations. Hence we have extended feature models with quality attributes and associated metrics to facilitate the choice of an optimal configuration, at deployment as well as at run time. We have proposed several strategies and heuristics offering different properties regarding optimality of results and execution efficiency [41].

This year we have conducted further experiments to evaluate the optimization algorithm and the metrics. In particular, we studied the prediction accuracy of the additive metrics for estimating two properties of interest: frame processing time and reconfiguration time. The goal was to compare predicted against measured properties of a running system. We used a simple video chain implemented with OpenCV components (acquisition, filtering, various detections, and visualisation) and we tested it on a video sample of 48s (i.e., about 1350 frames). We defined a feature model for this chain, which exhibits 14 valid configurations.

We first computed the properties of each component in isolation (based on a set of repetitive measurements), then we measured the actual frame processing time and reconfiguration time, and finally we compared the estimated and actual values.
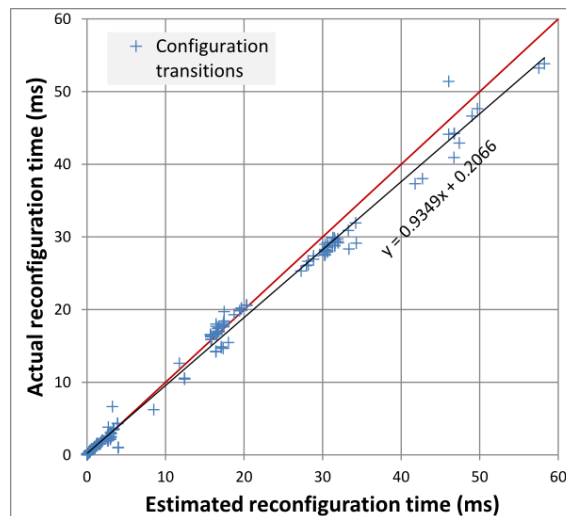
Figure 36. Measured and estimated reconfiguration time. Black line represents regression curve and red line the ideal trend

Figure 36 displays the results for reconfiguration time. In our experiment, we have a total of 210 transitions over a set of 15 valid configurations (14 system configuration + one for system shutdown). For frame processing time, we achieved an accuracy of 96.7% on average, and for reconfiguration time the accuracy was between 90.5% and 87.6%.

Introducing metrics in feature models is a precious help to reduce the huge set of valid configurations after a dynamic context change and to provide a real time selection of an appropriate running configuration. However, more evaluation remains to be done on other runtime properties and for other video chains.

6.25.2. Configuration Adaptation at Run Time

To react to environment changes we favor the "model at run-time" approach. Our current prototype ressorts to tools and technologies which were readily available. This made possible a proof of concepts. However, this induces several redundant representations of the same data, consistency problems, coexistence of several formalisms or programming languages, and superfluous back and forth inter-module communications. This year we started to design a more homogeneous and better integrated prototype. The two key points are, first, a component management framework, second, a feature model managemen tool.

This year, we have continued to develop our OSGi-like component framework, but more adapted to real time and compatible with our extended feature models. Concerning feature model management, we started to study a FAMILIAR replacement that integrates smoothly into the C++ video analysis chain instead of being a separate (Java) tool. Moreover, the new tool should cope with our feature extensions (e.g., quality attributes).

6.26. Scenario Analysis Module

Participants: Annie Ressouche, Sabine Moisan, Jean-Paul Rigault, Daniel Gaffé, Omar Abdalla.

Keywords: Synchronous Modelling, Model checking, Mealy machine, Cognitive systems.

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. 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 [80] 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.

Setting up our tools on top of an existing language such as LUSTRE was convenient for rapid prototyping. However, it appeared delicate for efficiency reasons on the one hand, but also because it is a closed environment, difficult to customize. Hence we developed our own language LE and its environment CLEM (see section 6.27). This year, we focus on the expression of scenario models in CLEM through the internship of Omar Adballa [51] and we define in CLEM a specific back end to generate recognition engines (see figure 37). However, mastering all aspects of this environment will allow the user scenario description language to rely directly on the semantics of LE and not on its syntax. This reduces the number of necessary translations.

Currently, SAM implements an "exact" algorithm in the sense that it generates, at each instant, *all* possible scenario instances although many of them will freeze, still holding system resources. We have started scalability studies to evaluate the risk of combinatorial explosion. In parallel we enriched the synchronous scenario descriptions to reduce the number of generated scenario instances as well as the number of instances to awake at each instant. We are currently modifying our recognition engine generator to take advantage of this supplementaty information.

6.27. The Clem Workflow

Participants: Annie Ressouche, Daniel Gaffé, Mohamed Bouatira, Ines Sarray.

Keywords: Synchronous languages, Synchronous Modelling, Model checking, Mealy machine.

This research axis concerns the theoretical study of a synchronous language LE with modular compilation and the development of a toolkit (see Figure 37) 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 focus on the improvement of both LE language and compiler concerning data handling and in the generation of back-ends required by other research axis of the team (see 6.26 and 6.28). We also improve the design of a new simulator for LE programs which integrates our new approach.

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). Last year, we chose an algebra which is a bilattice and we showed that it is well suited to solve our problem. To compute the partial orders of equation systems, we introduced two ways : a CPM (Critical Path Method) like algorithm, efficient and a "fix point" approach that allows us to show that we can compute partial orders locally and in an incremental way, or globally (thanks to the uniqueness if fix points). We introduced "abstract" equation systems and our method allows us to compute their partial orders. We defined a new intermediate format LEA (see figure 37) to record these sorted abstract eaquation systems and they will be expanded latter in concrete equation systems with a refinement operation. We apply this technique to the compilation of Grafcet language. Our work in under publication in two journal papers.

In CLEM, we added types and data a few years ago and this year we complete this addition, we know can express automata where control and output signals are valued. From last year, we rely on CLEM both to design SAM (see section 6.26) and to perform validation in a component-based middleware (see section 6.28). To this aim, we generate now two specific output formats dedicated to these applications [54]

Finally, in CLEM, we generate an independent intermediate code (LEC) before specific target generations. This code represents the semantics of programs with 4-valued equation systems. In our design flow, we need to simulate programs at this level. This year, we complete a simulator begun last year but which did not integrate the data part of the language. The simulator GUI has been designed again in Qt and the simulator takes into account the values carried by signals. This work has been done by Mohamed Bouatira during his internship.

6.28. Multiple Services for Device Adaptive Platform for Scenario Recognition

Participants: Annie Ressouche, Daniel Gaffé, Ines Sarray, Jean-Yves Tigli.

Keywords: Synchronous Modelling, Model checking, Mealy machine, Ubiquitous Computing.

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 [91], [90], [73]) 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 behavioral model of the critical component and model checking techniques apply to verify safety properties concerning this critical component. Thus, we consider that the critical component is validated.
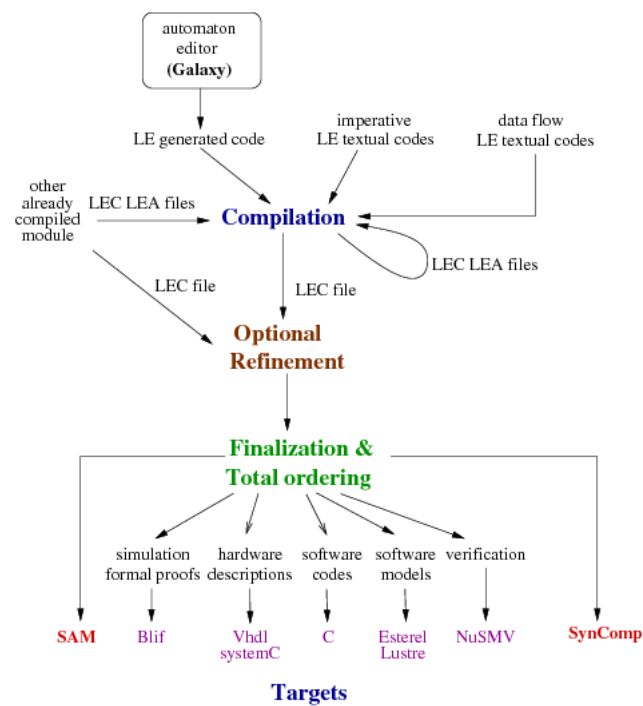
Figure 37. The CLEM *toolkit*

In [83], [82], we showed that an efficient means to define the synchronous components which allow to validate critical component behaviours, is to specify them with Mealy machines. Previously, we used a classical synchronous language (Lustre) to specify synchronous components, but the integration of the synchronous component code into WComp was not straightforward because Lustre compiler is not opened and cannot integrate new target code needed by WComp. This year, we rely on CLEM to design synchronous monitor as CLEM automata and we extend CLEM to generate automatically the internal code of WComp(C#).

On another hand, for some critical components, we can be led to introduce several synchronous monitors, each of them being related to a sub assembly. Then, some outputs of these synchronous monitors can be linked to the same input of a critical component. To face this problem, we introduced in [83], [82] a composition under constraints operation, which composed all the synchronous monitors linked to a critical component according to a set of constraints telling us how the respective outputs of monitors which access the same input are combined. We proved that this operation preserved safety properties, however it cannot ensure adaptivity and incrementality. This year, we have introduced a new way to compose synchronous monitors. We introduce an additional monitor (called constrain monitor) which express as a CLEM Mealy machine (set of equations) the constraints and we perform a usual parallel composition of all the monitors. Moreover, we define a syntactic mean (language DCL) to describe constraints in a generic way and we can derive automatically the constraint monitor for each specific application. In such a setting, we are able to manage the appearance and the desappearance of synchronous monitors.

Moreover, in middleware as WComp, communication is asynchronous while synchronous monitors respect the synchrony paradigm. Thus, we need to introduce in the middleware a means to transform asynchronous events into synchronous entries for synchronous monitors and the opposite to connect again the synchronous events into asynchronous design. To this aim, we introduced in WComp, specific components that receive asynchronous events and generate synchronous ones. Indeed, the part of this component is to decide how asynchronous events will be gather into a synchronous instant (logical time). To this aim, the user can choose between different politics. Then, all the events composing an instant are serialized and deserialized by the synchronous monitor. The desynchronisation operation performs the opposite way[54].

<div style="text-align:center">

TITANE Project-Team

</div>

6. New Results

6.1. Analysis

6.1.1. *Parametric Object Detection in Large Scenes*
Participant: Florent Lafarge.

Point processes are a natural extension of Markov Random Fields (MRF), designed to handle parametric objects. They have shown efficiency and competitiveness for tackling object extraction problems in vision. Simulating these stochastic models is however a difficult task. The performances of the existing samplers are limited in terms of computation time and convergence stability, especially on large scenes. We propose a new sampling procedure based on a Monte Carlo formalism [8]. Our algorithm exploits the Markovian property of point processes to perform the sampling in parallel. This procedure is embedded into a data-driven mechanism so that the points are distributed in the scene in function of spatial information extracted from the input data. The performances of the sampler are analyzed through a set of experiments on various object detection problems from large scenes, including comparisons to the existing algorithms. The sampler is also tested as optimization algorithm for MRF-based labeling problems.

6.2. Reconstruction

6.2.1. *Indoor Scene Reconstruction*
Participants: Sven Oesau, Florent Lafarge, Pierre Alliez.

In collaboration with EADS ASTRIUM

We contributed a method for automatic reconstruction of permanent structures of indoor scenes, such as walls, floors and ceilings, from raw point clouds acquired by laser scanners [6]. Our approach employs graph-cut to solve an inside/outside labeling of a space decomposition. To allow for an accurate reconstruction the space decomposition is aligned with permanent structures. A Hough Transform is applied for extracting the wall directions while allowing a flexible reconstruction of scenes. The graph-cut formulation takes into account data consistency through an inside/outside prediction for the cells of the space decomposition by stochastic ray casting, while favoring low geometric complexity of the model. Our algorithm produces watertight reconstructed models of multi-level buildings and complex scenes.

6.2.2. *State of the Art in Surface Reconstruction from Point Clouds*
Participant: Pierre Alliez.

In collaboration with Matthew Berger, Andrea Tagliasacchi, Lee Seversky, Joshua Levine, Andrei Sharf and Claudio Silva.

The area of surface reconstruction has seen substantial progress in the past two decades. The traditional problem addressed by surface reconstruction is to recover the digital representation of a physical shape that has been scanned, where the scanned data contains a wide variety of defects. While much of the earlier work has been focused on reconstructing a piece-wise smooth representation of the original shape, recent work has taken on more specialized priors to address significantly challenging data imperfections, where the reconstruction can take on different representations – not necessarily the explicit geometry. This state-of-the-art report surveys the field of surface reconstruction, providing a categorization with respect to priors, data imperfections, and reconstruction output. By considering a holistic view of surface reconstruction, this report provides a detailed characterization of the field, highlights similarities between diverse reconstruction techniques, and provides directions for future work in surface reconstruction [11].

6.2.3. Robust Shape Reconstruction and Optimal Transportation
Participants: Simon Giraudot, Pierre Alliez.

In collaboration with David Cohen-Steiner.

We describe a framework for robust shape reconstruction from raw point sets, based on optimal transportation between measures, where the input point sets are seen as distribution of masses. In addition to robustness to defect-laden point sets, hampered with noise and outliers, our approach can reconstruct smooth closed shapes as well as piecewise smooth shapes with boundaries [10].

6.3. Approximation

6.3.1. Zometool Shape Approximation
Participants: Henrik Zimmer, Florent Lafarge, Pierre Alliez.

In collaboration with Leif Kobbelt.

We contributed an algorithm that approximates 2-manifold surfaces with Zometool models while preserving their topology. Zometool is a popular hands-on mathematical modeling system used in teaching, research and for recreational model assemblies at home. This construction system relies on a single node type with a small, fixed set of directions and only nine different edge types in its basic form. While being naturally well suited for modeling symmetries, various polytopes or visualizing molecular structures, the inherent discreteness of the system poses difficult constraints on any algorithmic approach to support the modeling of freeform shapes. We contribute a set of local, topology preserving Zome mesh modification operators enabling the efficient exploration of the space of 2-manifold Zome models around a given input shape. Starting from a rough initial approximation, the operators are iteratively selected within a stochastic framework guided by an energy functional measuring the quality of the approximation. We demonstrate our approach on a number of designs and also describe parameters which are used to explore different complexities and enable coarse approximations [15].

6.3.2. CGALmesh: a Generic Framework for Delaunay Mesh Generation
Participants: Pierre Alliez, Clement Jamin.

In collaboration with Jean-Daniel Boissonnat and Mariette Yvinec.

CGALmesh is the mesh generation software package of the Computational Geometry Algorithm Library (CGAL). It generates isotropic simplicial meshes – surface triangular meshes or volume tetrahedral meshes – from input surfaces, 3D domains as well as 3D multi-domains, with or without sharp features. The underlying meshing algorithm relies on restricted Delaunay triangulations to approximate domains and surfaces, and on Delaunay refinement to ensure both approximation accuracy and mesh quality. CGALmesh provides guarantees on approximation quality as well as on the size and shape of the mesh elements. It provides four optional mesh optimization algorithms to further improve the mesh quality. A distinctive property of CGALmesh is its high flexibility with respect to the input domain representation. Such a flexibility is achieved through a careful software design, gathering into a single abstract concept, denoted by the oracle, all required interface features between the meshing engine and the input domain. We already provide oracles for domains defined by polyhedral and implicit surfaces [5].

6.3.3. Level-of-Detail Quad Meshing
Participant: David Bommes.

In collaboration with Hans-Christian Ebke and Leif Kobbelt from RWTH Aachen.

The most effective and popular tools for obtaining feature aligned quad meshes from triangular input meshes are based on cross field guided parametrization. These methods are incarnations of a conceptual three-step pipeline: (1) cross field computation, (2) field- guided surface parametrization, (3) quad mesh extraction. While in most meshing scenarios the user prescribes a desired target quad size or edge length, this information is typically taken into account from step 2 onwards only, but not in the cross field computation step. This turns into a problem in the presence of small scale geometric or topological features or noise in the input mesh: closely placed singularities are induced in the cross field, which are not properly reproducible by vertices in a quad mesh with the pre- scribed edge length, causing severe distortions or even failure of the meshing algorithm. We reformulate the construction of cross fields as well as field-guided parameterizations in a scale-aware manner which effectively suppresses densely spaced features and noise of geometric as well as topological kind. Dominant large-scale features are adequately preserved in the output by relying on the unaltered input mesh as the computational domain [4].

6.3.4. *Mesh Watermarking based on a Constrained Optimization Framework*
Participants: Xavier Rolland-Nevière, Pierre Alliez.

In collaboration with Technicolor and Gwenael Doerr.

A watermarking strategy for triangle surface meshes consists in modifying the vertex positions along the radial directions, in order to adjust the distribution of radial distances and thereby encode the desired payload. To guarantee that watermark embedding does not alter the center of mass, prior work formulated this task as a quadratic programming problem. We contribute a generalization of this formulation with: (i) integral reference primitives, (ii) arbitrary relocation directions to alter the vertex positions, and (iii) alternate distortion metrics to minimize the perceptual impact of the embedding process. These variants are evaluated against a range of attacks and we report both improved robustness performances, in particular for simplification attacks, and improved control over the embedding distortion [9].

6.3.5. *Robust 3D Watermarking*
Participants: Xavier Rolland-Nevière, Pierre Alliez.

In collaboration with Technicolor, thesis co-advised by Pierre Alliez and Gwenael Doerr.

3D models are valuable assets widely used in the industry and likely to face piracy issues. This dissertation deals with robust mesh watermarking that is used for traitor-tracing. Following a review of state-of-the-art 3D watermarking systems, the robustness of several content adaptation transforms are evaluated. An embedding domain robust against pose is investigated, with a thickness estimation based on a robust distance function to a point cloud constructed from some mesh diameters. A benchmark showcases the performance of this domain that provides a basis for robust watermarking in 3D animations. For static meshes, modulating the radial distances is an efficient approach to watermarking. It has been formulated as a quadratic programming problem minimizing the geometric distortion while embedding the payload in the radial distances. This formulation is leveraged to create a robust watermarking framework, with the integration of the spread-transform, integral reference primitives, arbitrarily selected relocation directions and alternate metrics to minimize the distortion perceived. Benchmarking results showcase the benefits of these add-ons w.r.t the fidelity vs. robustness watermarking trade-off. The watermark security is then investigated with two obfuscation mechanisms and a series of attacks that highlight the remaining limitations. A resynchronization approach is finally integrated to deal with cropping attacks. The resynchronization embeds landmarks in a configuration that conveys synchronization information that will be lost after cropping. During the decoding, this information is blindly retrieved and significant robustness improvements are achieved [2].

6.3.6. *Spread Transform and Roughness-based Shaping to Improve 3D Watermarking based on Quadratic Programming*
Participants: Xavier Rolland-Nevière, Pierre Alliez.

In collaboration with Technicolor and Gwenael Doerr.

Modulating the distances between the vertices and the center of mass of a triangular mesh is a popular approach to watermark 3D objects. Prior work has formulated this approach as a quadratic programming problem which minimizes the geometric distortion while embedding the watermark payload in the histogram of distances. To enhance this framework, we introduce two watermarking components, namely the spread transform and perceptual shaping based on roughness information. Benchmarking results showcase the benefits of these add-ons with respect to the fidelity-robustness trade-off [13].

WIMMICS Project-Team

6. New Results

6.1. Highlights of the Year

BEST PAPERS AWARDS :

[52] **IEEE/WIC/ACM International Conference on Intelligent Agent Technology, IAT**. A. G. B. TETTAMANZI, C. DA COSTA PEREIRA.

[45] **10th IEEE International Conference Beyond Databases, Architectures, and Structures (BDAS 2014)**. T. H. H. NGUYEN, N. LE THANH.

6.2. Argumentation Theory

6.2.1. *Combining Argumentation Theory and Normative Reasoning with Natural Language Processing*

Participants: Serena Villata, Elena Cabrio, Fabien Gandon.

We have proposed a methodology to identify and classify the semantic relations holding among the possible different answers obtained for a certain query on DBpedia language specific chapters. The goal is to reconcile information provided by language specific DBpedia chapters to obtain a consistent results set. The results of this research have been published at the LREC conference [29]. This classification has then been exploited in another work, together with Elena Cabrio and Alessio Palmero Aprosio (FBK Trento, Italy), where Serena Villata has worked on an extension of QAKiS, the system for open domain Question Answering over linked data, that allows to query DBpedia multilingual chapters. Such chapters can contain different information with respect to the English version, e.g. they provide more specificity on certain topics, or fill information gaps. In particular, she extended the results presented last year embedding the new identified relations among the different answers, using argumentation theory to reconcile the information and further improving the system's performances. A demo of the new argumentation module is available online [0]. This work has also been presented at the International Semantic Web Conference demo session [85].

Moreover, we have proposed, together with Alessio Palmero Aprosio, a system called NLL2RDF to translate in an automated way licenses, such as GPL, in natural language into a machine-readable version using the RDF language. The system is available online [0]. The results of this research have been presented at the European Semantic Web conference [26].

Finally, we have published the benchmark of natural language arguments called NoDE. The benchmark is available online [0]. The results of this research have been presented at the 15th International Workshop on Non-Monotonic Reasoning (NMR 2014) [28] and at the 5th Conference on Computational Argumentation [27] (COMMA 2014 - demo).

6.2.2. *Argumentation and Legal Reasoning*

Participant: Serena Villata.

---

[0] http://qakis.org/qakis2/
[0] http://www.airpedia.org/nll2rdf/
[0] http://www-sop.inria.fr/NoDE/

Together with Leendert van der Torre (University of Luxembourg), we proposed a framework for reasoning about norms using argumentation theory. Norms regulate our everyday life, and are used to assess the conformance of our behavior with respect to the regulations holding in specific contexts. Given the profound importance of norms in our lives, it is fundamental to understand which norms are valid in certain environments, how to interpret them, the legal conclusions of such norms, which norms can be derived from the existing ones, etc. In order to understand norms, people discuss about them to assess the validity or applicability of a certain norm subject to particular conditions, to derive the obligations and permissions to be enforced, or claim that a certain normative conclusion cannot be derived from the existing regulations. Several frameworks have been proposed for legal argumentation, but no comprehensive formal model of legal reasoning from arguments has been proposed yet. The goal of this work is to enrich legal argumentation with a formal account of deontic modalities. These results have been published at the 5th Conference on Computational Argumentation [55] (COMMA 2014).

Moreover, together with Guido Boella (University of Torino, Italy), Pietro Baroni and Massimiliano Giacomin (University of Brescia, Italy), Federico Cerutti (University of Aberdeen, UK), Leendert van der Torre (University of Luxembourg), we have studied also the dynamics of argumentation framework and this research has lead to a publication in the Artificial Intelligence journal [15].

6.3. License for the Web of Data

6.3.1. *Reasoning about rights and licenses in the Web of Data*
Participant: Serena Villata.

In the domain of Linked Open Data a need is emerging for developing automated frameworks able to generate the licensing terms associated to data coming from heterogeneous distributed sources. Together with Guido Governatori (NICTA, Australia) and Antonino Rotolo (University of Bologna, Italy), Serena Villata proposed and evaluated a deontic logic semantics which allows to define the deontic components of the licenses, i.e., permissions, obligations, and prohibitions, and generate a composite license compliant with the licensing items of the composed different licenses. The approach is evaluated using the SPINdle defeasible reasoner, where the proposed heuristics have been hard coded in the reasoner. The prosecution of this research line has seen the analysis of the compatibility of a set of licensing terms (always using SPINdle), the analysis of the role of licenses associated to vocabularies, and the development of the Licentia suite of services to reason over licenses and help users to deal with such kind of information. The results of this research line have been published at the International Semantic Web Conference demo session [87].

6.4. Logic and Cognition

6.4.1. *Logical Foundations of Cognitive Agents*
Participants: Andrea Tettamanzi, Serena Villata.

Together with Célia da Costa Pereira of I3S, we have investigated syntactic belief revision operators [16] and goal-generation mechanisms [30] to make the practical implementation of a general BDI (Belief-Desire-Intention) model of agency based on possibility theory. Furthermore, we took part in a joint investigation with a research team, led by Cristiano Castelfranchi, of the CNR-ISTC in Rome on the issue of trust in multi-agent systems [21]. We also employed agent-based simulation to test a theory of human stupidity proposed by the late Italian economist Carlo Cipolla [52]; our paper won the Best Paper Award at IAT 2014.

6.4.2. *RDF Mining*
Participants: Andrea Tettamanzi, Catherine Faron Zucker, Fabien Gandon.

We carried on our investigation in an approach to RDF mining based on grammatical evolution and possibility theory, whose aim is to mine large RDF graphs by automatically generating and testing OWL 2 axioms based on the known facts. In particular, we addressed the problem of testing candidate OWL 2 axioms against the fact contained in an RDF base and proposed a novel scoring heuristics based on falsification and possibility theory [53].

6.4.3. Sentiment Analysis
Participant: Andrea Tettamanzi.

Together with Célia da Costa Pereira of I3S and Mauro Dragoni of FBK, Trento, who visited our team for three months from April to June 2014, we have proposed a novel approach to concept-level sentiment analysis based on fuzzy logic. Our system [31], [89] participated in the Semantic Web Evaluation Challenge (SemWebEval) at ESWC 2014 and was the winner for Task 1 and Most Innovative Approach.

6.4.4. Data and Knowledge Integration and Extraction
Participant: Andrea Tettamanzi.

Together with Somsack Inthasone, Nicolas Pasquier and Célia da Costa Pereira of I3S, we developed a data warehouse collecting data for research on biodiversity [38].

6.4.5. Miscellaneous
Participant: Andrea Tettamanzi.

A work on electorcardiographic signal classification using evolutionary algorithms and neural networks carried out while still at the University of Milan, got published as a book chapter [65].

6.5. Natural Language Processing

6.5.1. Bridging Natural Language Processing and the Web of Data for Multimedia Question Answering
Participants: Elena Cabrio, Fabien Gandon, Yoann Moise.

Differently from search engines, the goal of Question Answering (QA) is to return precise answers to users' natural language questions, extracting information from both documentary text and advanced media content. Up to now, QA research has largely focused on text, mainly targeting factual and list questions. The goal of our work was instead to exploit structured data and metadata describing multimedia content on Linked Open Data to provide a richer and more complete answer to the user, combining textual information with other media content.

We implemented an extension of our QAKiS system [0] to boost the answer visualization adding multimedia content. More specifically, once QAKiS outputs the textual answer(s) to the question asked by the user, the user can click on "more details" to have further information on the retrieved entity. Three main types of additional (and multimedia) content are then displayed: *i)* additional textual information providing a description of the retrieved entity (extracted from DBpedia), and a structured Information Card containing a set of relevant properties of the entity; *ii)* images (extracted from Flickr) and relevant videos (extracted from YouTube); *iii)* entity geo-localization: a pointer on a map is shown for questions asking about a place, together with its points of interest. The results of this research have been published at ESWC 2014 - Demo/poster paper [84].

6.5.2. SMILK - Social Media Intelligence and Linked Knowledge
Participants: Elena Cabrio, Fabien Gandon, Fabrice Jauvat.

Automated Natural Language Processing (NLP), Web Open Data (Linked Open Data) and social networks are the three topics of the SMILK ANR LabCom including their coupling studied in three ways: texts and Linked Data, Linked Data and social resources, texts and social resources. The purpose of this LabCom is indeed to develop research and technologies on the one hand, retrieve, analyze, and reason about linking data from textual Web resources and other to use open Web data taking into account the social structures and interactions in order to improve the analysis and understanding of textual resources.

---

[0]http://qakis.org

As a first step in this direction, during the internship of Fabrice Jauvat we have developed a prototype of a system that - given free text (in particular in the cosmetics domain, extracted from a forum, a magazine, or a Web site) - can first recognize the named entities launching in parallel the RENCO system (developed by our partner in the LabCom), and NERD [0], and then connect them to DBpedia, so that additional information on the entity can be extracted. For instance, if the word "J'adore" is detected in the text, it is recognized as a named entity and connected to its DBpedia page, so that information about the fact that it is a perfume, that its brand is Dior, and so on, can be automatically extracted and shown to the user.

6.5.3. *Ontology-Based and Natural Language Chatbot System in the Commercial Domain*
Participants: Amine Hallili, Catherine Faron Zucker, Elena Cabrio, Fabien Gandon.

This work is done within a Cifre PhD Thesis colocated in the Wimmics team and with SynchroNext Company located in Nice. The work consists in modelling and implementing ontology-based natural language Chatbot in commercial domain which consists of

- The design of a commercial knowledge base using the websites' APIs and web services (e.g. Amazon1, eBay2, BestBuy3),
- Interpreting and handling links between users' natural language questions by constructing relational graph,
- Generation and visualization of textual and media answers.

A. Hallili attended the ESSLLI summer school where a poster was accepted [93].

6.5.4. *Editor of Formal Lexicographic Definitions*
Participants: Maxime Lefrançois, Romain Gugert, Alain Giboin, Fabien Gandon.

Last year a prototype of a GUI of an editor of formal dictionary definitions aimed at lexicographers was developed based on the formalism of Units Graphs and on Meaning-Text Theory. This year, the prototype was demonstrated during the IC 2014 conference [60]. The prototype was also described in a paper reporting the knowledge engineering methodology for representing lexicographic definitions it supports [39].

6.6. Collaborative Software Development Platforms
Participant: Isabelle Mirbel.

Today's Web has given rise to several platforms serving the purpose of collaborative software development. Thanks to these environments, it is possible, among others, for anyone to suggest new requirements for a software under development. A lot of requirements are thus proposed by users and it becomes difficult, after a while, for the persons in charge of the software which development is hosted by the platform to understand this large set of new requirements in its entirety. Therefore we proposed a tool to make large sets of requirements posted on collaborative software development platforms better workable despite the poor content of requirement body. Our aim was to propose an approach to automatically group similar requirements together in order to propose a limited number of requirement categories, thus improving the review process. As requirements expressed on collaborative software development platforms are usually very short and their content not very structured, we proposed to exploit relationships between stakeholders and already processed requirements to break the whole set of new requirements into meaningful categories. Our tool relies on Semantic Web languages and Formal Concept Analysis to provide a 3 steps data analysis process. The data is first extracted from the platform and translated into RDF, then stakeholders' past activities are analyzed to finally get stakeholder categories in order to improve the review of newly posted requirements.

According to the experiments that we conducted, we noticed some limitations in our approach. When the contributing stakeholders are newbies with no previous participation in any blueprint or bug and when there is no sufficient number of evaluated blueprints or bugs. To cope with this limitation, we plan to evaluate stakeholders reputation by looking at their activities on the whole collaborative software development platform (and not only the project under consideration). The results of this research have been published in [23].

---

[0]http://nerd.eurecom.fr/

In today's software development methodologies, User Stories (US) are mostly used as primary requirements artifacts. They are used to express requirements from a final user point of view and at a low abstraction basis using natural language. Over the years, several informal templates have been proposed by agile methods practitioners or academics to guide requirements gathering. Consequently, these templates are used in an ad-hoc manner, each modeler having idiosyncratic preferences. In this context, we performed a study of templates found in literature in order to propose a unified model. We also proposed an RDFS translation of this model in order to allow the annotation of user stories, thus providing search and reasoning capabilities to agile methods practitioners. The results of this research have been published in [56].

6.7. Workflow Management Systems

6.7.1. Semantic Mappings with a control flow-based business workflow

Participants: Thi Hoa Hue Nguyen, Nhan Le Thanh.

The aim of this PhD work is to improve Coloured Petri Nets (CPNs) and Ontology engineering to support the development of business process and business workflow definitions of various fields. To realize this objective, in the first time, we propose an ontological approach for representing business models in a meta-knowledge base. We introduce four basic types of manipulation operations on process models used to develop and modify business workflow patterns. In the second time we propose a formal definition of semantic constraints and an $O(n3)$-time algorithm for detecting redundant and conflicting constraints. By relying on the CPN Ontology and sets of semantic constraints, workflow processes are semantically created. Finally, we show how to check the semantic correctness of workflow processes with the SPARQL query language [45], [20], [46].

6.7.2. Extraction Mechanisms and Semantic Enrichment of Short Messages in Social Networks

Participants: Amosse Edouard, Nhan Le Thanh.

In this PhD, the work is focused on text processing in social networks and the main objectives are focused on the analysis of the spatial aspect, context enrichment and spatiotemporal analysis of short text messages.

During the first half of the year, we have mainly worked on positioning the research subject beside the state of the art as well as determining relevant domain. After analyzing several works on short text analysis in many domains such as Semantic Web, Data Mining and Natural Language Processing, we have identified a lack in the representation of the spatial aspect. Indeed, the spatial properties of items shared among online communities can be seen on three different aspects: i) The location of the resources which can be identified by its URI/URL, ii) The producer's location, iii) The location related to the content of the messages.

Most existent works have considered as identical the producer's location and the event described by the content of the message, which can lead to wrong results in many cases. For example, a user can be in the United States while describing an event in Africa. The SIOC ontology is one of the most known for representing items shared among online communities; we have proposed an extension of this ontology in which the spatial aspects are clearly represented. However, there exists a big challenge in finding the relevant location that can be associated to the content of a message. We are currently working on an approach that combines NLP technics and GIS to identify the spatial location of an item by analyzing its content.

6.7.3. Ontology-Based Workflow Management Systems

Participants: Tuan Anh Pham, Nhan Le Thanh.

The main objective of the PhD work is to develop a Shared Workflow Management System (SWMS) using ontology engineering. Everybody can share a semi-complete workflow which is called *Workflow template*, and other people can modify and complete it to use in their system. This customized workflow is called *Personalized workflow*. The challenges of a SWMS is to be simple, easy to use, friendly with the user and not too heavy. But it must have all functions of a WMS. There are three major challenge in this work: How to allow the users to customize the workflow template to correspond to their requirements, but their changes must be compliance with the predefined rules in the workflow template? How to build an execution model to evaluate step by step a personalized workflow ?

6.7.4. Model Spatio-Temporal Dedicated Social Networks
Participants: Amel Ben Othmane, Nhan Le Thanh.

The research aims of the work are to: i) model spatio-temporal, dedicated social networks using semantic web models (ontologies) taking into account spatial, temporal, social and dedicated dimensions. ii) overcome limitations of traditional Recommender Systems and improve the quality of recommendation by exploiting context (time, location, goal, etc.) and social ties.

The following tasks, proposed in the first year planning, are completed or almost finished, and are highly relevant to the current work, despite the different initial overall aim:

- Literature Review Report in Social networks, Spatio-temporal networks, Dedicated networks, Activities modeling techniques, Recommender systems.
- Elaboration of the requirements of an "ideal system" and presentation of an initial approach. The approach named the 5ws approach tries to answer those five questions: who must do, what, when, where and why?
- Implementation of the approach with Protégé.
- Extend semantic sensor network ontology to meet our requirements. In fact we use this ontology to enrich data from sensors networks which are used to measure different metrics describing physical activities (speed, heart rates, distance, etc.).

The following developments are ongoing: Adaptation of recommender systems for activities recommendation and reusing multi-dimensional recommendation model.

6.8. Modeling Team Processes

6.8.1. Modeling and Assessing Coordination Processes
Participants: Alain Giboin, Isabelle Mirbel.

This work is done in collaboration with Pierre Robillard (Polytechnique Montréal).

Last year we worked on an assessment method of the quality of team dynamics based on a taxonomy of episodes of interactions encountered in software development teams [99] – the CoDyMA (Collaborative Dynamics Measurement and Analysis) method. Precisely, we proposed an analysis procedure of episodes based on the Formal Concept Analysis (FCA) approach. This year, we proposed to enrich the CoDyMA method with a procedure for assessing the quality of coordination interactions and the quality of coordination artifacts within a development team. The procedure is based on the "Coordinative Artifacts" Framework [101], [100].

6.8.2. Modeling Multimodal Grounding Processes in Design Teams
Participant: Alain Giboin.

This work is done with Aurore Defays (Université de Liège).

Grounding is the process used by participants to a collective activity to coordinate both the content and process of their communication to be successful [96]. Grounding is also defined as the process of elaborating and maintaining the Common Ground (i.e., mutual knowledge, mutual beliefs, and mutual assumptions) necessary to participants' mutual understanding [97]. Multimodal grounding is the process of grounding using several perceptual modalities. Last year we improved the methodology of analysis of multimodal grounding proposed in [98].

6.9. Semantic and Temporal Analysis of Online Communities
Participant: Zide Meng.

The objective of the OCKTOPUS ANR project is to increase the potential social and economic benefit of the large and quickly growing amounts of user-generated content, by transforming it into useful knowledge. Since user communities are the basic of user-generated content sites, we start with community detection problem, which is a fundamental research point in social network analysis. Based on the preliminary experience from the previous year, we made several progress this year and published the results in international conferences, specifically:

- Topic based interested group detection:
  By analyzing a dataset extracted from the popular question answer site "StackOverflow", we proposed a heuristic method to enrich questions' tag. We also introduced a tag tree based model to extract topics from questions' tags, then we used the detected topics to label users in order to detect interest groups. We conducted experiments on the dataset and compared with related method. Results show that the proposed method is much simple and fast. This work has been published in [43].

- Question Answer social media management
  We proposed a question answer social media system based on social network analysis and social media mining to manage the two main resources in question answer sites: users and contents. We also presented a vocabulary used to formalize both the level of interest and the expertise of users on topics. We tested QASM on a dataset extracted from the popular "StackOverflow" site. We showed how the formalized knowledge is used to find relevant experts for a question. This work has been published in [95].

6.9.1. *Temporal analysis in User and Topic:*

We are planning to introduce temporal analysis into our research problem. According to the previous work, the potential direction could be topic evolution and user interest evolution. We believe this work could benefit community management in question answer sites, for example topic trend detection or user interest management.

6.10. P2P Media Streaming

Participant: Gaspard Perrot.

The *Heave-Ho* project won the Inria 2014 *Boost Your Code* contest.

The goal of the project is to design an overlay network for P2P media streaming based on new HTML5 technologies such as WebRTC. While conventional Internet applications encounter problems with scaling up as the number of visitors grows, the Heave-Ho project aims to enable website's users to share the resource directly among them. The proposed solution is a perfect fit for real-time video broadcasting. In traditional server/client architecture the server can only handle a limited number of requests; if there are too many clients, some of them will not have access to the video. Using a P2P system, the video can be broadcasted to more clients. The use of sharing techniques based on user location can also cut data transfer costs directly at ISP level, thereby reducing the risk of problems such as data rate limits [0].

6.11. Discovery Hub

Participants: Nicolas Marie, Fabien Gandon, Emilie Palagi, Alain Giboin.

In the context of the Discovery Hub project [0], we performed an ergonomic evaluation and redesign of the graphical and textual User Interfaces (UIs) displaying, and allowing to interact with, the explanations provided by Discovery Hub to justify the results it has retrieved. We did user tests of the existing UIs and designed several mockups improving the UIs by taking users' feedback into account.

---

[0] https://www.inria.fr/en/content/view/full/61889
[0] http://www.discoveryhub.co

We also performed a user-centered evaluation of the quality of the results retrieved by the 4 algorithms of Discovery Hub. We decided to focus on the quality of the results and not on the UI. Thus, specific criteria of the quality of the results were defined in this evaluation: the surpriseness and the interestingness of the results. A result is considered as:

- surprising if the user discovers an unknown resource or relation between the topic searched and the selected result, or if she discovers something unexpected ;

- interesting if the user thinks it is similar to the topic explored or if she thinks she will remind or reuse it later on.

We are currently developing an ergonomic method for evaluating exploratory search systems (ESSs) in general. We are performing a first test of the method on Discovery Hub [41].

6.12. Knowledge Graphs

6.12.1. SPARQL Template Transformation Language

Participants: Olivier Corby, Catherine Faron-Zucker Song.

We finalized the design and implementation of SPARQL Template Transformation Language [0] (STTL) [58], [73]. STTL is an extension of SPARQL with a `template {} where {}` clause that enables a transformation to generate presentation format for RDF. For example, it is possible to generate Turtle, OWL or SPIN-to-SPARQL syntax.

We designed a new service in Corese server that returns HTML. Using STTL transformations that generate HTML, we are able to set up light weight Semantic Web servers on top of local RDF Datasets or remote Datasets such as DBpedia [0].

6.12.2. RDF Serialization and Introspection

Participant: Olivier Corby.

We started a work on RDF serialization of (Java) objects for Semantic Web system introspection. In conjunction with the overloading of SPARQL named graph pattern, we are able to query the system on several internal status such as graph index, triple provenance, property path triples, etc.

6.12.3. OWL 2 RL

Participant: Olivier Corby.

We dramatically optimized Corese Inference Rule Engine and we were able to run OWL 2 RL [0] rule base on the FMA ontology (Fundational Model of Anatomy [0]) with interesting performance. The initial OWL graph contains 1.74 million triples, the final graph contains 13.46 million triples and the rule engine runs in 3 minutes.

6.12.4. Rules for the Web of Data

Participants: Oumy Seye, Olivier Corby, Catherine Faron Zucker, Fabien Gandon.

This year we focused on the validation, the update of rules bases and the optimization of the reasonning. The goal of this work is to detect some inconsistencies in selected rule bases with respect to ontology and offer users to correct this. We built a set of SPARQL queries enabling (1) to build specific rule bases for a given context or application, (2) to optimize inference engines based on rule selection with respect to target RDF data sources, (3) to validate and update rule bases. We propose another optimization of the inference engines based on graph of rules dependencies and rules aplication ordering. This work is published in [62].

---

[0]http://ns.inria.fr/sparql-template/
[0]http://corese.inria.fr
[0]http://www.w3.org/TR/owl2-profiles/#OWL_2_RL
[0]http://sig.biostr.washington.edu/projects/fma/release/v3.2.1/alt_formats.html

6.12.5. KGRAM

Participants: Olivier Corby, Fuqi Song.

We received a two years grant from Inria to support the development of the Corese platform. This action aims at enhancing Corese software for conforming to latest W3C standards and facilitating its usage in distributed environment, we integrated several open source parsers to Corese, including JSON-LD, RDFa, TriG and N-Quads. Corese now is able to process RDF dataset in these formats. A Firefox extension called RDF Triple Collector (RTC) was developed, it can extract triple from web pages (annotated using RDFa), upload triples to Corese server and query data using SPARQL endpoint. A prototype of LDP 1.1 (Linked Data Platform) is implemented using RTC as data collector.

Besides, with the purpose of improving Corese query performance and carrying out research work on distributed environment, we proposed and developed a heuristic-based query planning method within Corese. The approach includes 3 main steps: 1) generate extended SPARQL query triple pattern Graph (ESG), 2) estimate the cost of ESG using pre-defined heuristics and cost models and 3) search ESG to find a good query plan and rewrite the SPARQL query. The approach was evaluated using BSBM benchmark, the results suggest that the developed method optimized 60% of the query execution time averagely [77].

6.13. Sociocultural Ontologies

Participants: Papa Fary Diallo, Olivier Corby, Isabelle Mirbel.

6.13.1. Sociocultural Ontology : Upper-level and Domain Ontologies

We propose a process of sociocultural ontology development in order to promise and preserve the culture of a country through sharing the customs and history of different localities. This can be compared with the construction of a platform straddling "corporate memory" and a "social network", but applied in the context of a country. This process is based on the Vygotskian Framework, a theory of Russian psychologist Lev Vygotsky. We worked on an upper-level ontology and mapped it on the Linked Open Data (LOD) cloud. We designed a sociocultural domain ontology for the Senegalese context and the platform design on top of Semantic Mediawiki (SMW). This allows Senegalese communities to share and co-construct their sociocultural knowledge. This work is published in [59].

6.13.2. Human Time Ontology

In the second step of the PhD thesis of P. F. Diallo, we focus on the consideration of the time in the modeled knowledge. The main objective of this work is to provide a vocabulary (ontology) to handle temporal information on semantic data. Thus, the first step was to create a meta-language which handles temporal knowledge representation in the socio-cultural field that can be used in a wider area. This meta-language allows 1) to model cyclic knowledge (non-convex interval), 2) knowledge about calendar, 3) convex intervals 4) modeling absolute and relative time, 5) modeling relations between intervals, 6) distinction between open and closed intervals, 7) concepts such as time stamps and 8) to set different time granularities. The second step is to propose an RDFS representation of this meta-language. Thus this representation, Human Time Ontology (HuTO), allows us to model complex time statement which are a date, an interval (convex and non-convex), relative and absolute time. HuTO allows also temporal data annotation which is the representation of temporal notions on knowledge (expressed as RDF triple) and allow to reason over it.

HuTO allows us to use a resource as a temporal marker for dating another resource. Our ontology allows the relative dating which is to determine the relative order of resources, without necessarily determining their absolute time. A major contribution of HuTO is the modeling of non-convex intervals but also requests that can treat all types of intervals. For temporal annotation data, HuTO provides an approach that can link two models: one for temporal information and another for knowledge of the modeled area. This approach facilitates the information retrieval when it is on temporal or non-temporal data. Thus HuTO can annotate resources, triples or named graphs.

ZENITH Project-Team

6. New Results

6.1. Highlights of the Year

- Patrick Valduriez received the 2014 Innovation Prize from Inria – Académie des sciences – Dassault Systems.
- Miguel Liroz-Gistau received the best presentation award from the Grid5000 Spring School 2014 in Lyon for his talk on "Using Grid5000 for MapReduce Experiments".
- Triton, a new common lab. (i-lab) has been created between Zenith and Beepeers (beepeers.com) to work on a platform for developing social networks in mobile/Web environments.
- 127 research groups worldwide registered to the LifeCLEF 2014 evaluation campaign chaired by Alexis Joly.

6.2. Big Data Integration

6.2.1. *Probabilistic Data Integration*

Participants: Reza Akbarinia, Naser Ayat, Patrick Valduriez.

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.

An important problem that arises in big data integration is that of Entity Resolution (ER). ER is the process of identifying tuples that represent the same real-world entity. The problem of *entity resolution over probabilistic data* (which we call ERPD) arises in many distributed application domains that have to deal with probabilistic data, ranging from sensor databases to scientific data management. The ERPD problem can be formally defined as follows. Let e be an uncertain entity represented by multiple possible alternatives, i.e., tuples, each with a membership probability. Let D be an uncertain database composed of a set of tuples each associated with a membership probability. Then, given e, D, and a similarity function F, the problem is to find the entity-tuple pair (t, t_i) (where $t \in e, t_i \in D$) such that (t, t_i) has the highest cumulative probability to be the most similar in all possible worlds. This entity-tuple pair is called the *most probable match pair* of e and D, denoted by MPMP(e, D).

Many real-life applications produce uncertain data distributed among a number of databases. Dealing with the ERPD problem for distributed data is quite important for such applications. A straightforward approach for answering distributed ERPD queries is to ask all distributed nodes to send their databases to a central node that deals with the problem of ER by using one of the existing centralized solutions. However, this approach is very expensive and does not scale well neither in the size of databases, nor in the number of nodes.

In [24], we propose an efficient solution for the ERPD problem. Our contributions are summarized as follows. We adapted the possible worlds semantics of probabilistic data to define the problem of ERPD based on both similarity and probability of tuples. We proposed a PTIME algorithm for the ERPD problem. This algorithm is applicable to a large class of the similarity functions, where the similarity score of two tuples depends only on their attributes i.e., context-free functions. For the rest of similarity functions (i.e., context-sensitive), we proposed a Monte Carlo approximation algorithm. We also proposed a parallel version of our Monte Carlo algorithm using the MapReduce framework. We conducted an extensive experimental study to evaluate our approach for ERPD over both real and synthetic datasets. The results show the effectiveness of our algorithms.

Another topic of interest is the integration of large astronomy data catalogs. The main challenge in such integration, besides the huge amount of catalog data to be merged, is the weak identification of sky objects, which leads to ambiguities in object matching amongst catalogs. In cite [30], we present the NACluster algorithm. NACluster considers a Euclidian metric space and distance function to drive disambiguation amongst objects in various catalogs and extends the traditional k-means algorithm to deal with the dynamic creation of new clusters, representing real sky objects. NACluster shows F-measure results steadily superior to the Q3C join operator matching results, which is its closest competitor.

6.2.2. *CloudMdsQL, a query language for heterogeneous data stores*
Participants: Carlyna Bondiombouy, Boyan Kolev, Oleksandra Levchenko, Patrick Valduriez.

The blooming of different cloud data management infrastructures, specialized for different kinds of data and tasks, has led to a wide diversification of DBMS interfaces and the loss of a common programming paradigm. The CoherentPaaS European project addresses this problem, by providing a common programming language and holistic coherence across different cloud data stores.

In this context, we have started the design of a Cloud Multi-datastore Query Language (CloudMdsQL), and its query engine. CloudMdsQL is a functional SQLlike language, capable of querying multiple heterogeneous data stores (relational and NoSQL) within a single query that may contain embedded invocations to each data store's native query interface. Thus, CloudMdsQL unifies a quite diverse set of data management technologies while preserving the expressivity of their local query languages. Our experimental validation, with three data stores (graph, document and relational) and representative queries, shows that CloudMdsQL satisfies the five important requirements for a cloud multidatabase query language.

6.2.3. *Semantic Data Integration using Bio-Ontologies*
Participants: Emmanuel Castanier, Patrick Valduriez.

Biologist have adopted ontologies for several reasons: (1) to provide canonical representation of scientific knowledge; (2) to annotate experimental data to enable interpretation, comparison, and discovery across databases; (3) to facilitate knowledge-based applications for decision support, natural language processing and data integration. The challenge is to automatically process complex databases and generate mappings using relevant ontologies in a way that scales up for many resources and ontologies, while being easy to use for the biomedical community, customizable to fit specific needs and smart, in order to leverage the knowledge contained in ontologies.

The National Center for Biomedical Ontology (NCBO) has developed a popular ontology-based annotation workflow. To address the above challenge, we have integrated the NCBO annotator with our WebSmatch tool and the Biosemantic tool from IRD to perform semantic annotation using bio-ontologies [47]. The resulting tool provides very useful capabilities. First, it can convert SQL database schemas to RDF/RDFS with Biosemantic. Second, it can annotate with the NCBO annotator and WebSmatch using the NCBO resources index. Third, the NCBO annotator relies on WebSmatch to create mappings between elements of schemas and ontological concepts, and uses ontologies properties (i.e. subsomption, transitivity) to enhance matching techniques.

Unlike the bio-medical domain which has accepted ontologies as a means to manage (integrate) knowledge, the agronomic sciences is yet to exploit its full potential. To this end, we are currently developing an RDF knowledge base, Agronomic Linked Data (AgroLD) [50]. The knowledge base is designed to integrate data from various publically available plant centric data sources. The aim of AgroLD project is to collaborate with domain experts in bridging the gap between technology and its potential users to enhance biological research.

6.3. Distributed Indexing and Searching

6.3.1. *Query Reformulation in P2P Data Management Systems*
Participant: Esther Pacitti.

We consider peer-to-peer data management systems (PDMS), where each peer maintains mappings between its schema and some acquaintances, along with social links with peer friends. In this context, we deal with reformulating conjunctive queries from a peer's schema into other peer's schemas. Precisely, queries against a peer node are rewritten into queries against other nodes using schema mappings thus obtaining query rewritings. Unfortunately, not all the obtained rewritings are relevant to a given query, as the information gain may be negligible or the peer is not worth exploring. On the other hand, the existence of social links with peer friends might be useful to get relevant rewritings.

In [19], we propose a new notion of "relevance" of a query with respect to a mapping that encompasses both a local relevance (the relevance of the query wrt. the mapping) and a global relevance (the relevance of the query wrt. the entire network). Based on this notion, we design a new query reformulation approach for social PDMS which achieves great accuracy and flexibility. We combine several techniques: (i) social links are expressed as FOAF (Friend of a Friend) links to characterize peer's friendship; (ii) concise mapping summaries are used to obtain mapping descriptions; (iii) local semantic views are special views that contain information about mappings captured from the network by using gossiping techniques. Our experimental evaluation, based on a prototype on top of PeerSim and a simulated network demonstrate that our solution yields greater recall, compared to traditional query translation approaches proposed in the literature.

6.3.2. *Diversified and Distributed Recommendation for Scientific Data*

Participants: Esther Pacitti, Maximilien Servajean.

Recommendation is becoming a popular mechanism to help users find relevant information in large-scale data (scientific data, web). Different diversification techniques have been proposed to avoid redundancy in the process of recommendation. Intuitively, the goal of recommendation diversification is to identify a list of items that are dissimilar, but nonetheless relevant to the user's interests.

The main goal of this work [39], [17] is to define a new diversified search and recommendation solution suited for scientific data (i.e., plant phenotyping, botanical data). We first propose an original profile diversification scoring function that enables to address the problem of returning redundant items, and enhances the quality of diversification compared to the state-of-the-art solutions. We believe our work is the first to investigate profile diversity to address the problem of returning highly popular but too-focused items.Through experimental evaluation using two benchmarks we showed that our scoring function presents the best compromise between diversity and relevancy. Next, to implement our new scoring function, we propose a Top-k threshold-based algorithm that exploits a candidate list to achieve diversification. However this algorithm is greedy and does not scale up well.To overcome this limitation, we propose several techniques to improve performance. First, we simplify the scoring model to reduce its computational complexity. Second, we propose two techniques to reduce the number of items in the candidate list, and therefore the number of diversified scores to compute. Third, we propose different indexing scores (i.e., the score used to sort the items in the inverted lists) that take into account the diversification of items, and using them, we developed an adaptive indexing approach to reduce the number of accesses in the index dynamically based on the queries workload. We evaluated the performance of our techniques through experimentation. The results show that they enable to reduce the response time up to 12 times compared to a baseline greedy diversification algorithm.

We also address the problem of distributed and diversified recommendation (P2P and multi-site) that fits very well in different application scenarios. We propose a new scoring function (usefulness) to cluster relevant users over a distributed overlay. We analyzed the new clustering algorithm in details, and we studied its behavior with an experimental evaluation using different datasets. Compared with state-of-the-art solutions, we obtain major gains in recall (order of 3 times).

6.4. Scientific Workflows

6.4.1. *Reuse of Scientific Workflows*

Participant: Sarah Cohen-Boulakia.

With the increasing popularity of scientific workflows, public and private repositories are gaining importance as a means to share, find, and reuse such workflows. As the sizes of workflows repositories grow, methods to compare the scientific workflows stored in them become a necessity, for instance, to allow duplicate detection or similarity search. Scientific workflows are complex objects, and their comparison entails a number of distinct steps from comparing atomic elements to comparison of the workflows as a whole. Various studies have implemented methods for scientific workflow comparison and came up with often contradicting conclusions upon which algorithms work best. Comparing these results is cumbersome, as the original studies mixed different approaches for different steps and used different evaluation data and metrics.

We first contribute to the field [26] by (i) comparing in isolation different approaches taken at each step of scientific workflow comparison, reporting on an number of unexpected findings, (ii) investigating how these can best be combined into aggregated measures, and (iii) making available a gold standard of over 2000 similarity ratings contributed by 15 workflow experts on a corpus of 1500 workflows and re-implementations of all methods we evaluated.

Then, we present a novel and intuitive workflow similarity measure that is based on layer decomposition [40]. Layer decomposition accounts for the directed dataflow underlying scientific workflows, a property which has not been adequately considered in previous methods. We comparatively evaluate our algorithm using our gold standard and show that it a) delivers the best results for similarity search, b) has a much lower runtime than other, often highly complex competitors in structure-aware workflow comparison, and c) can be stacked easily with even faster, structure-agnostic approaches to further reduce runtime while retaining result quality.

6.4.2. *Processing Scientific Workflows in Multi-site cloud*
Participants: Ji Liu, Esther Pacitti, Patrick Valduriez.

As the scale of the data increases, scientific workflow management systems (SWfMSs) need to support workflow execution in High Performance Computing (HPC) environments. Because of various benefits, cloud emerges as an appropriate infrastructure for workflow execution. However, it is difficult to execute some scientific workflows in one cloud site because of geographical distribution of scientists, data and computing resources. Therefore, a scientific workflow often needs to be partitioned and executed in a multisite environment.

In [46], we define a multisite cloud architecture that is composed of traditional clouds, e.g., a pay-per-use cloud service such as Amazon EC2, private data-centers, e.g. a cloud of a scientific organization like Inria, COPPE or LNCC, and client desktop machines that have authorized access to the data-centers. We can model this architecture as a distributed system on the Internet, each site having its own computer cluster, data and programs. An important requirement is to provide distribution transparency for advanced services (i.e., workflow management, data analysis), to ease their scalability and elasticity. Current solutions for multisite clouds typically rely on application specific overlays that map the output of one task at a site to the input of another in a pipeline fashion. Instead, we define fully distributed services for data storage, intersite data movement and task scheduling.

Also, SWfMSs generally execute a scientific workflow in parallel within one site. In [38], we propose a non-intrusive approach to execute scientific workflows in a multisite cloud with three workflow partitioning techniques. We describe an experimental validation using an adaptation of Chiron SWfMS for Microsoft Azure multisite cloud. The experiment results reveal the efficiency of our partitioning techniques, and their superiority in different environments.

6.4.3. *Data-centric Iteration in Dynamic Workflows*
Participant: Patrick Valduriez.

Dynamic workflows are scientific workflows supporting computational science simulations, typically using dynamic processes based on runtime scientific data analyses. They require the ability of adapting the workflow, at runtime, based on user input and dynamic steering. Supporting data-centric iteration is an important step towards dynamic workflows because user interaction with workflows is iterative. However, current support for iteration in scientific workflows is static and does not allow for changing data at runtime.

In [20], we propose a solution based on algebraic operators and a dynamic execution model to enable workflow adaptation based on user input and dynamic steering. We introduce the concept of iteration lineage that makes provenance data management consistent with dynamic iterative workflow changes. Lineage enables scientists to interact with workflow data and configuration at runtime through an API that triggers steering. We evaluate our approach using a novel and real large-scale workflow for uncertainty quantification on a 640-core cluster. The results show impressive execution time savings from 2.5 to 24 days, compared to non-iterative workflow execution. We verify that the maximum overhead introduced by our iterative model is less than 5% of execution time. Also, our proposed steering algorithms are very efficient and run in less than 1 millisecond, in the worst-case scenario.

6.5. Scalable Query Processing

6.5.1. *Big Data Partitioning*

Participants: Reza Akbarinia, Miguel Liroz, Esther Pacitti, Patrick Valduriez.

The amount of data that is captured or generated by modern computing devices has augmented exponentially over the last years. For processing this *big data*, parallel computing has been a major solution in both industry and research. This is why, the MapReduce framework, which provides automatic distribution parallelization and fault-tolerance in a transparent way over lowcost machines, has become one of the standards in big data analysis.

For processing a big dataset over a cluster of nodes, one main step is data partitioning (or fragmentation) to divide the dataset to the nodes. In [23], we consider applications with very large databases, where data items are continuously appended. Thus, the development of efficient data partitioning is one of the main requirements to yield good performance. In particular, this problem is harder in the case of some scientific databases, such as astronomical catalogs. The complexity of the schema limits the applicability of traditional automatic approaches based on the basic partitioning techniques. The high dynamicity makes the usage of graph-based approaches impractical, as they require to consider the whole dataset in order to come up with a good partitioning scheme. In our work, we propose *DynPart* and *DynPartGroup*, two dynamic partitioning algorithms for continuously growing databases [23]. These algorithms efficiently adapt the data partitioning to the arrival of new data elements by taking into account the affinity of new data with queries and fragments. In contrast to existing static approaches, our approach offers constant execution time, no matter the size of the database, while obtaining very good partitioning efficiency. We validate our solution through experimentation over real-world data; the results show its effectiveness.

6.5.2. *Scalable Query Processing with Big Data*

Participants: Reza Akbarinia, Miguel Liroz, Patrick Valduriez.

We address the problem of data skew in MapReduce parallel processing framework. There are many cases where because of skew intermediate data, a high percentage of processing in the reduce side of MapReduce is done by a few nodes, or even one node, while the others remain idle.There have been some attempts to address this problem of data skew, but only for specific cases. In particular, there is no solution when all or most of the intermediate values correspond to a single key, or to a set of keys that are fewer than the number of reduce workers.

In this work, we propose FP-Hadoop, a system that makes the reduce side of MapReduce more parallel, and can efficiently deal with the problem of reduce side data skew. We extended the programming model of MapReduce to allow the collaboration of reduce workers on processing the values of an intermediate key, without affecting the correctness of the final results. In FP-Hadoop, the reduce function is replaced by two functions: intermediate reduce and final reduce. There are three phases, each phase corresponding to one of the functions: map, intermediate reduce and final reduce phases. In the intermediate reduce phase, the intermediate reduce function, which usually includes the main load of reducing in MapReduce jobs, is executed by reduce workers in a collaborative way, even if all values belong to only one intermediate key. This allows performing a big part of the reducing work by using the computing resources of all workers,

even in the case of highly skewed data. We implemented a prototype of FP-Hadoop by modifying Hadoop's code, and conducted extensive experiments over synthetic and real datasets. The results show that FP-Hadoop makes MapReduce job processing much faster and more parallel, and can efficiently deal with skewed data. We achieve excellent performance gains compared to native Hadoop, e.g. more than 10 times in reduce time and 5 times in total execution time.

6.6. Data Stream Mining

6.6.1. *Summarizing Uncertain Data Streams*
Participants: Reza Akbarinia, Florent Masseglia.

In recent years, there has been a growing interest for probabilistic data management. In [41], we focus on probabilistic time series where a main characteristic is the high volumes of data, calling for efficient compression techniques. To date, most work on probabilistic data reduction has provided synopses that minimize the error of representation w.r.t. the original data. However, in most cases, the compressed data will be meaningless for usual queries involving aggregation operators such as SUM or AVG. We propose *PHA* (Probabilistic Histogram Aggregation), a compression technique whose objective is to minimize the error of such queries over compressed probabilistic data. We incorporate the aggregation operator given by the end-user directly in the compression technique, and obtain much lower error in the long term. We also adopt a global error aware strategy in order to manage large sets of probabilistic time series, where the available memory is carefully balanced between the series, according to their individual variability.

6.6.2. *An Anti-Bouncing Data Stream Model*
Participant: Florent Masseglia.

Usage mining is a significant research area with applications in various fields. However, Web usage data is usually considered streaming, due to its high volumes and rates. Because of these characteristics, we only have access, at any point in time, to a small fraction of the stream. When the data is observed through such a limited window, it is challenging to give a reliable description of the recent usage data. In [28] we show that data intralinkings, i.e., a usage record (event) may be associated with other records (events) in the same dataset, are common for Web usage streams. Therefore, in order to have a more authentic grasp of Web usage behaviors, the corresponding data stream models for Web usage streams should be able to process such intralinkings. We study the important consequences of the constraints and intralinkings, through the "bounce rate" problem and the clustering of usage streams. Then we propose the user-centric ABS (the Anti-Bouncing Stream) model which combines the advantages of previous models but avoids their drawbacks. First, ABS is the first data stream model that is able to seize the intralinkings between the Web usage records. It is also the first user-centric data stream model that can associate the usage records for the users in the Web usage streams. Second, owing to its simple but effective management principle, the data in ABS is available at any time for analysis. Under the same resource constraints as existing models in the literature, ABS can better model the recent data. Third, ABS can better measure the bounce rates for Web usage streams. We demonstrate its superiority through a theoretical study and experiments on two real-world data sets.

6.6.3. *Autonomic Intrusion Detection: Adaptively Detecting Anomalies over Unlabeled Audit Data Streams*
Participant: Florent Masseglia.

In [27], we propose a novel framework of autonomic intrusion detection that fulfills online and adaptive intrusion detection over unlabeled HTTP traffic streams in computer networks. The framework holds potential for self-managing: self-labeling, self-updating and self-adapting. Our framework employs the Affinity Propagation (AP) algorithm to learn a subject's behaviors through dynamical clustering of the streaming data. It automatically labels the data and adapts to normal behavior changes while identifying anomalies. Two large real HTTP traffic streams collected in our institute as well as a set of benchmark KDD'99 data are used to validate the framework and the method. The test results show that the autonomic model achieves better results in terms of effectiveness and efficiency compared to adaptive Sequential Karhunen-Loeve method and static AP as well as three other static anomaly detection methods, namely k-NN, PCA and SVM.

6.7. Scalable Data Analysis

6.7.1. *Retrieval of Large-scale Visual Entities*

Participants: Valentin Leveau, Alexis Joly, Patrick Valduriez.

In [37], we consider the problem of recognizing legal entities in visual contents in a similar way to named-entity recognizers for text documents. Whereas previous works were restricted to the recognition of a few tens of logotypes, we generalize the problem to the recognition of thousands of legal persons, each being modeled by a rich corporate identity automatically built from web images. We therefore introduce a new geometrically-consistent instance-based classification method that has several benefits over state-of-the-art instance classification methods including an efficient training phase reduced to a simple indexing process with a linear time and space complexity, but also the easy management of multi-labeled images, the fine grained localisation of the recognized patterns or the possibility of dynamically inserting additional training images in an incremental way. Experiments show that our method achieves better results than state-of-the-art techniques while being much more scalable, notably on an automatic web crawl of 5,824 legal entities which demonstrates the scalability of the approach.

6.7.2. *Content-based Life Species Identification in Large Multimedia Collections*

Participants: Alexis Joly, Julien Champ, Jean-Christophe Lombardo.

Building accurate knowledge of the identity, the geographic distribution and the evolution of living species is essential for a sustainable development of humanity as well as for biodiversity conservation. In this context, using crowdsourced data collection and multimedia identification tools is considered as one of the most promising solution. With the recent advances in digital devices/equipment, network bandwidth and information storage capacities, the production of multimedia data has indeed become an easy task. The emergence of citizen sciences and social networking tools has actually fostered the creation of large and structured communities of nature observers (e.g. e-bird, xeno-canto, Tela Botanica, etc.) who started to produce outstanding collections of multimedia records. Unfortunately, the performance of the state-of-the-art multimedia analysis techniques on such data is still not well understood and is far from reaching the real world's requirements in terms of identification tools. We therefore created LifeCLEF [36], [35], [31], [42], a new lab of the CLEF international forum [0] that evaluates these challenges in the continuity of the image-based plant identification task that we organized since 2011 within the ImageCLEF [0] lab. LifeCLEF is organized around 3 complementary tasks (PlantCLEF, BirdCLEF, FishCLEF), each being based on large and real-world data, as well as realistic scenarios established in collaboration with biologists and environmental stakeholders. 127 research groups worldwide did registered to the 2014 pilot campaign and downloaded the data. 22 of them crossed the finish line by submitting runs and papers to the workshop.

Besides the organization of the campaign, we also participated to two tasks in order to evaluate the content-based retrieval technologies developed within ZENITH. We notably implemented a new method [34] for the bird task based on the dense indexing of MFCC features and the offline pruning of the non-discriminant ones. To make such strategy scalable to the 30M of MFCC features extracted from the tens of thousands audio recordings of the training set, we used high dimensional hashing techniques coupled with an efficient approximate nearest neighbors search algorithm with controlled quality. Further improvements were obtained by (i) using a sliding classier with max pooling, (ii) weighting the query features according to their semantic coherence, and (iii) making use of the metadata to filter incoherent species. Results did show the effectiveness of the proposed technique which ranked 3rd among the 10 participating groups (some of them with years of experience in bioacoustic).

We finally investigated new interactive identification methods in [29], by extending classical faceted search mechanisms to the use of so called visual facets. The principle is to automatically build comprehensive visual illustrations of the expert data available in classical structured botanical dataset by building a visual matching graph of the related pictures and choosing the most connected ones. Additional facets can then be built automatically by clustering the graph and solving incompleteness issues.

---

[0]http://www.clef-initiative.eu/

[0]http://www.imageclef.org/

6.7.3. A look inside the Pl@ntNet experience

Participants: Alexis Joly, Julien Champ, Jean-Christophe Lombardo.

Pl@ntNet is an innovative participatory sensing platform relying on image-based plants identication as a mean to enlist non-expert contributors and facilitate the production of botanical observation data [22]. 18 months after the public launch of the iOS public application (and 6 months after the release of the Android version [32]), we carried out a self-critical evaluation of the experience with regard to the requirements of a sustainable and effective ecological surveillance tool (to appear in Multimedia Systems journal). Thanks to usage data analytics, we first demonstrated the attractiveness of the developed multimedia system (with more than 300K end-users and several thousands of users daily) as well as the nice self-improving capacities of the whole collaborative workflow (1.5 millions of observations were collected). We also pointed out the current limitations of the approach towards producing timely and accurate distribution maps of plants at a very large scale. We discussed in particular two main issues:

1. Data validation bottleneck: within the current workflow, only a few percentage of the observations are validated to avoid submerging the volunteer experts who actively do this job thanks to the collaborative web tools. There is consequently a need of smarter task assignment and recommendation mechanisms that would better balance the collaborative workload across all users and improves the serendipity.

2. Bias of the produced data: The temporal and geographical distribution of the observations is highly correlated with human activity. High densities of observations are more determined by population density and humans behavior than by plants density. This issue inevitably arises in any participatory sensing system but when the objective is to monitor noise nuisance or air quality, the concentration of the observations in the cities is less critical. There is therefore a need to build new data analytics methods compensating the bias through long-term statistics and the use of contextual information.