RESEARCH CENTER
Nancy - Grand Est

FIELD

Activity Report 2015

Section New Results

Edition: 2016-03-21

ALICE Project-Team

7. New Results

7.1. Dihedral Angle-Based Maps of Tetrahedral Meshes

Participants: Nicolas Ray, Bruno Lévy.

This work is a collaboration with Gilles-Philippe Paillé (visiting), Pierre Poulin (U. de Montréal) and Alla Sheffer (UBC).

Given a 2D triangulation, it is well known that it is reasonably easy to reconstruct the shape of all the triangles from the sole data of the angles at the triangle corners, provided that they satisfy some constraints. In this project, we studied how this idea can be generalized in the volumetric setting. In other words, we proposed a geometric representation of a tetrahedral mesh that is solely based on dihedral angles, and what are the constraints that these dihedral angles need to satisfy to make that possible. We first show that the shape of a tetrahedral mesh is completely defined by its dihedral angles. This proof leads to a set of angular constraints that must be satisfied for an immersion to exist in $\mathbb{R}^3$. This formulation lets us easily specify conditions to avoid inverted tetrahedra and multiply-covered vertices, thus leading to locally injective maps. We then present a constrained optimization method that modifies input angles when they do not satisfy constraints. Additionally, we develop a fast spectral reconstruction method to robustly recover positions from dihedral angles. We demonstrate the applicability of our representation with examples of volume parameterization, shape interpolation, mesh optimization, connectivity shapes, and mesh compression. This work has been published in Transactions on Graphics [17].

7.2. Hexahedral-dominant Remeshing

Participants: Dmitry Sokolov, Nicolas Ray, Bruno Lévy, Maxence Reberol.

Representing the geometry of complex objects in a computer is usually achieved by a mesh: the object is decomposed in cells that have a simple geometry. Each cell is defined by a set of facets. The simplest choice is to use meshes with tetrahedral cells that are relatively easy to produce and to work with. However, some applications involving numerical simulations better work with hexahedral cells. Such hexahedral meshes are very difficult to produce, even when it is completely done by a designer.

Our objective is to relax the intrinsic difficulties of full hexahedral remeshing by allowing the process to generate a few tetrahedra in the hexahedral mesh (hexahedral-dominant meshes). Our approach is a two steps procedure that defines the desired orientation of the hexahedra with a frame field, then integrates this frame field to generate a deformed 3D grid inside the object. Hexahedra are generated where the grid is non degenerated and not too distorted, and tetrahedra will fill the remaining volume.

7.2.1. Frame Field Generation

A frame field must define the orientation of a cube (the less deformed hexahedron) everywhere inside the object. This object is very difficult to manipulate because it has to be invariant by rotation of 90 degrees around each of its facet normal vector. In [26] we have designed a fast algorithm that is able to define a smooth frame field constrained to be aligned with the object boundary. We represent frames by spherical harmonics (as introduced in [33]) and greatly improve the state of the art thanks to an expression of the boundary constraints that keep the objective function of the optimisation problem very close to quadratic.

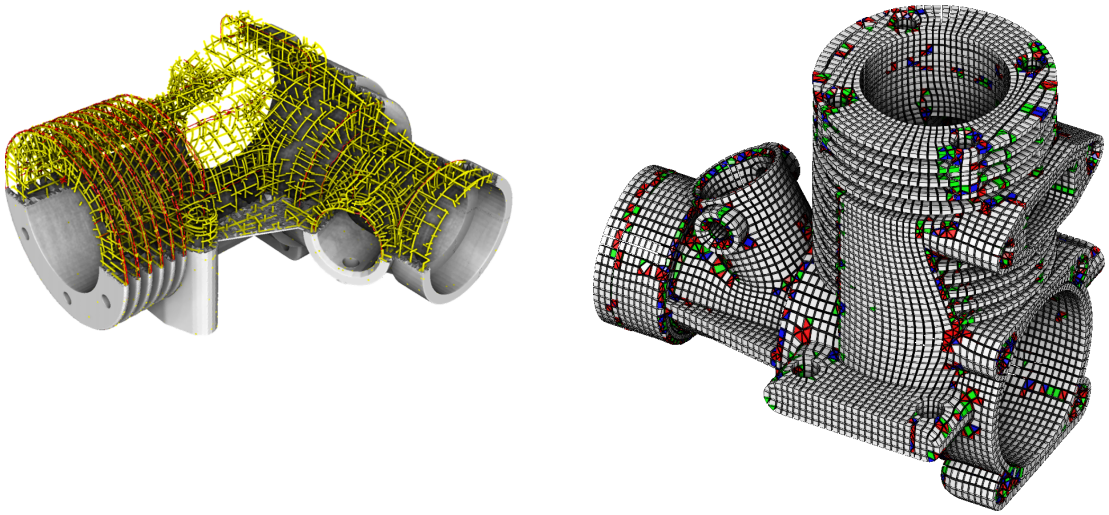

Figure 1. Our Hexahedral dominant meshes generation pipeline have two steps: the generation of a smooth frame field (yellow curves on the left image), and the construction of a mesh with hexahedron that aligns with the frame field.

7.2.2. Generation of Hexahedral-dominant Meshes

The generation of the hex dominant meshes is performed in two steps: place a deformed 3D grid inside the object such that it is aligned with the frame field, then use it to produce hexahedra and fill the rest of the volume with tetrahedra. We developed two solutions for the first step: a 3D extension of PGP [40] and an adaptation of Cubecover [39]. Both solutions have pros and cons so we plan to make them cooperate in a near future. The conversion of this result in an hexahedral-dominant mesh is a very complex problem for which we have a fair solution: we extract a point set from the deformed 3D grid, generate a tetrahedral mesh of the object that is constrained to include the point set in its vertices. From this tetrahedral mesh, we merge sets of tetrahedra into hexahedra with an extension of [37]. We are now working on an alternative solution that will generate hexahedra directly from the deformed 3D grid, and extract the boundary of the rest of the volume as a 2D mesh. From this mesh, we will try to produce more hexahedra by adapting existing combinatorial methods [27].

7.2.3. Impact on FEM Performance

It is admitted by our scientific community that hexahedral meshes are better than tetrahedral meshes for some FEM simulation. We would like to demonstrate evidence of this belief, including fair comparisons with equal running time and/or result accuracy, but the best function basis for each case. For hexahedral dominant meshes, we want to determine if the benefit of using hexahedra deserves having specific function bases devoted to properly link tetrahedral and hexahedral elements. We are developing a new function basis, tailored to non-conformal mixed hexahedra-tetrahedra meshes. Using a combination of tri-linear and quadratic hexahedra, it is possible to construct a space of continuous functions even on a non-conformal mesh. We are now proceeding to analyse the properties of this function space, both theoretically and experimentally. This topic is addressed in the (ongoing) Ph.D. thesis of Maxence Reberol.

7.3. Semi-discrete Optimal Transport in 3D

Participant: Bruno Lévy.

This work introduces a practical algorithm to compute the optimal transport map between a piecewise linear density and a sum of Dirac masses in 3D. In this semi-discrete setting, Aurenhammer *et al.* showed that the optimal transport map is determined by the weights of a power diagram [28]. The optimal weights are computed by minimizing a convex objective function with a quasi-Newton method. To evaluate the value and gradient of this objective function, we propose an efficient and robust algorithm that computes at each iteration the intersection between a power diagram and the tetrahedral mesh that defines the measure. Like in the multilevel proposed by Mérigot, we use a hierarchical algorithm, that uses nested point sets to discretize the source measure.
We think this work may lead to interesting discretizations of the physics, that include the conservation laws (conservation of energy, conservation of momentum ...) deep in their definition, as explained by Jean-David Benamou and Yann Brenier in their fluid dynamics formulation of optimal transport [30].

This work was published in the journal *Mathematical Modeling and Analysis* [10].

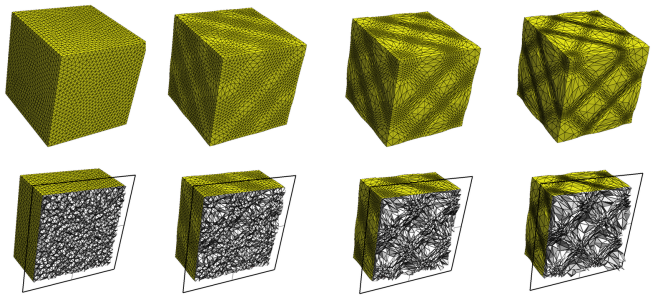

Figure 2. Semi-discrete optimal transport from a constant density to a varying one (product of sines).

7.4. By-example Synthesis of Structurally Sound Shapes

Participants: Jonas Martínez Bayona, Jérémie Dumas, Sylvain Lefebvre.

This is a collaboration with Li-Yi Wei (HKU) on a first project, An Lu (Inria/TU. Munich), Jun Wu (TU. Munich) and Christian Dick (TU. Munich) on a second project.

This work is at the heart of the ERC ShapeForge and considers the by-example synthesis of shapes under structural constraints. We considered two views of the problem that lead to different methodologies.

In a first approach, our goal is to cover a surface with a pattern – an operation akin to texturing in Computer Graphics. The pattern is however used to define the final shape, by determining which parts of the surface are solid or empty. The method operates on a thin voxel shell and does not require any parametrization of the input surface. The pattern is synthesized using a novel formulation for by-example pattern synthesis along surfaces. It is analyzed for structural weaknesses and this information is fed back to the pattern synthesizer, so that seamless reinforcements are added to the structure. We collaborated with researchers from T.U. Munich to analyze the structural behaviour of our structures, and developed a fast evaluation scheme that can be used within our optimization loop to guarantee structural soundness of the resulting design. The work was published in ACM Transactions on Graphics in 2015 [9] (proceedings of SIGGRAPH 2015).

In a second approach we considered the synthesis of shapes that are as rigid as possible under specific boundary conditions and using a prescribed amount of material, while resembling a given input example pattern, as illustrated in Figure 3 . Our method is inspired by the field of topology optimization, where rigid shapes are optimized but without any appearance constraints. Our algorithm generates shapes that resemble the input exemplar while being within a user specified percentage of the most rigid shape obtained without the

appearance objective. The work was published in ACM Transactions on Graphics in 2015 [12] (proceedings of SIGGRAPH Asia 2015).

Figure 3. Left. A chair automatically synthesized from a load scenario and an example pattern. The rigidity of the chair is within controlled bounds of a shape optimized without appearance objective. Right. A table design automatically synthesized.

7.5. Modeling for Fabrication

We pursued our research regarding automatic modeling techniques for fabrication, where an algorithm takes into account fabrication constraints to simplify the modeling process. This year we have worked on three projects in this area: the modeling of mechanisms from incomplete 2D definitions, the modeling of self-supporting tight enclosures to assist the fabrication process, and the interactive sculpting of support-free objects.

7.5.1. 3D Fabrication of 2D Mechanisms

Participants: Jean Hergel, Sylvain Lefebvre.

This project considered the automatic modeling of 3D mechanisms from an under-specified 2D model of the mechanism. Our approach casts the synthesis problem as an edge orientation problem in a graph, where graph nodes represent parts of the mechanisms and edges capture their interactions as analyzed by the 2D simulation of the mechanism. The edge orientation determines which parts include which others. Once all inclusions have been determined, we formulate a CSP to solve for the layering problem: each part is assigned a depth 'layer' in 3D. We finally compute the final geometry through CSG (boolean combinations of shapes). This work has been published in Computer Graphics Forum (proceedings of Eurographics 2015) [8]. It received an honorable mention from the best paper committee.

7.5.2. Self-supporting Tight Enclosures

Participants: Samuel Hornus, Sylvain Lefebvre, Frédéric Claux, Jérémie Dumas.

The aim of this project was to develop a technique to automatically generated a tight enclosure in the free space around an object. The challenge was to ensure that the enclosure stays close to the object and be as thin as possible while still being printable without collapsing. Such an enclosure finds at least two important applications : 1. as a protective skin to avoid artifacts when 3D-printing a multi-material object. 2. for generating as-large-as-possible cavities inside the printed object in order to minimized material usage and print time. The work is available as an Inria technical report [22].

7.5.3. Interactive Sculpting of Support-free Objects

Participants: Tim-Christopher Reiner, Sylvain Lefebvre.

Tim Reiner, former PhD student at the Karlsruhe Institute of Technology, joined the team on a Post-Doc position to explore new ideas in the context of modeling, rendering, and fabrication. Starting in March 2015, he developed a voxel-based environment for interactive modeling. In a research project together with Sylvain Lefebvre, our team has derived novel techniques for sculpting support-free 3D shapes. These shapes have the property that they do not require support structures during fabrication on fused deposition modeling or resin-based printers. This work is currently under review.

7.6. Intersection Detection via Gauss Maps; a Review and New Techniques

Participant: Samuel Hornus.

We have revisited the problem of deciding whether two convex objects intersect or not. A systematic view of the problem for polyhedra led us to a unified view of several techniques developed in the computer graphics community and to a new and very fast technique specialized to pairs of tetrahedra. A novel view of the problem as a minimization problem over the 2-sphere led us to the description of new interesting links between the set of planes separating two objects and the silhouette edges of their Minkowski difference. From there, we devised a new algorithm for separating two arbitrary convex objects that is a little bit faster and much more robust than the state of the art technique of Gilbert, Johnson and Keerthi [31]. The work has been summarized in [21].

7.7. Fractal Geometry

Participant: Dmitry Sokolov.

This is a collaboration with Christian Gentil (LE2I), Gilles Gouatay (LSIS), Anton Mishkinis (LE2I).

Additive manufacturing enables for the first time the physical realization of objects having complex geometries. Good approximations of fractals, in particular, can now be manufactured in a variety of materials, including metals. The application domains of fabricated fractal geometries are vast, from the be design of “fractal” micro-strip antennas, to the creation of meta-materials.

The main challenge with traditional fractals is the control of the resulting geometry. For example, it is quite challenging to get the desired shape using the system of fractal homeomorphisms proposed by Barnsley [29]. We elaborate here a new type of modeling system, using the facilities of existing CAGD software, while extending their capabilities and their application areas. This new type of modeling system will offer designers (engineers in industry) and creators (visual artists, stylists, designers, architects, etc.) new opportunities to design and produce a quick mock-up, a prototype or a single object. Our approach is to expand the possibilities of a standard CAD system by including fractal shapes while preserving ease of use for end users.

This year we published two papers on the subject [20], [16].

BIGS Project-Team

7. New Results

7.1. Stochastic modeling

7.1.1. Tumor growth modeling

Participants: P. Vallois, S. Wantz-Mézières
External collaborator: J-S. Giet (IECL,Université de Lorraine)

A cancer tumor can be represented for simplicity as an aggregate of cancer cells, each cell behaving according to the same discrete model and independently of the others. Therefore to measure its size evolution, it seems natural to use tools coming from dynamics of population, for instance the logistic model. This deterministic framework is well-known but the stochastic one is worthy of interest. We work with a model in which we suppose that the size V_t at time t of the tumor is a diffusion process of the type :

$$\begin{cases} dV_t = r\,V_t\,\left(1-\frac{V_t}{\kappa}\right) - c\,V_t + \beta\,V_t\,dB_t \\ V_0 = v > 0 \end{cases} \tag{1}$$

where $(B_t)_{t\geq 0}$ is a standard brownian motion starting from zero. Then (i) We define a family of time continuous Markov chains which models the evolution of the rate of malignant cells and approximate (under some conditions) the diffusion process (V_t). (ii) We study in depth the solution to equation (1). This diffusion process lives in a domain delimited by two boundaries: 0 and $\kappa > 0$. In this stochastic setting, the role of κ is not so clear and we contribute to understand it. We describe the asymptotic behavior of the diffusion according to the values of the parameters. The tools we resort to are boundary classification criteria and Laplace transform of the hitting time to biological worthwhile level. We are able in particular to express the mean of the hitting time. We have an accepted paper in the journal Theory of Stochastic Processes [70].

7.1.2. A Multitype Branching Process Model of Heterogeneous Damages in vitro Cancer Cell Populations Treated by Radiotherapy

Participants: T. Bastogne, P. Vallois
External collaborator: S. Pinel (CRAN, Université de Lorraine)

Cancer is the result of inter-dependent multi-scale phenomena and this is mainly why the understanding of its spread is still an unsolved problem. In integrative biology, mathematical models play a central role; they help biologists and clinicians to answer complex questions through numerical simulations and statistical analyses. The main issue here is to better understand and describe the role of cell damage heterogeneity and associated mutant cell phenotypes in the therapeutic responses of cancer cell populations submitted to a radiotherapy sessions during *in vitro* experiments. The cell heterogeneity is often described as randomness in mathematical modeling and different representations, such as Markov chains, branching processes and even stochastic differential equations, have been recently used. Conversely to these previous studies, which only focused on the steady-state responses of cell populations, we are interested by modeling the transient behavior after treatment and to identify the role of mutation heterogeneity in the global dynamic response of the cell populations. We propose to describe the survival response of anin vitro cancer cell culture treated by radiotherapy as a superposition of independent dynamics. Each cell is represented by a finite collection of cell mutation states with possible transitions between them. The population dynamics is given by an age-dependent multi-type branching process. From this representation, we obtain equations satisfied by the average size of the global survival population as well as the one of subpopulations associated with 10 mutation phenotypes. This work was presented via a poster communication in a international congress [40].

7.1.3. Modeling of response to chemotherapy in gliomas

Participant: S. Wantz-Mézières
External collaborators: M. Ben Abdallah, Yann Gaudeau, J.-M. Moureaux (CRAN, Université de Lorraine) and M. Blonski, L. Taillandier (CHU Nancy)

In the framework of a collaboration with neurologists (Luc Taillandier, Marie Blonski, CHU Nancy) and automaticians (Jean-Marie Moureaux, Yann Gaudeau, CRAN), around the thesis supervision of M. Ben Abdallah, our aim is to work out personalized therapeutic strategy in the monitoring of diffuse low-grade glioma patients. Regular monitoring with MRI are used to estimate the tumour volume ; we proposed a method by manual segmentation and statistically assessed its reproducibility by a subjective test. In order to design a decision-aid tool for the response to chemotherapy, our approach is phenomelogical and we used simple regression tools to model and predict the cinetics of the tumour growth. We identified two different models. These results open up many perspectives, the main one being the modeling by multi-factor models, including biological and anatomopathological factors. This work is currently in progress.

7.1.4. Photodynamic therapy

Participant: C. Lacaux
External collaborators: T. Obara and M. Thomassin (CRAN, Université de Lorraine), L. Vinckenboch (Fribourg)

Our project focuses on an innovative application: the interstitial PDT for the treatment of high-grade brain tumors. This strategy requires the installation of optical fibers to deliver the light directly into the tumor tissue to be treated, while nanoparticles are used to carry the photosensitizer into the cancer cells. In order to optimize the intra-cerebral position of our optical fiber, two fundamental questions have to be answered: (1) What is the optimal shape and position of the light source in order to optimize the damage on malignant cells? (2) Is there a way to identify the physical parameters of the tissue which drive the light propagation?

Notice that we are obviously not the first ones to address these issues, and there is nowadays a consensus in favor of the algorithm proposed by L. Wang and S. L. Jacques for the simulation of light transport in biological tissues. However, our starting point is the observation that the usual methods slightly lack of formalism and miss formal representations that answer the questions of identifiability. In [16], in the framework of homogeneous biological tissues, we propose an alternative MC method to Wang's algorithm. Then we also propose a variance reduction method. Interestingly enough, our formulation also allows us to design quite easily a Markov chain Monte Carlo (MCMC) method based on Metropolis-Hastings algorithm and to handle the inverse problem (of crucial importance for practitioners), consisting in estimating the optical coefficients of the tissue according to a series of measurements. We have compared the proposed MC and MCMC method and Wang's algorithm: we see that our MC method is much more consistent. However, MCMC methods induce quick mutations, which paves the way to very promising algorithms in the inhomogenous case. To handle the inverse problem, we derive a probabilistic representation of the variation of the fluence with respect to the absorption and scattering coefficients. This leads us to the implementation of a Levenberg-Marquardt type algorithm that gives an approximate solution to the inverse problem. Our results open the way for new improvements of Monte-Carlo methods in the context of light propagation. They should rather be seen as a starting point for new methods, including in inhomogeneous tissue. This work has been presented in several french seminars (Lille, Avignon, Paris Descartes, Orléans).

7.1.5. Time-changed extremal process as a random sup measure

Participant: C. Lacaux
External collaborator: G. Samorodnitsky (Cornell, USA)

In extreme value theory, one of the major topics is the study of the limiting behavior of the partial maxima of a stationary sequence. When this sequence is i.i.d., the unique limiting process is well-known and called the extremal process. Considering a long memory stable sequence, the limiting process is obtained as a simple power time change extremal process. Céline Lacaux and Gennady Samorodnistky have proved in [38] that this limiting process can also be interpreted as a restriction of a self-affine random sup measure. In addition, they have established that this random measure arises as a limit of the partial maxima of the same long memory stable sequence, but in a different space. Their results open the way to propose new self-similar processes with stationary max-increments. Céline Lacaux has presented this work in an invited session of the international conference *Extreme Value Analysis* at Ann Arbor (June 2015).

7.1.6. Modulus of continuity of some conditionally sub-Gaussian fields, application to stable random fields

Participant: C. Lacaux
External collaborator: H. Biermé (Poitiers)

Hermine Biermé and Céline Lacaux maintain their collaboration on the study of anisotropic random fields. They have extended their previous work in the framework of conditionally sub-Gaussian random series. For such anisotropic fields, they have obtained a modulus of continuity and a rate of uniform convergence. Their framework enables the study of study e.g., Gaussian fields, stable random fields and multi-stable random fields. As invited speaker, Céline Lacaux has presented this work in the international conference *Adventure in Self-similarity* at Cornell University (June 2015) [17]. Another of their works in progress deals with the simulation of anisotropic Gaussian random fields and the estimation of their parameters using quadratic variations.

7.1.7. DNA sequences analysis

Participants: P. Vallois
External collaborators: A. Lagnoux and S. Mercier (Toulouse)

Here we want to determine the sequences that are biologically interesting and compare the results using the single local score Hn and using the pair (Hn;Ln) where Ln is the length of the segment that realizes the best score. In that view, we work on the p-values associated to the observed samples.

7.1.8. Multicriteria Agregation for Health Economic Assessment

Participants: T. Bastogne, Y. Petot, P. Vallois

The framework of this work is the PhD thesis of Yann Petit. The first chapter of the thesis is a state of the art identifying the current challenges in medico-economic analyses. A review article should be submitted in spring 2016. We are currently working on the aggregation operators, based on fuzzy measures and the Choquet integral. Theoretical results have been obtained and a publication is planned to be submitted in the second half of 2016. Work continues by introducing probabilities. The next step will be to apply our theoretical results to real clinical cases.

7.1.9. Spatial and spatio-temporal modeling

Participant : A. Gégout-Petit
External collaborators: S. Li, L. Guerin-Dubrana (Inra Bordeaux)

In the framework of a collaboration with INRA Bordeaux about the esca-illness of vines, Anne Gégout-Petit with Shuxian Li developed different spatial models and spatio-temporal models for different purposes: (1) study the distribution and the dynamics of esca vines in order to tackle the aggregation and the potential spread of the illness (2) propose a spatio-temporal model in order to capture the dynamics of cases and measure the effects of environmental covariates. For this, we propose different hierarchic models with latent process associated with a bayesian inference. A part of the research has been submitted in a journal of biology [39]. Shuxian Li defended his PhD on December the 15th.

7.1.10. Stochastic modeling of fatigue crack propagation

Participants: R. Azaïs, A. Gégout-Petit
External collaborators: A.B. Abdessalem, M. Puiggali, M. Touzet (Bordeaux)

Fatigue crack propagation is a stochastic phenomenon due to the inherent uncertainties originating from material properties and environmental conditions. In a recent preprint [35], we propose to model and to predict the fatigue crack growth by a piecewise-deterministic Markov process associated with deterministic crack laws of the literature, namely the Paris-Erdogan equation defined by $\mathrm{d}a/\mathrm{d}N = C(\Delta K)^m$ and the Forman equation given by $\mathrm{d}a/\mathrm{d}N = C(\Delta K)^m/(K_c(1-R)-\Delta K)$, where a is the crack length, N denotes the number of cyclic mechanical loads, ΔK is the range of the stress intensity factor and C, m, K_c and R are different parameters. We introduce a regime-switching model to express the transition between Paris' regime and rapid propagation which occurs just before failure. We also investigate the prediction of the fatigue crack path and its variability based on measurements taken at the beginning of the propagation. This work has also been presented in an international conference [25].

7.2. Estimation and control for stochastic processes

7.2.1. Inference for dynamical systems driven by Gaussian noises

Participant: S. Tindel
External collaborators: K. Chouk, A. Deya, Y. Hu, L. Khoa, D. Nualart, E. Nualart, F. Xu. (US)

The problem of estimating the coefficients of a general differential equation driven by a Gaussian process is still largely unsolved. To be more specific, the most general ($\mathbb{R}$-valued) equation handled up to now as far as parameter estimation is concerned is of the form:

$$X_t^\theta = a + \theta \int_0^t b(X_u)\, du + B_t,$$

where θ is the unknown parameter, b is a smooth enough coefficient and B is a one-dimensional fractional Brownian motion. In contrast with this simple situation, our applications of interest (motivated by some anomalous diffusion phenomenon in proteins fluctuations) require the analysis of the following $\mathbb{R}^n$-valued equation:

$$X_t^\theta = a + \int_0^t b(\theta; X_u)\, du + \int_0^t \sigma(\theta; X_u)\, dB_t, \tag{2}$$

where θ enters non linearly in the coefficient, where σ is a non-trivial diffusion term and B is a d-dimensional fractional Brownian motion. We have thus decided to tackle this important scientific challenge first.

To this aim, here are the steps we have focused on in 2015:

- Some limit theorems for general functionals of Gaussian sequences [6], or for functionals of a Brownian motion [3], which give some insight on the asymptotic behavior of systems like (2).
- Extension of pathwise stochastic integration to processes indexed by the plane in [1], which helps to the definition of noisy systems such as partial differential equations.
- Definition of new systems driven by a (spatial) fractional Brownian motion, such as the stochastic PDE considered in [37].
- The local asymptotic normality obtained for the system (2), which implies a lower bound on general estimators of the coefficient θ. This is the contents of the preprint [41].

7.2.2. Optimal estimation of the jump rate of a piecewise-deterministic Markov process

Participants: R. Azaïs, A. Muller-Gueudin

A piecewise-deterministic Markov process is a stochastic process whose behavior is governed by an ordinary differential equation punctuated by random jumps occurring at random times. In a recent preprint [33], we focus on the nonparametric estimation problem of the jump rate for such a stochastic model observed within a long time interval under an ergodicity condition. More precisely, we introduce an uncountable class (indexed by the deterministic flow) of recursive kernel estimates of the jump rate and we establish their strong pointwise consistency as well as their asymptotic normality. In addition, we propose to choose among this class the estimator with the minimal variance, which is unfortunately unknown and thus remains to be estimated. We also discuss the choice of the bandwidth parameters by cross-validation methods. This paper has also been presented in two national workshops.

7.2.3. Estimation and optimal control for the TCP process

Participant : R. Azaïs
External collaborators: N. Krell (Rennes), B. de Saporta (Montpellier)

In [33], we assume that the transition kernel is continuous with respect to the Lebesgue measure. This condition may be not satisfied in some applications, as for instance for the well-known TCP process that appears in the modeling of the famous Transmission Control Protocol used for data transmission over the Internet. As a consequence, we propose to investigate estimation followed by optimal control for this ergodic process. The particular framework defined by this process allows us to define an optimal policy for the estimation of its jump rate. We obtain at present an efficient method for estimating the moments of the conditional distribution of the inter-congestion times in an optimal way. This work is currently in progress.

7.2.4. Estimation of integrals from a Markov design

Participant : R. Azaïs
External participants: B. Delyon, F. Portier

Monte-Carlo methods for estimating an integral assume that the distribution of the random design is known. Unfortunately, some applications generate a design whose density function f is unknown. In this case, a solution is to perform the classical Monte-Carlo estimate of the integral by replacing f by a leave-one-out kernel estimator, and one may expect the convergence

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\varphi(X_i)}{\widehat{f}^{(-i)}(X_i)} \to \int \varphi \mathrm{d}\lambda,$$

when the number n of independent data X_i goes to infinity. This difficult question has been investigated by François Portier and Bernard Delyon in a recent paper. We propose to extend this work to the more general case of a Markov design. This new model includes a large variety of applications, in particular in biology and climatology. Indeed, the data $(X_i, \varphi(X_i))$ are often obtained from a measuring instrument that is launched in its environment and thus follows a random walk in it. A paper on this work will be submitted soon.

7.2.5. Method of control for radiotherapy treatment using Decision Markov Processes

Participants : R. Azaïs, B. Scherrer, S. Tindel, S. Wantz-Mézières

In recent years, Bastogne, Keinj and Vallois designed a Markov model of the evolution of cells under a radiotherapy treatment. We are currently investigating the problem of optimizing the radiotherapy intensity sequence in order to kill as many cancerous cells as possible while preserving as many healthy cells, a problem that fits into the stochastic optimal control problem. Our preliminary efforts suggest that, since we are dealing with large populations of cells, the problem can be well approximated by a limit deterministic optimal control problem. We can solve this problem numerically with a Pontryagine approach, and symbolically (in the simplest cases) by identifying the critical points of some multivariate polynomials. The latter approach allows us to validate the fact that the former actually finds globally optimal solutions. This is a work in progress.

7.2.6. Numerical approximate schemes for large optimal control problems and zero-sum two player games

Participant: B. Scherrer
External collaborators: V. Gabillon, M. Ghavamzadeh, M. Geist, B. Lesner, J. Perolat, O. Pietquin, M. Tagorti

We have provided in [23] (ICML 2015) the first finite-sample analysis of the LSTD(λ) algorithm aimed at approximating the value of some fixed policy in a large MDP, through the approximation of the projected fixed point of the linear Bellman equation from samples. This analysis highlights the influence of the main parameter λ of the algorithm.

The long version of our previous work on the analysis of an approximate modified policy iteration for optimal control and its application to the Tetris domain is now published in JMLR [13]. The extension of this algorithm family for computing approximately-optimal non-stationary policies allows to improve the dependency with respect to the discount factor: we provide such improved bounds in [19], as well as examples that show that our analysis is tight (and cannot be further improved).

An original analysis of the variation of the approximate modified policy iteration for computing approximate Nash equilibria in the more general setting of two-player zero-sum games was published in ICML 2015 [22].

7.3. Algorithms and estimation for graph data

7.3.1. Modelling of networks of multiagent systems

Participant: A. Muller-Gueudin
External collaborators: A. Girard, S. Martin, I.C. Morarescu (CRAN, Nancy)

We relate here a starting of collaboration with researchers in Automatics in Nancy. We consider here networks, modeled as a graph with nodes and edges representing the agents and their interconnections, respectively. The objective is to study the evolution of the opinion of all the agents. The connectivity of the network, persistence of links and interactions reciprocity influence the convergence speed towards a consensus. The problem of consensus or synchronization is motivated by different applications as communication networks, power and transport grids, decentralized computing networks, and social or biological networks. We then consider networks of interconnected dynamical systems, called agents, that are partitioned into several clusters. Most of the agents can only update their state in a continuous way using only inner-cluster agent states. On top of this, few agents also have the peculiarity to rarely update their states in a discrete way by reseting it using states from agents outside their clusters. In social networks, the opinion of each individual evolves by taking into account the opinions of the members belonging to its community. Nevertheless, one or several individuals can change their opinions by interacting with individuals outside its community. These inter-cluster interactions can be seen as resets of the opinions. This leads us to a network dynamics that is expressed in term of reset systems. We suppose that the reset instants arrive stochastically following a Poisson renewal process. We have an accepted paper in the journal IEEE Transactions on Automatic Control [10].

7.3.2. Microbial interaction inference by network analysis

Participants: A. Gégout-Petit, A. Muller-Gueudin
External collaborators: A. Deveau (INRA Nancy), C. Raïssy (Inria Orpailleur)

The objective is to characterize microbial interactions in a particular environment: the truffles.

The truffle provides a habitat for complex bacterial communities. The role for bacteria in the development of truffles has been suggested but very little is known regarding the structure and the functional potential of the truffle's bacterial communities along truffle maturation. In a mathematical point of view, two micro-organisms are connected if they are not independent, conditionally to the other micro-organisms. Several models fit into this setting, especially the gaussian graphical models, the bayesians networks, and the graphical log-linear models. But the data, which can be zeros inflated, need developments and we have to proposed new models. Moreover, we are confronted to the problem that $n \ll p$, that is the sample size is much smaller that the number of variables ($n = 30, p = 200$). Last year, thanks to a financially supported project (PEPS), we have began a collaboration between statisticians and data-miners. The first approches have been notified in a report [31]. The statistical methodologies developed for this project could also be applied to human health (for instance identification of network between bacteria inside colon).

7.3.3. Lossy compression of unordered trees

Participant: R. Azaïs
External collaborators: J-B. Durand, C. Godin

A classical compression method for trees is to represent them by directed acyclic graphs. This approach exploits subtree repeats in the structure and is efficient only for trees with a high level of redundancy. The class of self-nested trees presents remarkable compression properties by this method because of the systematic repetition of subtrees. In particular, the compressed version of a self-nested tree T is a linear directed acyclic graph with only $1 + \text{height}(T)$ nodes. Unfortunately, it should be noted that trees without a high level of redundancy are often insufficiently compressed by this procedure. In a paper recently submitted for publication in an international conference [32], we introduce a lossy compression method that consists in computing in polynomial time for trees with bounded outdegree the reduction of a self-nested structure that closely approximates the initial data. We prove on a simulated dataset that the error rate of this lossy compression method is always better than the loss involved in a previous algorithm of the literature, while the compression rates are equivalent.

7.3.4. Inference for critical Galton-Watson trees from their Harris process

Participant: R. Azaïs
External collaborator: A. Genadot (Inria CQFD Bordeaux)

Galton-Watson trees are an elementary model for the genealogy of a branching population and thus play a central role in biology. Critical Galton-Watson trees are generated from a sibling distribution μ whose theoretical expectation $\sum k\mu(k)$ is equal to 1. Under this assumption, the well-known Harris process of a tree conditioned on having n nodes converges to a Brownian excursion characterized by the variance $\sigma^2 = \sum (k-1)^2 \mu(k)$ of μ. We propose to exploit this asymptotic approximation to define a new estimate of the unknown parameter of interest σ^2 based on a least-square method. In particular, this new technique allows us to take into account the behavior of the Harris path with respect to its asymptotic theoretical expectation. In certain cases, we obtain a better confidence interval than the classical approach. A paper on this work is in preparation.

7.4. Regression and machine learning

7.4.1. Uniform asymptotic certainty bands for the conditional cumulative distribution function

Participants: S. Ferrigno, A. Muller-Gueudin
External collaborator: M. Maumy-Bertrand (IRMA, Strasbourg)

In this work with Myriam Maumy-Bertrand (IRMA, Strasbourg), we study the conditional cumulative distribution function and a nonparametric estimator associated to this function. The conditional cumulative distribution function has the advantages of completely characterizing the law of the random considered variable, allowing to obtain the regression function, the density function, the moments and the conditional quantile function. As a nonparametric estimator of this function, we focus on local polynomial techniques described in Fan and Gijbels [64]. In particular, we use the local linear estimation of the conditional cumulative distribution function.

The objective of this work is to establish uniform asymptotic certainty bands for the conditional cumulative distribution function. To this aim, we give exact rate of strong uniform consistency for the local linear estimator of this function. We show that limit laws of the logarithm are useful in the construction of uniform asymptotic certainty bands for the conditional distribution function. In particular, we use a single bootstrap to construct sharp uniform asymptotic bands of this estimator.

We illustrate our results with simulations and a study of fetal growth which is based on 694 fetuses (carefully selected by exclusion of multiple pregnancies, malformed, macerated or serious ill fetuses, or those with chromosomal abnormalities) autopsied in fetopathologic units of the "Service de foetopathologie et de placentologie" of the Maternité Régionale Universitaire (CHU Nancy, France) between 1996 and 2013.

We have presented our results in two international conferences with proceedings in Lille in June 2015 ("47èmes Journées de Statistique de la SFdS") [21] and London in December 2015 ("CM Statistics") [36].

7.4.2. Omnibus tests for regression models.

Participants: R. Azaïs, S. Ferrigno
External collaborator: M-J. Martinez Marcoux (LJK, Grenoble)

The aim of this collaboration with Marie-José Martinez Marcoux (LJK, Grenoble) is to compare, through simulations, several methods to test the validity of a regression model. These tests can be "directional" in that they are designed to detect departures from mainly one given assumption of the model (for example the regression function, the variance or the error) or global (for example the conditional distribution function). The establishment of such statistical tests require the use of nonparametric estimators various functions (regression, variance, cumulative distribution function). The idea would then be able to build a tool (package R) that allows a user to test the validity of the model it uses through different methods and varying parameters associated with modeling. This work is currently in progress.

7.4.3. Data analysis techniques: a tool for cumulative exposure assessment

Participant: J-M. Monnez
External collaborators : W. Kihal, B. Lalloué, C. Padilla, D, S. Zmirou-Navier

Everyone is subject to environmental exposures from various sources, with negative health impacts (air, water and soil contamination, noise, etc.) or with positive effects (e.g. green space). Studies considering such complex environmental settings in a global manner are rare. We propose to use statistical factor and cluster analyses to create a composite exposure index with a data-driven approach, in view to assess the environmental burden experienced by populations. We illustrate this approach in a large French metropolitan area. The study was carried out in the Great Lyon area (France, 1.2 M inhabitants) at the census Block Group (BG) scale. We used as environmental indicators ambient air NO2 annual concentrations, noise levels and proximity to green spaces, to industrial plants, to polluted sites and to road traffic. They were synthesized using Multiple Factor Analysis (MFA), a data-driven technique without a priori modeling, followed by a Hierarchical Clustering to create BG classes. The first components of the MFA explained, respectively, 30, 14, 11 and 9% of the total variance. Clustering in five classes group: (1) a particular type of large BGs without population; (2) BGs of green residential areas, with less negative exposures than average; (3) BGs of residential areas near midtown; (4) BGs close to industries; and (5) midtown urban BGs, with higher negative exposures than average and less green spaces. Other numbers of classes were tested in order to assess a variety of clustering. We present an approach using statistical factor and cluster analyses techniques, which seem overlooked to assess cumulative

exposure in complex environmental settings. Although it cannot be applied directly for risk or health effect assessment, the resulting index can help to identify hot spots of cumulative exposure, to prioritize urban policies or to compare the environmental burden across study areas in an epidemiological framework [9].

7.4.4. Online Partial Principal Component Analysis of a Data Stream

Participant: J-M. Monnez
External collaborator: R. Bar (EDF, R & D)

Consider a data stream and suppose that each data vector is a realization of a random vector whose expectation varies with time, the law of the centered data vector being stationary. Consider the principal component analysis (PCA) of this centered vector called partial PCA. In this study are defined online estimators of direction vectors of the first principal axes by stochastic approximation processes using a data batch at each step or all the data until the current step. This extends a former result obtained by the second author by using one data vector at each step. This is applied to partial generalized canonical correlation analysis by defining a stochastic approximation process of the metric involved in this case using all the data until the current step. If the expectation of the data vector varies according to a linear model, a stochastic approximation process of the model parameters is used. All these processes can be performed in parallel.

Moreover, several incremental procedures of linear and logistic regression of a data stream were defined and tested and compared on existing batch data files and on simulated data streams.

7.4.5. Prognostic Value of Estimated Plasma Volume in Heart Failure.

Participant: J-M. Monnez
External collaborators: E. Albuisson, B. Pitt, P. Rossignol, F. Zannad (CHU, Nancy)

The purpose of this study was to assess the prognostic value of the estimation of plasma volume or of its variation beyond clinical examination in a post-hoc analysis of EPHESUS (Eplerenone Post-Acute Myocardial Infarction Heart Failure Efficacy and Survival Study).

Assessing congestion after discharge is challenging but of paramount importance to optimize patient management and to prevent hospital readmissions.

The present analysis was performed in a subset of 4,957 patients with available data (within a full dataset of 6,632 patients). The study endpoint was cardiovascular death or hospitalization for heart failure (HF) between months 1 and 3 after post-acute myocardial infarction HF. Estimated plasma volume variation (ΔePVS) between baseline and month 1 was estimated by the Strauss formula, which includes hemoglobin and hematocrit ratios. Other potential predictors, including congestion surrogates, hemodynamic and renal variables, and medical history variables, were tested.

An instantaneous estimation of plasma volume at month 1 was defined and also tested.

Multivariate analysis was performed with stepwise logistic regression. ΔePVS was selected in the model. The corresponding prognostic gain measured by integrated discrimination improvement was significant. Nevertheless, instantaneous estimation of plasma volume at month 1 was found to be a better predictor than ΔePVS. LDA with mixed variables was also performed and confirmed these results.

In HF complicating myocardial infarction, congestion as assessed by the Strauss formula and an instantaneous derived measurement of plasma volume provided a predictive value of early cardiovascular events beyond routine clinical assessment. Prospective trials to assess congestion management guided by this simple tool to monitor plasma volume are warranted [4].

7.4.6. Death or hospitalization scoring for heart failure patients

Participant: J-M. Monnez
External collaborator: E. Albuisson (CHU Nancy)

The purpose of this study was to define an event - death or hospitalization - score for heart failure patients based on the observation of biological, clinical and medical historical variables. Some of them were transformed or winsorized. Two methods of statistical learning were performed, logistic regression and linear discriminant analysis, with a stepwise selection of variables. Aggregation of classifiers by bagging was used. Finally a score taking values between 0 and 100 was established.

7.4.7. A simultaneous stepwise covariate selection and clustering algorithm to discriminate a response variable with numerous values

Participant: J-M. Monnez
External collaborator: O. Collignon (LIH, Luxembourg)

In supervised learning the number of values of a response variable to predict can be high. Also clustering them in a few clusters can be useful to perform relevant supervised classification analyses. On the other hand selecting relevant covariates is a crucial step to build robust and efficient prediction models, especially when too many covariates are available in regard to the overall sample size. As a first attempt to solve these problems, we had already devised in a previous study an algorithm that simultaneously clusters the levels of a categorical response variable in a limited number of clusters and selects forward the best covariates by alternate minimization of Wilks' Lambda. In this paper we first extend the former version of the algorithm to a more general framework where Wilks's Lambda can be replaced by any model selection criterion. We also turned forward selection into stepwise selection in order to remove covariates while the procedure processes if necessary. Finally an application of our algorithm to real datasets from peanut allergy studies allowed confirming previously published results and suggesting new discoveries.

7.4.8. Statistical Analyses of Cell Impedance Signals in High-Throughput Cell Analysis

Participant: T. Bastogne, L. Batista
External Collaborator: El-Hadi Djermoune (Université de Lorraine, CRAN)

With the advent of high-throughput technologies, life scientists are starting to grapple with massive data sets, encountering challenges with handling, processing and moving information that were once the domain of astronomers and high-energy physicists [91]. We particularly focus the statistical analysis of large batch of time series with applications in the preclinical research in Cancerology. Our original contribution consists in developing new dynamical system identification methods suited to the processing of those type of data. System identification is a data-driven modeling approach more and more used in biology and biomedicine. In this application context, each assay is always repeated to estimate the response variability. The inference of the modeling conclusions to the whole population requires to account for the inter-individual variability within the modeling procedure. One solution consists in using mixed effects models but up to now no similar approach exists in the field of dynamical system identification. Therefore, our objective is to develop a new identification method integrating mixed effects within an ARX (Auto Regressive model with eXternal inputs) model structure. The parameter estimation step relies on the EM (Expectation-Maximisation) algorithm. First simulation results show the relevance of this solution compared with a classical procedure of system identification repeated for each subject. This work and derived was accepted in conference papers [34] [24] [18].

CAMUS Team

7. New Results

7.1. Formal Proofs for an Ordering Relation in Explicitly Parallel Programs

Participants: Alain Ketterlin, Éric Violard.

This project is a collaborative work with the COMPSYS Inria Team, in Lyon. Participants are: Paul Feautrier, Tomofumi Yuki.

The growing need to make use of available parallelism has led to new explicitly parallel language constructs. These constructs are usually grouped under the term *Task Parallelism*, because they aim to go beyond "simple" *Data Parallelism* (i.e., loop and array-based parallelism). Prominent examples of languages integrating task parallelism are X10 (http://x10-lang.org) and variants, Cilk (http://supertech.csail.mit.edu/cilk/), and recent versions of OpenMP (http://www.openmp.org). Most of the work on such languages has focused on efficient run-time support for *tasks*, in contrast with *threads*, i.e., for programs generating potentially large numbers of distinct tasks with explicit (but arbitrary) ordering between the tasks. However, little attention has been given to the static analysis and optimization of explicitly parallel programs, probably because their properties are much harder to formalize, compared to their sequential counterpart. Starting with the work of our colleagues Paul Feautrier and Tomofumi Yuki, from the Compsys team in Lyon, we have advanced the formalization and formally proved several properties of some fundamental building blocks for the analysis of certain classes of explicitly parallel programs.

Task parallelism is usually based on a few syntactic constructs to represent tasks and their synchronization. We use X10's terminology (and syntax, with simplifications), but the corresponding constructs of other languages is usually obvious. Across all languages one finds a construct to start (or *spawn*) an asynchronous task, named `async` in X10, and a "container" construct, named `finish` in X10, whose role is to wait for the completion of all task spawned during the execution of its body. Given that these constructs allow the parallel execution of pieces of the program, a first question arises: is there a static (i.e., compile-time) way to decide whether two given statements are ordered, i.e., that the first necessarily executes before the other. Feautrier and Yuki (with colleagues) have defined such a criterion for programs made of `async` and `finish` [33], along with arbitrary statements and for-loops, defining the so-called *polyhedral fragment* of X10. The resulting (partial) relation, called *happens-before*, opens the door to various static analyses, like data-dependence analysis, which are at the heart of a range of optimization techniques. Here is a quick example:

```
finish
for i in ...
    async
    for j in ...
        S(i,j)
```

$\texttt{S}(i,j)$ happens before $\texttt{S}(i',j')$ iff $i = i' \wedge j < j'$

The resulting condition, $i = i' \wedge j < j'$, defines exactly the situation in which two statement executions are ordered, and can be seen as an appropriate extension of the lexicographic order to explicitly parallel programs.

Our work on this basis has been to take the formal definition of happens-before (HB), and implement it in Coq (https://coq.inria.fr). The goal was first to prove various properties of the relation, like transitivity, and second to provide a formal proof of both correctness and completeness of HB itself. The first part has been fairly immediate, due to the high representative power of Coq. The second part took more time, and involved several new contributions. The major part of the work went into defining a formal semantics for the fragment of X10 needed by the definition of HB. Given the semantics, it was possible to obtain the relation between a program and its trace(s), and then to prove that HB is correct (i.e., if HB states that one statement executes before another, then these statements appear in order in all possible traces of the program), and that HB is complete (i.e., that statements that are always ordered in traces are actually recognized as such by HB). The complete proof scripts are available on the Inria forge (gforge.inria.fr), under the `x10-coq` project.

Further work has also started on extending *happens-before* to X10 programs using synchronization primitives called *clocks*, which are basically *barriers*, where distinct tasks can wait for each other. Since an unrestricted use of synchronization barriers can lead to deadlocks, X10 introduces "implicit clocks", which are introduced (and scoped) by a `finish` construct, on which a task can "register", and whose scoping rules ensure that any program point can only use the single "nearest" clock. These restrictions offer termination guarantees, which in turn enables a sound *happens-before* relation between statement instances. The "clock-less" HB relation can then be modified to take into account the additional ordering imposed by clocks. We have started work to update the semantics to the case of implicit clocks, and to formalize this extension in Coq.

7.2. Validity Conditions for Transformations of Non-Affine Programs

Participants: Alain Ketterlin, Philippe Clauss.

This project is a collaborative work with the CORSE Inria Team, in Grenoble. Participant is: Fabrice Rastello.

Representing loop nests with the help of the polyhedral model has been a powerful and fruitful strategy to enable automatic optimization and parallelization. However, this model places strong requirements on the input program, and in many cases these requirements are hard to meet. Because they are based on linear programming, polyhedral techniques require every constraint to be affine in loop counters and parameters. While this is easily verified for loop bounds in a large majority of programs, the same constraint imposed to memory access functions is often too strong. There are several reasons for this. First, programmers often linearize multi-dimensional arrays, turning straightforward accesses like `t[i][j]` into `t1[i*n+j]`, with the unfortunate effect of placing their program outside the scope of the polyhedral model. Second, optimization often happens late in the compilation process (or even during just-in-time compilation at run-time), where multi-dimensional array accesses have been transformed by the compiler itself, for the needs of its earlier passes. Third, complex data storage strategies for certain classes of arrays, e.g., band or triangular matrices, may introduce non-linear access functions, and this non-linearity must be taken into account, e.g., for locality optimization. And fourth, some access functions are almost completely unspecified, like in the case of indirect accesses (`t[s[i]]`) or abstract mappings (`t[f(i)]`).

Our goal is to extend polyhedral analysis techniques to cover at least some of these cases, and see how far we can push the limits of the fundamental algorithms beyond pure linearity. We have started by considering the case of multi-dimensional array linearization, where the code doesn't provide access functions for all (original) dimensions, but rather a single access function, which is linear in loop counters but contains parametric coefficients. Here is an example illustrating our initial target, which is taken from the `gemver` program in the `polybench` suite:

```
  for (i = 0; i < n; i++)
    for (j = 0; j < n; j++)
      // Was: A[i][j] = A[i][j] + u1[i] * v1[j] + ...;
S1:   *(n*i+A+j) = *(n*i+A+j) + *(u1+i) * *(v1+j) + ...;
  for (i = 0; i < n; i++)
    for (j = 0; j < n; j++)
      // Was: x[i] = x[i] + beta * A[j][i] * y[j];
```

```
S2:   *(x+i) = *(x+i) + beta * *(n*j+A+i) * *(y+j);
  // ...
```

The original form of the statements appear in comments, but what finally reaches the compiler is much more convoluted: basically, every array access appears as a pointer access whose effective address is a polynomial function mixing counters (`i`, `j`), array base addresses (`A`, `u1`, `v1`, `x`, `y`), and size parameters (`n`). In some other cases, the arrays have been "locally" linearized, i.e., the code still displays different arrays, but their inner dimensions have been linearized. In our example, statement `S1` would appear as:

```
      // Was: A[i][j] = A[i][j] + u1[i] * v1[j] + ...;
S1:   A[n*i+j] = A[n*i+j] + u1[i] + v1[j] + ...;
```

This is an important special case in practice, and its particular structure helps a lot, for example, when data dependence analysis is needed.

Extending current polyhedral techniques to deal with non-affine accesses is a formidable endeavor, requiring the adaptation of the many algorithms developed over decades for analysis, scheduling, and code generation. Rather, we have started by studying a specific task, with immediate practical impact: given a non-affine loop nest *and* a specific desired transformation, what are the conditions under which this condition is valid? It is not unreasonable to expect the transformation to be provided by other means than pure analysis, for instance to be suggested by profiling data. In this case, the problem we are left with is the one of testing whether the given transformation is valid. This in turn requires testing the emptyness of a "problematic system". For any given loop nest, this can be written as:

$$\begin{array}{lll} \bigvee\limits_{(A,A')} & \exists (v, v') \text{s.t.} & \\ & v \in \mathcal{D}_A \wedge v' \in \mathcal{D}_{A'} & \text{(domain)} \\ \wedge & v \prec_{lex} v' & \text{(originalschedule)} \\ \wedge & A(v) = A'(v) & \text{(sameaccesslocation)} \\ \wedge & T_A(v) \neg \prec_{lex} T_{A'}(v') & \text{(transformedschedule)} \end{array}$$

where A and A' range over pairs of potentially conflicting accesses, v and v' are iteration vectors, $\mathcal{D}_A$ and $\mathcal{D}_{A'}$ are iteration domains, $A(v)$ and $A'(v')$ are access functions, and T_A and $T'_{A'}$ are schedules. The condition under which the transformation is valid is the projection of this set on parameter dimensions, i.e., the elimination of all variables representing counters. The difficulty of this comes from the non-affine condition expressing the equality of access functions.

Building on previous work, we have devised a projection procedure that eliminates all counters and leaves a (usually complex) condition on parameters. We have also developed several simplification strategies, applied during elimination and also on the final result, that overall produces a test deciding whether the targeted transformation can be applied. For instance, on the fully linearized version of the previous examples, when deciding whether the following transformation is legal:

$$T_{\mathtt{S1}}(0, i, j) = (0, i, j) \qquad T_{\mathtt{S2}}(1, i, j) = (0, j, i)$$

i.e., interchanging the second loop (around `S2`) and then applying fusion on both depth-2 loops, our elimination and simplification procedure produces the following run-time test:

```
if ( ((y+n >= x+2) && (x+n >= y+2))
|| ((n >= 2) && (n*n+A >= x+1) && (x >= A+1))
|| ((n >= 2) && (u1+n >= x+1) && (x+n >= u1+2))
|| ((n >= 2) && (n+v1 >= x+1) && (x+n >= v1+1))
|| ((n*n+A >= y+1) && (y >= A+1) && (n >= 2)) || ...)
  // Transformation invalid: run the original version...
else
  // Transformation valid: run the transformed version...
```

The reader may want to verify that this test actually corresponds to verifying that “arrays” do not overlap, but only as far as the given transformation requires it.

A systematic evaluation of our procedure on a benchmark suite has shown that the resulting tests are both accurate and incur very little run-time overhead. The overall mechanism compares favorably with alternative techniques aiming at dealing with non-affine access functions, which consist in statically reconstructing array dimensions [30]. This part of our work is ready for publication. However, to be completely competitive with alternative approaches, we need to find ways to complete the polyhedral compilation chain, with a prior effective scheduling algorithm and a posterior code generation algorithm.

7.3. Automatic Parallelization of Nonlinear Loops

Participants: Aravind Sukumaran-Rajam, Philippe Clauss.

Runtime code optimization and speculative execution are becoming increasingly prominent to leverage performance in the current multi- and many-core era. However, a wider and more efficient use of such techniques is mainly hampered by the prohibitive time overhead induced by centralized data race detection, dynamic code behavior modeling, and code generation. Most of the existing Thread Level Speculation (TLS) systems rely on naively slicing the target loops into chunks and trying to execute the chunks in parallel with the help of a centralized performance-penalizing verification module that takes care of data races. Due to the lack of a data dependence model, these speculative systems are not capable of doing advanced transformations, and, more importantly, the chances of rollback are high. The polyhedral model is a well- known mathematical model to analyze and optimize loop nests. The current state-of-art tools limit the application of the polyhedral model to static control codes. Thus, none of these tools can generally handle codes with while loops, indirect memory accesses, or pointers. Apollo (Automatic POLyhedral Loop Optimizer) is a framework that goes one step beyond and applies the polyhedral model dynamically by using TLS. Apollo can predict, at runtime, whether the codes are behaving linearly or not, and it applies polyhedral transformations on-the-fly.

Apollo has been extended to handle codes whose memory accesses and loop bounds are not necessarily linear [23], [14]. The proposed extension consists of modeling memory addresses that are accessed either as "tubes" obtained through linear regression, or as ranges. More generally, this approach expands the applicability of the polyhedral model at runtime to a wider class of codes. Plugging together both linear and nonlinear accesses to the dependence prediction model enables the application of polyhedral loop optimizing transformations even for nonlinear code kernels while also allowing a low-cost speculation verification.

This work takes part of Aravind Sukumaran-Rajam’s PhD thesis that has been defended November the 5th, 2015 [13].

7.4. Dynamic Code Generation for Speculative Polyhedral Optimization

Participants: Juan Manuel Martinez Caamano, Philippe Clauss.

We have developped a new runtime code generation technique for speculative loop optimization and parallelization, that allows to generate on-the-fly codes resulting from any polyhedral optimizing transformation of loop nests, such as tiling, skewing, loop fission, loop fusion or loop interchange, without introducing a penalizing time overhead. The proposed strategy is based on the generation of code bones at compile-time, which are parametrized code snippets either dedicated to speculation management or to computations of the original target program. These code bones are then instantiated and assembled at runtime to constitute the speculatively-optimized code, as soon as an optimizing polyhedral transformation has been determined. Their granularity threshold is sufficient to apply any polyhedral transformation, while still enabling fast runtime code generation. This strategy has been implemented in the speculative loop parallelizing framework Apollo.

7.5. The XFOR Programming Structure

Participants: Imen Fassi, Philippe Clauss, Cédric Bastoul.

We have proposed a new programming control structure called “xfor” or “multifor”, providing users a way to schedule explicitly the statements of a loop nest, and take advantage of optimization and parallelization opportunities that are not easily attainable using the standard programming structures, or using automatic optimizing compilers [19]. This is the PhD work of Imen Fassi, who started her work in 2013 and who defended her thesis November the 27th, 2015 [12].

It has been shown that xfor programs often reach better performance than programs optimized by fully automatic polyhedral compilers like Pluto [29]. It has also been shown that different versions of codes may perform very differently, although their memory behaviors are very similar. By analyzing further the origins of such performance differences, we noticed five important gaps in the currently adopted and well-established code optimization strategies [18], [19]: insufficient data locality optimization, excess of conditional branches in the generated code, too verbose code with too many machine instructions, data locality optimization resulting in processor stalls, and finally missed vectorization opportunities.

To ease and extend the usage of the XFOR structure, we have developed:

- Xfor-Wizard, which is a programming environment for XFOR programs, assisting users in writing XFOR codes and applying optimizing transformations. Automatic dependence analysis and comparisons against a referential code (XFOR-loops or classic for-loops) are achieved to order to help the user in ensuring semantic correctness of the written code.
- XFORGEN, which is a tool to automatically generate an XFOR code that is equivalent to for-loops that have been automatically transformed using a static polyhedral compiler. The generated XFOR code exhibits the parameters of the transformations that have been applied and thus can be modified for further optimizations.

7.6. Dynamic Optimization of Binary Code

Participants: Philippe Clauss, Alain Ketterlin.

This project is a collaborative work with the ALF Inria Team, in Rennes. Participants are: Erven Rohou and Nabil Hallou.

Automatic code optimizations have traditionally focused on source-to-source transformation tools and compiler IR-level techniques. Sophisticated techniques have been developed for some classes of programs, and rapid progress is made in the field. However, there is a persistent hiatus between software vendors having to distribute generic programs, and end-users running them on a variety of hardware platforms, with varying levels of optimization opportunities. The next decade may well see an increasing variety of hardware, as it has already started to appear particularly in the embedded systems market. At the same time, one can expect more and more architecture-specific automatic optimization techniques.

Unfortunately, many “old” executables are still being used although they have been originally compiled for now outdated processor chips. Several reasons contribute to this situation:

- commercial software is typically sold without source code (hence no possibility to recompile) and targets slightly old hardware to guarantee a large base of compatible machines;
- though not commercial, the same applies to most Linux distributions [0] – for example Fedora 16 (released Nov 2011) is supported by Pentium III (May 1999) [0];
- with the widespread cloud computing and compute servers, users have no guarantee as to where their code runs, forcing them to target the oldest compatible hardware in the pool of available machines.

All this argues in favor of binary-to-binary optimizing transformations. Such transformations can be applied either statically, i.e., before executing the target code, or dynamically, i.e., while the target code is running.

[0] with the exception of Gentoo that recompiles every installed package

[0] http://docs.fedoraproject.org/en-US/Fedora/16/html/Release_Notes/sect-Release_Notes-Welcome_to_Fedora_16.html

Dynamic optimization is mostly addressing adaptability to various architectures and execution environments. If practical, dynamic optimization should be preferred because it eliminates several difficulties associated with static optimization. For instance, when deploying an application in the cloud, the executable file may be handled by various processor architectures providing varying levels of optimization opportunities. Providing numerous different adapted binary versions cannot be a general solution. Another point is related to interactions between applications running simultaneously on shared hardware, where adaptation may be required to adjust to the varying availability of the resources. Finally, most code optimizations have a basic cost that has to be recouped by the gain they provide. Depending on the input data processed by the target code, an optimizing transformation may or may not be profitable.

We distinguish two classes of binary transformations:

1. code transformations that can be handled directly by analyzing and modifying the original binary code. We call such transformations *low-level binary transformations*;
2. code transformations that require a higher level of abstraction of the code in order to generate a very different, but semantically equivalent, optimized code. We call such transformations *high-level binary transformations*.

While we target both classes of transformations, the first was addressed by focusing on SSE to AVX transformations of vectorized codes [20].

In this work, we focus on SIMD ISA extensions, and in particular on the x86 SSE and AVX capabilities. Compared to SSE, AVX provides wider registers, new instructions, and new addressing formats. AVX has been first supported in 2011 by the Intel Sandy Bridge and AMD Bulldozer architectures. However, most existing applications take advantage only of SSE and miss significant opportunities. We show that it is possible to automatically convert SSE to AVX whenever profitable. The key characteristics of our approach are:

- we apply the transformation at run-time, i.e. when the hardware is known;
- we only transform hot loops (detected through very lightweight profiling), thus minimizing the overhead;
- we do *not* implement a vectorization algorithm in a dynamic optimizer, instead we recognize already statically vectorized loops, and convert them to a more powerful ISA at low cost.

For high-level binary transformations, we also focus on hot loops and loop nests appearing in executable codes. There is an important literature addressing automatic loop optimization and parallelization techniques. Such optimizations include combinations of loop interchange, loop fusion and fission, loop skewing, loop shifting and loop tiling. However, they are mostly applied at compile-time, either on the source code, or on an intermediate representation form of the code. The most advanced techniques are related to the polyhedral model.

Applying such advanced loop optimizing transformations at runtime, on a currently running binary code, without any previous knowledge, is our challenging goal. The same goal has been addressed in [8], but not at runtime. In this work, the binary code is analyzed and transformed without any constraint regarding the related time overhead. Candidate loops are identified regarding their compliance to the polyhedral model: the loop bounds and memory references must be convertible into linear functions of the loop indices. Then, compliant loop nests are translated into an equivalent program in C source code, in order to be used as input for the source-to-source polyhedral compiler Pluto [29]. The resulting optimized code is then compiled and re-injected into the original binary code.

While a similar approach should be considered to reach the same goal at runtime, it must be handled differently regarding three main issues:

1. At runtime, the time overhead of the employed analysis and optimization techniques must be small. Thus, any translation to source code, that would require costly steps for the de-compilation/re-compilation phases, must be avoided.
2. Static approaches, as the one presented in [8], can only handle loops that are syntactically compliant with the polyhedral model. However, it has been shown, with the Apollo framework, that loops

may exhibit a compliant behavior at runtime. Since we target runtime optimizations, we also can take advantage of the information that is only available at runtime, and maybe also use speculative techniques.

3. Binary codes may hide some interesting properties of the embedded loops, and may need very complex analysis techniques for discovering such properties. In short, a whole compiler for binary codes would be required.

To address these issues, we are currently investigating the strategy consisting first of translating, at runtime, any selected loop nest into the LLVM [0] intermediate representation form (LLVM-IR). This representation offers several advantages:

- Analysis and transformation passes of the LLVM compiler can be used on-the-fly, in order to discover and compute relevant information, and to safely transform the code;
- The LLVM just-in-time compiler can be used to compile the optimized code, which is in LLVM-IR, as an executable;
- Existing tools for loop optimization can be used, as Polly [0], for static polyhedral-compliant loops, or Apollo, for dynamic polyhedral-compliant loops.

Hence, this strategy requires a fast binary-to-LLVM-IR translator. For this purpose, we are currently using and extending McSema [0], which is a library for translating the semantics of native code to LLVM-IR. McSema supports translation of x86 machine code, including integer, floating point, and SSE instructions. Control flow recovery is separated from translation, permitting the use of custom control flow recovery front-ends.

For McSema to be able to handle mostly any code, we had to parametrize carefully its translation rules, and also to add some x86 SSE instructions that were not handled. McSema was recently plugged to the Padrone platform. Thus, any hot loop nest is now automatically converted into LLVM-IR, as illustrated in Figure 2 .

Instead of taking as input a binary file, McSema takes as input a code extract containing a hot loop nest, thanks to the code address provided by Padrone. Then, McSema builds the control flow graph of the input code and generates a corresponding LLVM-IR. The next step is to plug the polyhedral LLVM compiler Polly (phases *Canonicalication* to *CodeGeneration* in Figure 2), in order to generate automatically an optimized version of the target loop nest, that will be then compiled using the LLVM just-in-time compiler and re-injected in the running code.

7.7. Combining Locking and Data Management Interfaces

Participants: Jens Gustedt, Mariem Saied, Daniel Salas.

Handling data consistency in parallel and distributed settings is a challenging task, in particular if we want to allow for an easy to handle asynchronism between tasks. Our publication [5] shows how to produce deadlock-free iterative programs that implement strong overlapping between communication, IO and computation. The collaboration with Soumeya Hernane has continued after her thesis defence in 2013. It extends distributed lock mechanisms and combines them with implicit data management, and resulted in a journal submission, see [26].

A new implementation (ORWL) of our ideas of combining control and data management in C has been undertaken, see 6.9 . In previous work it has demonstrated its efficiency for a large variety of platforms. In 2015, work on the ORWL model and library has gained vigor with the thesis of Mariem Saied (Inria) and Daniel Salas (INSERM). We also now collaborate on that subject with the TADAAM project team from Inria Bordeaux, where a postdoc has been hired through Inria funding.

In 2015, a new domain specific language (DSL) has been developed that largely eases the implementation of applications with ORWL. In its first version it provides an interface for stencil codes, but extensions towards other types of applications are on their way. In addition, work has been started to encapsulate imaging applications that use certain pipeline patterns to describe dependencies between computational task.

[0] http://llvm.org
[0] http://polly.llvm.org
[0] https://github.com/trailofbits/mcsema

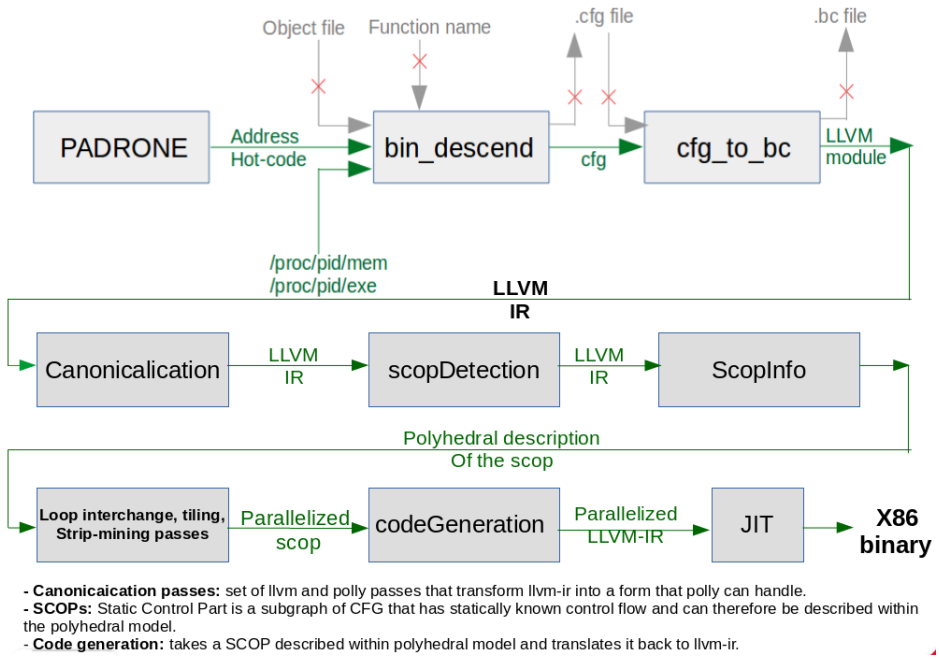

Figure 2. High-level Binary Loop Optimization through LLVM-IR

7.8. Efficient Execution of Polyhedral Codes on GPU and CPU+GPU Systems

Participants: Jean-François Dollinger, Vincent Loechner.

This is the main result of Jean-François Dollinger's PhD, started in 2012 and defended on July the 1st, 2015 [11].

Recent architectures complexity makes it difficult to statically predict the performance of a program. We have developped a reliable and accurate parallel loop nests execution time prediction method on GPUs for polyhedral codes. It is entirely automatic, and it based on three stages: static code generation, offline profiling on the target architecture, and online prediction.

In addition, we derived two techniques to fully exploit the computing resources at disposal on a computer. The first technique consists in jointly using all CPU cores and GPUs for executing a code. In order to achieve good performance, it is mandatory to consider load balance, in particular by predicting the execution time of a loop nest distributed on all those processing units. The runtime scheduler uses the profiling results to predict the execution times and adjust the parallel loop bounds to ensure load balance. The second technique puts CPU and GPU in a competition: instances of the considered code are simultaneously executed on CPU and GPU. The winner of the competition notifies its completion to the other instance, implying its termination.

7.9. Interactive Code Restructuring

Participants: Cédric Bastoul, Oleksandr Zinenko, Stéphane Huot.

This work falls within the exploration and development of semi-automatic programs optimization techniques. It consists in designing and evaluating new visualization and interaction techniques for code restructuring, by defining and taking advantage of visual representations of the underlying mathematical model. The main goal is to assist programmers during program optimization tasks in a safe and efficient way, even if they neither

have expertise into code restructuring nor knowledge of the underlying theories. This project is an important step for the efficient use and wider acceptance of semi-automatic optimization techniques, which are still tedious to use and incomprehensible for most programmers. More generally, this research is also investigating new presentation and manipulation techniques for code, algorithms and programs, which could lead to many practical applications: collaboration, tracking and verification of changes, visual search in large amount of code, teaching, etc.

This is a rather new research direction which strengthens CAMUS's static parallelization and optimization issue. It is a joint work with two Inria teams specialized in interaction: EX-SITU at Inria Saclay (contact: Oleksandr Zinenko) and MJOLNIR at Inria Lille (contact: Stéphane Huot).

In 2015, we presented our interactive tool, *Clint*, that maps direct manipulation of the visual representation to polyhedral program transformations with real-time semantics preservation feedback. We conducted two user studies showing that Clint's visualization can be accurately understood by both experts and non-expert programmers, and that the parallelism can be extracted better from Clint's representation than from the source code in many cases [21]. We are planing a first release of that tool in the coming year.

7.10. Automatic Generation of Adaptive Simulation Codes

Participants: Cédric Bastoul, César Sabater.

Compiler automatic optimization and parallelization techniques are well suited for some classes of simulation or signal processing applications, however they usually don't take into account neither domain-specific knowledge nor the possibility to change or to remove some computations to achieve "good enough" results. Quite differently, production simulation and signal processing codes have adaptive capabilities: they are designed to compute precise results only where it matters if the complete problem is not tractable or if the computation time must be short. In this research, we design a new way to provide adaptive capabilities to compute-intensive codes automatically, inspired by Adaptive Mesh Refinement a classical numerical analysis technique to achieve precise computation only in pertinent areas. It relies on domain-specific knowledge provided through special pragmas by the programmer in the input code and on polyhedral compilation techniques, to continuously regenerate at runtime a code that performs heavy computations only where it matters at every moment. A case study on a fluid simulation application shows that our strategy enables dramatic computation savings in the optimized portion of the application while maintaining good precision, with a minimal effort from the programmer.

This research direction started in 2015 and complements our other efforts on dynamic optimization. We are in the process of a collaboration with Inria Nancy Grand Est team TONUS, specialized on applied mathematics (contact: Philippe Helluy), to bring models and techniques from this field to compilers. First results, investigated during the Inria Internship Program of César Sabater, have been presented to the SimRace international conference dedicated on industrial fluid simulation applications [16].

7.11. Polyhedral Compiler White-Boxing

Participants: Cédric Bastoul, Lénaïc Bagnères, Oleksandr Zinenko, Stéphane Huot.

While compilers offer a fair trade-off between productivity and executable performance in single-threaded execution, their optimizations remain fragile when addressing compute-intensive code for parallel architectures with deep memory hierarchies. Moreover, these optimizations operate as black boxes, impenetrable for the user, leaving them with no alternative to time-consuming and error-prone manual optimization in cases where an imprecise cost model or a weak analysis resulted in a bad optimization decision. To address this issue, we researched and designed a technique allowing to automatically translate an arbitrary polyhedral optimization, used internally by loop-level optimization frameworks of several modern compilers, into a sequence of comprehensible syntactic transformations as long as this optimization focuses on scheduling loop iterations. With our approach, we open the black box of the polyhedral frameworks enabling users to examine, refine, replay and even design complex optimizations semi-automatically in partnership with the compiler.

This research started in 2014 and we found the first solution in 2015. It has been conducted as a joint work between teams in compiler technologies (CAMUS and Inria Saclay's POSTALE team) and teams in interaction (EX-SITU at Inria Saclay and MJOLNIR at Inria Lille). The first paper on this has been accepted in 2015 to be presented in one of the top conferences on optimization techniques: CGO 2016 [15]. Subsequent work and a first release of the tool implementing the technique is planned during 2016.

CAPSID Project-Team

7. New Results

7.1. Annotating 3D Protein Domains

Many protein chains in the Protein Data Bank (PDB) are cross-referenced with EC numbers and Pfam domains. However, these annotations do not explicitly indicate any relation between EC numbers and Pfam domains. In order to address this limitation, we developed EC-DomainMiner, a recommender-based approach for associating EC (Enzyme Commission) numbers with Pfam domains [19]. EC-DomainMiner is able to infer automatically 20,179 associations between EC numbers and Pfam domains from existing EC-chain/Pfam-chain associations from the SIFTS database as well as EC-sequence/Pfam-sequence associations from UniProt databases.

7.2. Large-Scale Analysis of 3D Protein Interactions

As part of a continuing collaboration with a former doctoral student in the Orpailleur team, Anisah Ghoorah (now at the University of Mauritius), we used her KBDOCK database of all known PPIs to perform a large-scale statistical analysis of the secondary structure composition of known protein-protein binding sites [14]. This showed that some combinations of secondary structure features are significantly favoured, whereas other combinations are considerably dis-favoured. These findings could provide knowledge-based rules for the prediction of unsolved protein-protein interactions.

7.3. Predicting Drug Side Effects

Together with Harmonic Pharma SAS (a LORIA / Inria spin-out company), we developed the “GESSE” method for proposing new uses for existing therapeutic drug molecules by associating the Gaussian shapes of known drug molecules with their clinically observed side-effects [15].

7.4. Modeling a GPCR Receptor Complex

In collaboration with the BIOS team (INRA Tours) and the AMIB team (Inria Saclay – Île de France) we used our Hex protein docking software to help model a multi-component G-protein coupled receptor (GPCR) complex [12]. The resulting 3D structure was shown to be consistent with the known experimental data for the protein components of this trans-membrane molecular signaling system.

7.5. Modeling the Apelin Receptor

The Apelin receptor (ApelinR) is a GPCR which is important in regulating cardiovascular homeostasis. As part of an on-going collaboration with the Centre for Interdisciplinary Research (CIRB) at Collège de France, we modeled the interaction between the Apelin peptide and ApelinR [13]. This study provides new mechanistic insights which could lead to the development of therapeutic agents for the treatment of heart failure.

7.6. Identifying New Anti-Fungal Agents

In this collaboration with several Brasilian laboratories (at University of Mato Grosso State, University of Maringá, Embrapa, and University of Brasilia), we identified several novel small-molecule drug leads against the pathogenic fungus *Paracoccidioides lutzii* [17] which is a serious health threat, especially in Brasilian hospitals.

CARAMEL Project-Team

7. New Results

7.1. The Logjam attack against the discrete logarithm

Participants: Pierrick Gaudry, Emmanuel Thomé [contact], Paul Zimmermann.

Together with colleagues from the Prosecco project-team and with other colleagues, we exhibited a new attack again the TLS protocol when using discrete logarithms [15]. A proof-of-concept of the attack was demonstrated using the CADO-NFS software. This paper obtained the best paper award at the ACM CCS 2015 conference, and received significant media coverage both in the specialized and non-specialized press.

7.2. Other results related to discrete logarithm

Participant: Pierrick Gaudry [contact].

Our 2014 work [16], in collaboration with Barbulescu, Guillevic and Morain, improving the practical aspects of discrete logarithm computation in quadratic extensions and reducing the theoretical complexity in the "medium characteristic case" has been published in Eurocrypt 2015.

In collaboration with Barbulescu and Kleinjung we have proposed in [17] to revisit an old construction of Schirokauer for discrete logarithms in extension fields. It is well suited for problems coming from pairings where the primes often have a special form.

With Galbraith we wrote a survey about the discrete logarithm problem in the context of elliptic curves [13].

7.3. Fast arithmetic for faster integer multiplication

Participants: Svyatoslav Covanov [contact], Emmanuel Thomé.

The paper [20] describes an algorithm for the multiplication of two n-bit integers. It achieves the best asymptotic complexity bound $O(n \log n \cdot 4^{\log^* n})$ under a hypothesis on the distribution of generalized Fermat primes of the form $r^{2^\lambda} + 1$. This hypothesis states that there always exists a sufficiently small interval in which we can find such a prime. Experimental results give evidence in favor of this assumption. A journal submission is planned shortly.

7.4. Certificates for exact linear algebra computations

Participant: Emmanuel Thomé [contact].

The paper [21], in collaboration with Jean-Guillaume Dumas and Erich Kaltofen, is a preliminary version of a research work that has then been pursued, and that solves an open question of proving the correctness of some specific linear algebra computations. It emerged from practical techniques which had been used for this purpose for a while, and for which improvements were obtained. Submission plans for this work are yet to be finalized.

7.5. Computing Jacobi's theta function in quasi-linear time

Participant: Hugo Labrande [contact].

We designed a new algorithm that improves the complexity of computing the value of the Jacobi theta function, $\theta(z, \tau)$ to arbitrary precision [23]. The algorithm uses a quadratically convergent sequence similar to the complex AGM, as well as Newton's method; its complexity is $O(\mathcal{M}(n) \log n)$ for computing the value up to an error bounded by 2^{-n}, which is an improvement over the state-of-the-art complexity of $O(\mathcal{M}(n)\sqrt{n})$. Here, $\mathcal{M}(n)$ denotes the time taken by a multiplication of two n-bit numbers. We provide bounds on the loss of significant digits incurred during the computation. The algorithm was implemented using GNU MPC, showing practical improvement over (our optimized implementation of) existing algorithms for precision above approximately $300,000$ bits. The paper was submitted to *Mathematics of Computation*.

7.6. Construction of sparse polynomial systems with many positive solutions

Participant: Pierre-Jean Spaenlehauer [contact].

In collaboration with Frédéric Bihan (Univ. Savoie Mont-Blanc), we propose a variant of the classical Viro method to construct polynomial systems with prescribed monomial support and many solutions whose coordinates are all positive [19]. This is an asymptotic construction which has strong connections with tropical and convex geometry, and which involves computational problems such as low-rank matrix completion.

7.7. Small certificates of inconsistency of quadratic fewnomial systems

Participant: Pierre-Jean Spaenlehauer [contact].

In collaboration with Jean-Charles Faugère (EPI PolSys) and Jules Svartz (Min. de Éducation Nationale), we studied the problem of certifying the inconsistency of sparse quadratic polynomial systems. Finding certificates of inconsistency is a classical problem in computational commutative algebra, and these certificates are in general of size exponential in the input size. We identify families of quadratic fewnomial systems for which there exist certificates of size linear in the size of the input and we propose algorithms to compute them in polynomial time.

7.8. Cracking passphrases based on famous sentences

Participant: Hugo Labrande [contact].

We proposed a method to attack passwords based on famous sentences, which are rather widespread [18]: we showed a method to construct large dictionaries using only publicly-available sources (e.g. Wikipedia) and modest computing power. The resulting dictionaries were able to crack millions of passphrases, among which a 55-character long one, and some that do not appear to have been cracked before. Our work thus shows that using famous sentences as passwords is not secure at all, as any attacker, even those with low skills and very modest computational resources, can guess them.

CARTE Project-Team

7. New Results

7.1. Computability and Complexity

- **Complexity of stream functions and higher-order complexity.** We have pursued our works on higher-order complexity and the complexity of stream functions. Both notions are closely related as any function from natural numbers to natural numbers can be seen as a stream (an infinite list) of natural numbers:
  - A characterization of the class of Basic Feasible Functionals using term rewrite systems on streams and interpretation methods has been proposed in [13]. This result is part of Hugo Férée's PhD thesis for which he has obtained the Ackermann award.
  - In [14], we have provided some interpretation criteria useful to ensure two kinds of stream properties: space upper bounds and input/output upper bounds. Our space upper bounds criterion ensures global and local upper bounds on the size of each output stream element expressed in term of the maximal size of the input stream elements. The input/output upper bounds criterion considers instead the relations between the number of elements read from the input stream and the number of elements produced on the output stream.
  - The paper [21] has extended the light affine lambda calculus with inductive and coinductive data types using the category theory notions of (weak) initial algebra and coalgebra.
- **Complexity analysis of Object-Oriented programs.** We have proposed a type system based on non-interference and data ramification (tiering) principles in [22]. We have captured the set of functions computable in polynomial time on OO programs. The studied language is general enough to capture most OO constructs and the characterization is quite expressive as it allows the analysis of a combination of imperative loops and of data ramification scheme based on Bellantoni and Cook's safe recursion using function algebra.
- **Rice-like theorem for primitive recursive functions.** We have studied the following question: what are the properties of primitive recursive functions that are decidable (by a Turing machine), given a primitive recursive presentation of the function. We give a complete characterization of these properties. We show that they can be expressed as unions of elementary properties of being compressible. If $h : \mathbb{N} \to \mathbb{N}$ is a computable increasing unbounded function (like $\log(n)$ or 2^n), we say that a function $f : \mathbb{N} \to \mathbb{N}$ is h-compressible if for each n there is a primitive recursive program of size at most $h(n)$ computing a function that coincides with f on entries $0, 1, ..., n$. Whether f is h-compressible is decidable given a primitive recursive program for f, and every decidable property can be obtained as a combination of such elementary properties. This result actually holds for any class of total functions that admits a sound and complete programming language. An article is currently in preparation.
- **Parametrization of geometric figures.** During the master internship of Diego Nava Saucedo, we have studied the semi-computability of geometric figures. A figure is semi-computable if there is a program that semi-decides whether a pixel intersects the figure. Our goal is to understand the semi-computability of a figure in terms of the parameters describing the figure. It turns out that the usual ways of parameterizing simple figures such as triangles, squares or disks do not behave well in terms of semi-computability. We have actually proved that no *finite* parametrization behaves well.
- **Symbolic Dynamics on Groups.** In an effort to better understand the interplay of geometry and computability in tiling theory, E. Jeandel has studied tiling problems on general Cayley graphs, and has obtained a significant number of new results. He has proven that groups with an (strongly) aperiodic tiling system have decidable word problem [30], and provided examples of new groups (in particular monster groups) with such tiling systems, and proved that all nontrivial nilpotent groups

have an aperiodic tiling system and an undecidable domino problem [31]. He also showed how the new concept of translation-like actions from geometric group theory can be used to prove that many groups, in particular the Grigorchuk groups and most groups with a nontrivial center, have an undecidable domino problem [33].

- **The smallest aperiodic tileset.** In a joint work with Michael Rao, E. Jeandel has proven that there exists an aperiocic set of 11 Wang tiles [34], and furthermore that this number is optimal.

7.2. Quantum Computing

- **On Weak Odd Domination and Graph-based Quantum Secret Sharing.** In this work published in the journal Theoretical Computer Science [15], Simon Perdrix and his co-authors Sylvain Gravier, Jérôme Javelle and Mehdi Mhalla study weak odd domination in graphs and its application in quantum secret sharing. A weak odd dominated (WOD) set in a graph is a subset B of vertices for which there exists a distinct set of vertices C such that every vertex in B has an odd number of neighbors in C. They point out the connections of weak odd domination with odd domination, [σ,ρ]-domination, and perfect codes. They introduce bounds on κ(G), the maximum size of WOD sets of a graph G, and on κ'(G), the minimum size of non-WOD sets of G. Moreover, they prove that the corresponding decision problems are NP-complete. The study of weak odd domination is mainly motivated by the design of graph-based quantum secret sharing protocols: a graph G of order n corresponds to a secret sharing protocol whose threshold is κQ(G)=max(κ(G),n-κ'(G)). These graph-based protocols are very promising in terms of physical implementation, however all such graph-based protocols studied in the literature have quasi-unanimity thresholds (i.e. κQ(G)=n-o(n) where n is the order of the graph G underlying the protocol). In this paper, they show using probabilistic methods the existence of graphs with smaller κQ (i.e. κQ(G)$\leq$0.811n where n is the order of G). They also prove that deciding for a given graph G whether κQ(G)$\leq$k is NP-complete, which means that one cannot efficiently double check that a graph randomly generated has actually a κQ smaller than 0.811n.
- **Minimum Degree up to Local Complementation: Bounds, Parameterized Complexity, and Exact Algorithms.** In this work presented at ISAAC [25], David Cattaneo and Simon Perdrix introduce new upper bounds and exact algorithms for the local minimum degree. The author also prove the W[2]-membership of the corresponding decision problem. The local minimum degree of a graph is the minimum degree that can be reached by means of local complementation. For any n, there exist graphs of order n which have a local minimum degree at least $0.189n$, or at least $0.110n$ when restricted to bipartite graphs. Regarding the upper bound, they show that the local minimum degree is at most $3/8n + o(n)$ for general graphs and $n/4 + o(n)$ for bipartite graphs, improving the known $n/2$ upper bound. They also prove that the local minimum degree is smaller than half of the vertex cover number (up to a logarithmic term). The local minimum degree problem is NP-Complete and hard to approximate. They show that this problem, even when restricted to bipartite graphs, is in W[2] and FPT-equivalent to the EvenSet problem, whose W[1]-hardness is a long standing open question. Finally, they show that the local minimum degree is computed by a $O_*(1.938n)$-algorithm, and a $O_*(1.466n)$-algorithm for the bipartite graphs.
- **The ZX Calculus is incomplete for Clifford+T quantum mechanics.** The ZX calculus is a diagrammatic language for quantum mechanics and quantum information processing. In this paper[17], Simon Perdrix and Harny Wang prove that the ZX-calculus is not complete for the Clifford+T quantum mechanics. The completeness for this fragment has been stated as one of the main current open problems in categorical quantum mechanics. The ZX calculus was known to be incomplete for quantum mechanics, on the other hand, it has been proved complete for Clifford quantum mechanics (a.k.a. stabilizer quantum mechanics), and for single-qubit Clifford+T quantum mechanics. The question of the completeness of the ZX calculus for Clifford+T is a crucial step in the development of the ZX calculus because of its (approximate) universality for quantum mechanics (i.e. any unitary evolution can be approximated using Clifford and T gates only). They exhibit a property which is know to be true in Clifford+T quantum mechanics and prove that this equation cannot be derived

in the ZX calculus, by introducing a new sound interpretation of the ZX calculus in which this particular property does not hold. Finally, we propose to extend the language with a new axiom. This result has been presented as invited speakers in the conferences "Quantum Theory: from foundations to technologies" in Vaxjo Sweden, and "Higher TQFT and categorical quantum mechanics" at the Scrounger Institute in Vienna. The authors also presented these results at the workshop of the CNRS groupe de travail Informatique Quantique du GDR IM, in Grenoble.

- **Block Representation of Reversible Causal Graph Dynamics.** In this work presented at the conference on Foundation of computer science (FCT'15) [18], Pablo Arrighi, Simon Martiel and Simon Perdrix, consider a reversible version of the causal graph dynamics. Causal Graph Dynamics extend Cellular Automata to arbitrary, bounded-degree, time-varying graphs. The whole graph evolves in discrete time steps, and this global evolution is required to have a number of physics-like symmetries: shift-invariance (it acts everywhere the same) and causality (information has a bounded speed of propagation). We study a further physics-like symmetry, namely reversibility. More precisely, we show that Reversible Causal Graph Dynamics can be represented as finite-depth circuits of local reversible gates.
- **Reversibility in the Extended Measurement-based Quantum Computation.** In this work by Nidal Hamrit and Simon Perdrix has been presented at the conference on Reversible Computation in Grenoble [23]. When applied on some particular quantum entangled states, measurements are universal for quantum computing. In particular, despite the fondamental probabilistic evolution of quantum measurements, any unitary evolution can be simulated by a measurement-based quantum computer (MBQC). They consider the extended version of the MBQC where each measurement can occur not only in the X,Y-plane of the Bloch sphere but also in the X,Z- and Y,Z-planes. The existence of a gflow in the underlying graph of the computation is a necessary and sufficient condition for a certain kind of determinism. They extend the focused gflow (a gflow in a particular normal form) defined for the X,Y-plane to the extended case, and provide necessary and sufficient conditions for the existence of such normal forms.
- **Quantum Circuits for the Unitary Permutation Problem.** In this paper presented at TAMC'15 [20] Stefano Facchni and Simon Perdrix consider the *Unitary Permutation* problem which consists, given n quantum gates $U_1, ..., U_n$ and a permutation σ of $\{1, ..., n\}$, in applying the quantum gates in the order specified by σ, i.e., in performing $U_{\sigma(n)} \circ ... \circ U_{\sigma(1)}$. This problem has been introduced and investigated in [47] where two models of computations are considered. The first is the (standard) model of query complexity: the complexity measure is the number of calls to any of the quantum gates U_i in a quantum circuit which solves the problem. The second model is roughly speaking a model for higher order quantum computation, where quantum gates can be treated as objects of second order. In both model the existing bounds are improved, in particular the upper and lower bounds for the standard quantum circuit model are established by pointing out connections with the *permutation as substring* problem introduced by Karp.

CASSIS Project-Team

7. New Results

7.1. Automated Deduction

We develop general techniques which allow us to re-use available tools in order to build a new generation of solvers offering a good trade-off between expressiveness, flexibility, and scalability. We focus on the careful integration of combination techniques and rewriting techniques to design decision procedures for a wide range of verification problems.

7.1.1. Building and Verifying decision procedures

Participants: Alain Giorgetti, Olga Kouchnarenko, Christophe Ringeissen.

In the context of the PhD thesis by Elena Tushkanova (defended in 2013), we have developed a methodology to build decision procedures specified by using a superposition calculus [20] which is at the core of all equational theorem provers. This calculus is refutation complete: it provides a semi-decision procedure that halts on unsatisfiable inputs but may diverge on satisfiable ones. Fortunately, it may also terminate for some theories of interest in verification, and thus it becomes a decision procedure. To reason on the superposition calculus, a schematic superposition calculus has been developed to build the schematic form of the saturations allowing to automatically prove decidability of single theories and of their combinations. We have proposed a rule-based logical framework and a tool implementing a complete many-sorted schematic superposition calculus for arbitrary theories. By providing results for unit theories, arbitrary theories, and also for theories with counting operators, we show that this tool is very useful to derive decidability and combinability of theories of practical interest in verification.

7.1.2. Combination of Satisfiability Procedures

Participant: Christophe Ringeissen.

We have continued our work started with Paula Chocron (IIIA-CSIC, U. Barcelona) and Pascal Fontaine (project-team Veridis) on the extension of the Nelson-Oppen combination method to non-disjoint unions of theories. We investigate the case of theories connected via bridging functions [28]. The motivation is, e.g., to solve verification problems expressed in a combination of data structures connected to arithmetic with bridging functions such as the length of lists and the size of trees. We present a sound and complete combination procedure à la Nelson-Oppen for the theory of absolutely free data structures, including lists and trees. This combination procedure is then refined for standard interpretations. The resulting theory has a nice politeness property, enabling combinations with arbitrary decidable theories of elements.

To go beyond the case of absolutely free data structures, we study in [29] the problem of determining the data structure theories for which this combination method is sound and complete. Our completeness proof is based on a rewriting approach where the bridging function is defined as a term rewrite system, and the data structure theory is given by a basic congruence relation. Our contribution is to introduce a class of data structure theories that are combinable with a disjoint target theory via an inductively defined bridging function. This class includes the theory of equality, the theory of absolutely free data structures, and all the theories in between. Hence, our non-disjoint combination method applies to many classical data structure theories admitting a rewrite-based satisfiability procedure.

7.1.3. Unification Modulo Equational Theories

Participant: Christophe Ringeissen.

We investigate a hierarchical combination approach to the unification problem in non-disjoint unions of equational theories. In this approach, the idea is to extend a base theory with some additional axioms given by rewrite rules in such way that the unification algorithm known for the base theory can be reused without loss of completeness. Additional techniques are required to solve a combined problem by reducing it to a problem in the base theory. In [33] we show that the hierarchical combination approach applies successfully to some classes of syntactic theories, such as shallow theories since the required unification algorithms needed for the combination algorithm can always be obtained. We also consider the matching problem in syntactic extensions of a base theory. Due to the more restricted nature of the matching problem, we obtain several improvements over the unification problem.

7.1.4. Enumeration of Planar Proof Terms

Participant: Alain Giorgetti.

By the Curry-Howard isomorphism, simply typed lambda terms correspond to natural deduction proofs in minimal logic. Noam Zeilberger and Alain Giorgetti proved that normal planar lambda terms are in size-preserving bijection with rooted planar maps [21]. Although the formal aspect is not emphasized in the paper, the use of formal representations of both normal planar lambda terms and rooted planar maps, of logic programming and a proof assistant software helped much in more quickly finding the bijection.

7.1.5. Rewriting-based Mathematical Model Transformations

Participants: Walid Belkhir, Alain Giorgetti.

Since 2011 we collaborate with the Department “Temps-Fréquence” of the FEMTO-ST institute (Franche-Comté Electronique Mécanique Thermique et Optique - Sciences et Technologies, CNRS UMR 6174) on the formalization of asymptotic methods (based on two-scale convergence). The goal is to design a software, called *MEMSALab*, for the automatic derivation of multiscale models of arrays of micro- and nanosystems. In this domain a model is a partial differential equation. Multiscale methods approximate it by another partial differential equation which can be numerically simulated in a reasonable time. The challenge consists in taking into account a wide range of different physical features and geometries e.g. thin structures, periodic structures, multiple nested scales etc. In [24], we propose a method called "*by-extension-combination*", in which the asymptotic models are constructed incrementally so that model features can be included step by step. More precisely, a model derivation is formalised as a rewriting strategy, and its extension is formalised as a second-order rewriting strategy. Thus, our method amounts to defining an operation of combination over a class of second-order rewriting strategies. We illustrate the method by an example of an asymptotic model for the stationary heat equation in a Micro-Mirror Array developed for astrophysics.

7.2. Security Protocol Verification

The design of cryptographic protocols is error-prone. Without a careful analysis, subtle flaws may be discovered several years after the publication of a protocol, yielding potential harmful attacks. In this context, formal methods have proved their interest for obtaining good security guarantees. Many analysis techniques have been proposed in the literature [66]. We have edited a book [62] where each chapter presents an important and now standard analysis technique. We develop new techniques for richer primitives, wider classes of protocols and higher security guarantees. In Section 7.4.3 we consider derived testing techniques for verifying protocol implementations.

7.2.1. Design of Voting Protocols

Participants: Véronique Cortier, Stéphane Glondu, Steve Kremer, Peter Rønne.

Voting is a cornerstone of democracy and many voting systems have been proposed so far, from old paper ballot systems to purely electronic voting schemes. Although many works have been dedicated to standard protocols, very few address the challenging class of voting protocols.

One famous e-voting protocol is Helios, an open-source web-based end-to-end verifiable electronic voting system, used e.g., by UCL and the IACR association in real elections. One main advantage of Helios is its verifiability, up-to the ballot box (a dishonest ballot box may add ballots). We have defined a variant of Helios, named Belenios, that prevents from ballot stuffing, even against a dishonest ballot box. Our approach consists in introducing an additional authority that provides credentials that the ballot box can verify but not forge. Belenios[0] has been implemented by Stéphane Glondu (cf Section 6.1.3).

Helios as well as Belenios are not receipt-free, that is, a (malicious) voter can *prove* how they voted to any third party. Building upon a scheme proposed by G. Fuschbauer and David Pointcheval, we have enhanced Belenios with a receipt-free variant, called BeleniosRF. Now, the ballot box can re-randomize any (signed) ballot it receives. This way, a voter can no longer exhibit the randomness they used to build their ballot.

End-to-end verifiable voting schemes typically involves voters handling an encrypted ballot in order to confirm that their vote is accurately included in the tally. While this may be technically valid, from a public acceptance standpoint it may be problematic: many voters may not really understand the purpose of the encrypted ballot and the various checks that they can perform. In [61] we take a different approach and revisit an old idea: to provide each voter with a private tracking number. Votes are posted on a bulletin board in the clear along with their associated tracking number. This is appealing in that it provides voters with a very simple, intuitive way to verify their vote, in the clear. However, there are obvious drawbacks: we must ensure that no two voters are assigned the same tracker and we need to keep the trackers private. We propose a new scheme, called Selene, that addresses both of these problems: we ensure that voters get unique trackers and we close off the coercer's window of opportunity by ensuring that the voters only learn their tracking numbers after votes have been posted. The resulting scheme provides receipt-freeness, and indeed a good level of coercion-resistance while also providing a more immediately understandable form of verifiability. The cryptography is under the bonnet as far as the voter is concerned.

In 2010 Hao, Ryan and Zielinski proposed a simple decentralised e-voting protocol that only requires 2 rounds of communication. Thus, for k elections their protocol needs $2k$ rounds of communication. Observing that the first round of their protocol is aimed to establish the public-keys of the voters, we propose in [60] an extension of the protocol as a non-interactive e-voting scheme in the public-key setting (NIVS) in which the voters, after having published their public-keys, can use the corresponding secret-keys to participate in an arbitrary number of one-round elections. We first construct a NIVS with a standard tally function where the number of votes for each candidate is counted. Further, we present constructions for two alternative types of elections. Specifically in the first type (dead or alive elections) the tally shows if at least one voter cast a vote for the candidate. In the second one (elections by unanimity), the tally shows if all voters cast a vote for the candidate. Our constructions are based on bilinear groups of prime order. As definitional contribution we provide formal computational definitions for privacy and verifiability of NIVSs. We conclude by showing intriguing relations between our results, secure computation, electronic exams and conference management systems.

7.2.2. Analysis of Voting Protocols

Participants: Véronique Cortier, Catalin Dragan, Steve Kremer, Peter Rønne.

Properties. Even a basic property like ballot secrecy is difficult to define formally and several definitions co-exist. We studied all game-based privacy definitions of the literature and discovered that none of them was satisfactory: they were either limited (not fully modeling e-voting protocols), or too strong (incompatible with verifiability), or even flawed for a few of them [25]. Based on our findings, we have proposed a new game-based privacy definition BPRIV, proved that it implies simulation-based privacy and showed that it is realized by the Helios protocol [25].

Proof. Such a proof of privacy for Helios is done by hand and is error-prone. Moreover, there is not a single version of Helios. Instead, many slight variants of Helios may be considered (e.g. early and late weeding, weeding based on the identity or on the ciphertexts, mixnet or homomorphic tally, etc.). Each of these variants would require a new proof. Therefore, we are conducting a proof of Helios and Belenios through the Easycrypt framework. This first fully formal proof will cover most existing variants of Helios and Belenios.

[0] https://belenios.loria.fr

Analysis. Existing automated analysis techniques are inadequate to deal with commonly used cryptographic primitives, such as homomorphic encryption and mix-nets, as well as some fundamental security properties, such as verifiability. In collaboration with Matteo Maffei and Fabienne Eigner (Saarland University) we propose a novel approach based on refinement type systems for the automated analysis of two fundamental properties of e-voting protocols, namely, vote privacy and verifiability. We demonstrate the effectiveness of our approach by developing the first automated analysis of Helios using an off-the-shelf type-checker [32].

A challenging problem in e-voting is to provide guarantees when the voting platform itself is corrupted. Du-Vote [73] is a recently presented remote electronic voting scheme that aims to be malware tolerant, i.e., provide security even in the case where the platform used for voting has been compromised by dedicated malware. For this it uses an additional hardware token, similar to tokens distributed in the context of online banking. Du-Vote aims at providing vote privacy as long as either the vote platform or the vote server is honest. For verifiability, the security guarantees are even higher, as even if the token's software has been changed, and the platform and the server are colluding, attempts to change the election outcome should be detected with high probability. In recent work [41] we provide an extensive security analysis of Du-Vote and show several attacks on both privacy as well as verifiability. We also propose changes to the system that would avoid many of these attacks.

7.2.3. Other Families of Protocols

Participants: Véronique Cortier, Jannik Dreier, Alicia Filipiak, Steve Kremer, Ludovic Robin.

Secure Mobile Applications. There is a growing development of Secure Elements for Mobile Phone and Tablets. These Secure Elements are hosted in the SIM for example and can perform cryptographic operations. This opens the way for a much higher level of security in such environnements. However, how to use these secure elements is still very unclear. How keys will be registered in Secure Elements? Which applications may access to the keys and how is this enforced? Which part of the application should be deployed in a Secure Element? It is of course not possible to host an entire application in a Secure Element for size and performance issues. Alicia Filipiak has started a PhD in March 2015 to propose a model for secure mobile applications that make use of Secure Elements. This is a collaboration with Orange Labs (CIFRE). She has proposed a light and secure paiement application which is compatible with standard paiement systems (EMV). The proof of security is conducted in Tamarin, in order to cope with global states.

Protocols using low-entropy secrets. Many two factor authentication protocols consider an additional authentic, but low bandwidth channel to send a confirmation code. A typical example is to send such a code by SMS to a user's mobile phone. Given that such codes need to be copied by users they are short and therefore vulnerable to offline brute-force attacks. Ludovic Robin has started a PhD thesis in October 2014 and proposed a model to take into account an attacker's capability to run such brute-force attacks. While the problem is reminiscent to guessing attacks in password-based protocols, several subtle differences make this problem more difficult. Ludovic is adapting the decision procedure implemented in *Akiss* in order to decide protocol security in the presence of such an attacker.

Auction protocols. Auctions have a long history, having been recorded as early as 500 B.C.. Nowadays, electronic auctions have been a great success and are increasingly used in various applications. Many cryptographic protocols have been proposed to address the various security requirements of these electronic transactions, in particular to ensure privacy. Jannik Dreier, in collaboration with Pascal Lafourcade from Université d'Auvergne and Jean-Guillaume Dumas from Université Grenoble Alpes, recently performed a detailed analysis [15] of Brandt's auction protocol that computes the winner using homomorphic operations on a distributed ElGamal encryption of the bids. Jannik and his coauthors were able to show that this protocol – when using malleable interactive zero-knowledge proofs – is vulnerable to attacks by dishonest bidders. Such bidders can manipulate the publicly available data in a way that allows the seller to deduce all participants' bids. They developed an efficient parallelized implementation of the protocol and the attack to show its practicality.

7.2.4. Automated Verification of Indistinguishability Properties

Participants: Vincent Cheval, Rémy Chrétien, Véronique Cortier, Antoine Dallon, Jannik Dreier, Steve Kremer.

New emerging classes of protocols such as voting protocols often require to model less classical security properties, such as anonymity properties, strong versions of confidentiality and resistance to offline guessing attacks. Many of these properties can be modelled using the notion of indistinguishability by an adversary, which can be conveniently modeled using process equivalences.

Active case, bounded number of sessions. We previously proposed a procedure for approximating trace equivalence in the case of a bounded number of sessions, i.e., for a replication free fragment of a cryptographic process calculus. The procedure is implemented in the *Akiss* tool. While we proved soundness and correctness for any convergent rewrite system that has the finite variant property, termination of the procedure was still an open question. We have recently shown that the procedure indeed terminates for the class of subterm convergent rewrite systems. We are currently also working on an extension of *Akiss* in order to verify protocols that may use the exclusive or operator. This extensions requires us to reason modulo associativity and commutativity. While proving soundness and completeness of a naive extension of the existing procedure is a rather straightforward, the resulting procedure faces directly non-termination. We therefore adapt the resolution strategy to ensure termination on practical examples. While soundness is preserved we need to prove the completeness of the new resolution strategy.

When considering the equational theory corresponding to the standard primitives, Vincent Cheval has proposed a decision procedure for checking equivalence of set constraints, which yields a procedure for checking trace equivalence [69]. We have extended this decision procedure to the case where the attacker can observe the *time* of executions [27], capturing what is called *timing attacks*. To obtain decidability, we have shown how to reduce to a previous result to decide length trace equivalence, where the attacker no longer has access to execution times but can still compare the length of messages. As an application, we study several protocols that aim for privacy. In particular, we (automatically) detect an existing timing attack against the biometric passport and new timing attacks against the Private Authentication protocol.

Active case, unbounded number of sessions.

We have shown that for some classes of protocols, decidability of trace equivalence can be reduced to equivalence of deterministic pushdown automata [13]. Equivalence of deterministic pushdown automata is decidable [79] and the corresponding decision procedure has been recently implemented by Géraud Senizergues. Based on his tool, we have developed a tool for automatically checking equivalence, for an unbounded number of sessions.

For trace properties such as secrecy and authentication, it has been shown that it is sufficient to consider typically three agents, two honest and one dishonest agents [70]. This result no longer holds for equivalence properties. Antoine Dallon has recently started a PhD thesis on deciding equivalence properties. He has shown that it is sufficient to consider two honest agents and two dishonest agents for equivalence properties, for deterministic processes with standard primitives and without else branches. More generally, he shows how to bound the number of agents for arbitrary constructor theories and for protocols with simple else branches. These hypotheses are tight, and counter-examples are provided for non action-deterministic processes, non constructor theories, or protocols with complex else branches.

When proving security in symbolic settings for an unbounded number of sessions, a typical technique (like in the aforementioned result) consists in abstracting away fresh nonces and keys by a bounded set of constants. While this abstraction is clearly sound in the context of secrecy properties (for protocols without else branches), this is no longer the case for equivalence properties. We have shown how to soundly get rid of nonces in the context of equivalence properties [30]. We show that nonces can be replaced by constants provided that each nonce is associated to two constants (instead of typically one constant for secrecy properties). Our result holds for deterministic (simple) protocols and a large class of primitives that includes e.g. standard primitives, blind signatures, and zero-knowledge proofs.

Of course, our abstraction of nonces may introduce false attacks. To avoid this, it is necessary to consider protocols *with* nonce. We have provide the first decidability result for trace equivalence of security protocols, for an unbounded number of sessions and unlimited fresh nonces [31]. Our class encompasses most symmetric key protocols of the literature, in their tagged variant.

Decomposing equivalence. Unique decomposition has been a subject of interest in process algebra for a long time (for example in BPP or CCS in the 1980s), as it provides a normal form and useful cancellation properties. In recent work [16] Jannik Dreier, together with Cristian Ene and Yassine Lakhnech from Université Grenoble Alpes as well as Pascal Lafourcade from Université d'Auvergne, proved two parallel decomposition results for subsets of the applied π-calculus. They showed that every closed normed (i.e. with a finite shortest complete trace) process P can be decomposed uniquely into prime factors P_i with respect to strong labeled bisimilarity, i.e. such that $P \sim_l P_1|...|P_n$. Moreover, they proved that closed finite processes can be decomposed uniquely with respect to weak labeled bisimilarity. They also investigated whether efficient algorithms that compute the unique decompositions exist, which would be useful for the verification of equivalences. It turned out that the simpler problem of deciding whether a process is in its unique decomposition form is undecidable in general in both cases, due to potentially undecidable equational theories. Moreover, the unique decomposition remains undecidable even given an equational theory with a decidable word problem.

7.2.5. Securely Composing Protocols

Participants: Vincent Cheval, Véronique Cortier, Éric Le Morvan.

Protocols are often built in a modular way. For example, authentication protocols may assume pre-distributed keys or may assume secure channels. However, when an authentication protocol has been proved secure assuming pre-distributed keys, there is absolutely no guarantee that it remains secure when executing a real protocol for distributing the keys. During his PhD thesis, Éric Le Morvan has shown how to securely realize the three main types of channels: secure (unreadable and untappable), confidential (unreadable), and authenticated (untappable) channels [54].

7.3. Model-based Verification

We have investigated extensions of regular model-checking to new classes of rewrite relations on terms. We have studied specification and proof of modular imperative programs, as well as of modal workflows.

7.3.1. Tree Automata with Constraints

Participants: Pierre-Cyrille Héam, Olga Kouchnarenko.

Tree automata with constraints are widely used to tackle data base algorithmic problems, particularly to analyse queries over XML documents. The model of Tree Automata with Global Constraints (TAGED) has been introduced for these purposes. The membership problem for TAGED is known to be NP-complete. The emptiness problem for TAGED is known to be decidable and the best known algorithm in the general case is non elementary. Following our NP-hardness result [74], we are still working in collaboration with Vincent Hugot on the complexity of the emptiness problem.

7.3.2. Random Generation of Finite Automata

Participant: Pierre-Cyrille Héam.

Developing new algorithms and heuristics raises crucial evaluation issues, as improved worst-case complexity upper-bounds do not always transcribe into clear practical gains. A classical way for software performance evaluation is to randomly generate inputs.

In collaboration with Jean-Luc Joly, we investigate the problem of randomly and uniformly generating deterministic pushdown automata [40]. Based on a recursive counting approach, we propose a polynomial time algorithm for this purpose. The influence of the accepting condition on the generated automata is also experimentally studied.

Partially ordered automata are finite automata where simple loops have length one. We have used a Markov chain based approach [75] to randomly - and uniformly - generate deterministic partially ordered automata.

In [39] we address the problem of the uniform random generation of non deterministic automata (NFA) up to isomorphism. We show how to use a Monte-Carlo approach to uniformly sample a NFA. The main result is to show how to use the Metropolis-Hastings Algorithm to uniformly generate NFAs up to isomorphism. Using labeling techniques, we show that in practice it is possible to move into the modified Markov Chain efficiently, allowing the random generation of NFAs up to isomorphism with dozens of states. This general approach is also applied to several interesting subclasses of NFAs (up to isomorphism), such as NFAs having a unique initial states and a bounded output degree. Finally, we prove that for these interesting subclasses of NFAs, moving into the Metropolis Markov chain can be done in polynomial time.

7.3.3. Verification of Linear Temporal Patterns over Finite and Infinite Traces

Participants: Pierre-Cyrille Héam, Olga Kouchnarenko.

In the regular model-checking framework, reachability analysis can be guided by temporal logic properties, for instance to achieve the counter example guided abstraction refinement (CEGAR) objectives. A way to perform this analysis is to translate a temporal logic formula expressed on maximal rewriting words into a "rewrite proposition" – a propositional formula whose atoms are language comparisons, and then to generate semi-decision procedures based on (approximations of) the rewrite proposition. In collaboration with Vincent Hugot, we have investigated suitable semantics for LTL on maximal rewriting words and their influence on the feasibility of a translation. We have expended the work in [76] by providing a general translation scheme giving exact results for a fragment of LTL corresponding mainly to safety formulæ, and approximations for a larger fragment.

7.3.4. Constraint Solving for Verifying Modal Workflow Specifications

Participants: Hadrien Bride, Olga Kouchnarenko.

Workflow Petri nets are well suited for modelling and analysing discrete event systems exhibiting behaviours such as concurrency, conflict, and causal dependency between events. They represent finite or infinite-state processes, and several important verification problems, like reachability or soundness, are known to be decidable. Modal specifications introduced in [77] allow loose or partial specifications in a framework based on process algebras.

Our work in [26] aims at verifying modal specifications of coloured workflows with data assigned to the tokens and modified by transitions. To this end, executions of coloured workflow nets are modelled using constraint systems, and constraint solving is used to verify modal specifications specifying necessary or admissible behaviours. An implementation supporting the proposed approach and promising experimental results on an issue tracking system constitute a practical contribution.

7.4. Model-based Testing

Our research in Model-Based Testing (MBT) aims to extend the coverage of tests. The coverage refers to several artefacts: model, test scenario/property, and code of the program under test. The test generation uses various underlying techniques such as symbolic animation of models [71], or symbolic execution of programs by means of dedicated constraints, SMT solvers, or model-checkers.

7.4.1. Automated Test Generation from Behavioral Models

Participants: Fabrice Bouquet, Frédéric Dadeau, Elizabeta Fourneret, Jean-Marie Gauthier, Julien Lorrain, Alexandre Vernotte.

We have developed an original model-based testing approach that takes a behavioral view (modelled in UML) of the system under test and automatically generates test cases and executable test scripts according to model coverage criteria. We continue to extend this result to SysML specifications for validating embedded systems [35]. We apply this method on smartSurface [34]

We have investigated the use of a model-based testing approach for vulnerability testing in web applications. Our research prototype was able to detect vulnerabilities on already deployed web applications [80].

7.4.2. Scenario-Based Verification and Validation

Participants: Fabrice Bouquet, Frédéric Dadeau, Elizabeta Fourneret.

Test scenarios represent an abstract test case specification that aims at guiding the model animation in order to produce relevant test cases. Contrary to the previous section, this technique is not fully automated since it requires the user to design the scenario, in addition to the model.

We have proposed a dedicated formalism to express test properties. A test property is first translated into a finite state automaton which describes a monitor of its behaviors. We have also proposed dedicated property coverage criteria that can be used either to measure the property coverage of a given test suite, or to generate test cases, exercising nominal or robustness aspects of the property. This process has been fully tool-supported into an integrated software prototype[0]. This process has been designed during the ANR TASCCC project (2009-2012) and was continued during the ANR ASTRID OSEP project (2012-2013). The industrialization of this approach, and its integration within commercial test generation tools has started with the ANR ASTRID Maturation MBT_Sec project (2014-2015) in collaboration with the French DoD [46]. A technology transfer is currently in progress to integrate this technology into the Smartesting CertifyIt test generation environment.

Also, we have experimented the model approach to validate and to design Multi-Agent systems [51], [52].

7.4.3. Mutation-based Testing of Security Protocols

Participants: Frédéric Dadeau, Pierre-Cyrille Héam, Michaël Rusinowitch.

We have proposed a model-based penetration testing approach for security protocols [14]. This technique relies on the use of mutations of an original protocol, proved to be correct, for injecting realistic errors that may occur during the protocol implementation (e.g., re-use of existing keys, partial checking of received messages, incorrect formatting of sent messages, use of exponential/xor encryption, etc.). Mutations that lead to security flaws are used to build test cases, which are defined as a sequence of messages representing the behavior of the intruder. We have applied our technique on protocols designed in HLPSL, and implemented the protocol mutation tool jMuHLPSL that performs the mutations. The mutants are then analyzed by *CL-AtSe*.

7.4.4. Code and Contract-based Test Generation and Static Analysis

Participants: Fabrice Bouquet, Frédéric Dadeau, Alain Giorgetti.

With the CEA we have developed a test generation technique based on C code and formal specifications, to facilitate deductive verification, in a new tool named StaDy [43]. The tool integrates the concolic test generator PathCrawler within the static analysis platform Frama-C. StaDy is able to handle the ANSI C Specification Language (ACSL) of the framework and other Frama-C plug-ins are able to reuse results from the test generator. This tool is designed to be the foundation stone of modular static and dynamic analysis combinations in the Frama-C platform.

For bounded exhaustive unitary testing of functions on structured arrays we have designed and formally verified with Frama-C a library of sequential generators [43], [36]. A structured array is an array satisfying given constraints, such as being sorted or having no duplicate values. A sequential generator of structured arrays can be defined by two C functions: the first one computes an initial array, and the second one steps from one array to the next one according to some total order on the set of arrays. We formally specify with ACSL annotations that the generated arrays satisfy the prescribed structural constraints (soundness property) and that the generation is in increasing lexicographic order (progress property). We refine this specification into two programming and specification patterns: one for generation in lexicographic order and one for generation by filtering the output of another generator. After adding suitable loop invariants we automatically prove the soundness and progress properties of many generators with the Frama-C platform.

[0] A video of the prototype is available at: http://vimeo.com/53210102

7.5. Verification of Collaborative Systems

We investigate security problems occurring in decentralized systems. We develop general techniques to enforce read and update policies for controlling access to XML documents based on recursive DTDs (Document Type Definition). Moreover, we provide a necessary and sufficient condition for undoing safely replicated objects in order to enforce access control policies in an optimistic way. We investigate privacy issues for social networks in order to give more control to users over their personal data.

7.5.1. Automatic Analysis of Web Services Security

Participants: Walid Belkhir, Michaël Rusinowitch, Mathieu Turuani, Laurent Vigneron.

Automatic composition of web services is a challenging task. Many works have considered simplified automata models that abstract away from the structure of messages exchanged by the services. For the domain of secured services (using e.g., digital signing or timestamping) we have proposed an original approach to automated orchestration of services under security constraints. Given a community of services and a goal service, we reduce the problem of generating a mediator between a client and a service community to a security problem where an intruder should intercept and redirect messages from the service community and a client service till reaching a satisfying state.

In [12] we develop an alternative approach based on *parametrized automata*, a natural extension of finite-state automata over infinite alphabet. In this model the transitions are labeled with constants or variables that can be refreshed in some specified states. We show the applicability of our model to Web services handling data from an infinite domain. We reduce the Web service composition problem to the construction of a simulation of a target service by the asynchronous product of existing services, and prove that this construction is computable. We also show expressive equivalence and succinctness of parametrized automata with respect to Finite Memory Automata in [47] We now work on synthesizing composed services that satisfy required security policies.

7.5.2. Querying Security Views over XML Data

Participant: Abdessamad Imine.

To enforce access control over XML data, virtual security views are commonly used in many many applications and commercial database systems. Querying these views raises some serious problems. More precisely, user XPath queries posed on recursive views cannot be rewritten to be evaluated on the underlying XML data. Existing rewriting solutions are based on the non-standard language "Regular XPath" enabling recursion operator. However, query rewriting under Regular XPath can be of exponential size. In [17], we show that query rewriting is always possible for arbitrary security views (recursive or not) by using only the expressive power of the standard XPath. We propose a more expressive language to specify XML access control policies as well as an efficient algorithm to enforce such policies. Finally, we present our system, called SVMAX, that implements our solutions and we show that it scales well through an extensive experimental study based on real-life DTD.

7.5.3. Secure Computation in Social Networks

Participants: Younes Abid, Bao Thien Hoang, Abdessamad Imine, Huu Hiep Nguyen, Michaël Rusinowitch.

Online social networks are increasingly exploited as real platforms for creating social links and sharing data. They are used from organizing public opinion polls about any societal theme to publish social graph data for achieving in-depth studies. To securely perform these large-scale computations, we need the design of reliable protocols to ensure the data privacy. In [44], [9], we address the polling problem in social networks where users want to preserve the confidentiality of their votes, obtain the correct final result, and hide, if any, their misbehaviors. We present EPol, a simple decentralized polling protocol that is deployed on a family of social graphs that satisfy a property based on topological ordering. Using these graphs, we show that their structures enable low communication cost, ensure vote privacy and limit the impact of dishonest users on the accuracy of the polling output.

The problem of private publication of social graphs has attracted a lot of attention recently. In [50], we tackle the problem about the upper bounds of privacy budgets related to differentially private release of graphs. We provide such a bound and we prove that, with a privacy budget of $O(\log n)$, there exists an algorithm capable of releasing a noisy output graph with edge edit distance of $O(1)$ against the true graph. At the same time, the complexity of our algorithm $Top - m$ Filter is linear in the number of edges m. This lifts the limits of the state-of-the-art, which incur a complexity of $O(n^2)$ where n is the number of nodes and runnable only on small graphs.

Anonymous use of Social network do not prevent users from privacy risks resulting from infering and cross-checking information published by themselves or their relationhips. In [57], we have conducted a survey in order to measure sensitiveness of personal data published on social media and to analyze the users behaviors. We have shown that 76% of internet users that have answered the survey are vulnerable to identity or sensitive data disclosure.

7.5.4. Safe Protocols for Collaborative Applications

Participant: Abdessamad Imine.

The Operational Transformation (OT) approach, used in many collaborative editors, allows a group of users to concurrently update replicas of a shared object and exchange their updates in any order. The basic idea is to transform any received update operation before its execution on a replica of the object. Designing transformation functions for achieving convergence of object replicas is a critical and challenging issue. In this work, we investigate the existence of transformation functions [19]. From the theoretical point of view, two properties, named TP1 and TP2, are necessary and sufficient to ensure convergence. Using controller synthesis technique, we show that there are some transformation functions, which satisfy TP1 for the basic signatures of insert and delete operations. But, there is no transformation function, which satisfies both TP1 and TP2. Consequently, a transformation function which satisfies both TP1 and TP2 must necessarily have additional parameters in the signatures of some update operations. Accordingly, we provide a new transformation function and show formally that it ensures convergence.

COAST Project-Team

5. New Results

5.1. Probabilistic Partial Orderings

Participants: Jordi Martori Adrian, Pascal Urso.

Ordering events in a distributed system fundamentally consists in delaying event delivery. Partial ordering, such as FIFO and causal order, has many usage in practical distributed and collaborative systems and can be obtained in arbitrarily large and dynamic networks. However, partial orderings imply that messages cannot be sent and delivered as soon as produced.

In [14], we study the latency induced by such partial orderings. We obtain a probabilistic measure of the moment a message can be delivered according the different characteristics of the distributed system. Having such a measure helps to understand the systems behaviour and to design new protocols. For instance, our measure allows us to parametrize a naive, albeit efficient, fault-tolerant causal delivery mechanism. We experimentally validate our approach using Internet-scale production distribution latency including faults.

5.2. Effect of Delay on Group Performance

Participants: François Charoy, Claudia-Lavinia Ignat [contact], Gérald Oster.

We continued our work on studying the effect of delay in real-time collaborative editing. Delays exist between the execution of one user's modification and the visibility of this modification to the other users. Such delays are in part fundamental to the network, as well as arising from the consistency maintenance algorithms and underlying architecture of collaborative editors. Existing quantitative research on collaborative document editing does not examine either concern for delay or the efficacy of compensatory strategies.

In [12] we studied a collaborative note taking task where we introduced simulated delay. The study was done with 20 groups of 4 users which were asked to listen to a short interview and take notes. We found out a general effect of delay on performance related to the ability to manage redundancy and errors across the document. We interpret this finding as a compromised ability to maintain awareness of team member activity, and a reversion to independent work. Measures of common ground in accompanying chat indicate that groups with less experienced team members attempt to compensate for the effect of delay. In contrast, more experienced groups do not adjust their communication in response to delay, and their performance remains sensitive to the delay manipulation. Results of this study support our team assertion that delay associated with conventional consistency maintenance algorithms will impede group performance. Therefore, these results promote the use of novel algorithms such as CRDTs and motivate the pursuance of research and development on these approaches.

5.3. A CRDT Supporting Selective Undo for Collaborative Text Editing

Participants: Luc André, Claudia-Lavinia Ignat [contact].

Selective undo is an important feature in collaborative editors. With selective undo, a user can undo an earlier operation, regardless of when and where the operation was generated. Current systems that support selective undo are subject to two main limitations. Firstly, they only support undo of operations on atomic objects (e.g. characters or un- breakable lines). In the case of string-wise operations such as copy-paste, find-replace or select-delete, users can typically only undo earlier operations character by charac- ter. Secondly, selective undo may lead to undesirable effects. For example, a user first inserts a misspelled word and then makes a correction. The correction depends on the first insertion of the word. It is undesirable to undo the insertion alone and leave the correction behind as a groundless modification. In [15] we proposed a novel consistency maintenance approach relying on a layered commutative replicated data type (CRDT) that supports selective undo of string-wise operations in collaborative editing. This is the first work that manages undesirable effects of undo. Our performance study shows that it provides sufficient responsiveness to the end users.

5.4. A Trust-Based Formal Delegation Framework for Enterprise Social Networks

Participants: Ahmed Bouchami, Olivier Perrin [contact].

Collaborative environments raise major challenges to secure them. These challenges increase when it comes to the domain of Enterprise Social Networks (ESNs) as ESNs aim to incorporate the social technologies in an organization setup while asserting greater control of information security. In this context, the security challenges have taken a new shape as an ESN may not be limited to the boundaries of a single organization and users from different organizations can collaborate in a common federated environment.

We address the problem of the authorization's delegation in federated collaborative environments like ESNs with an approach based on event-calculus, a temporal logic programming formalism. While the traditional approaches are either user-centric or organization-centric, the approach bridges the gap between these two views and the proposed framework enhances the delegation scheme. We have proposed a behavior monitoring mechanism, that permits to assess principals' trust level within the federated collaborative environment [10].

5.5. Risk Management in the Cloud. Application to Business Process Deployment

Participants: Claude Godart [contact], Elio Goettelmann.

The lack of trust in cloud organizations is often seen as braking forces to SaaS developments. This work proposes an approach which supports a trust model and a business process model in order to allow the orchestration of trusted business process components in the cloud.

The contribution is threefold and consists in a method, a model and a framework. The method categorizes techniques to transform an existing business process into a risk-aware process model that takes into account security risks related to cloud environments. These techniques are partially described in the form of constraints to automatically support process transformation. The model formalizes the relations and the responsibilities between the different actors of the cloud. This allows to identify the different information required to assess and quantify security risks in cloud environments.

The framework is a comprehensive approach that decomposes a business process into fragments that can automatically be deployed on multiple clouds. The framework also integrates a selection algorithm that combines the security information of cloud offers and of the process with other quality of service criteria to generate an optimized configuration. It is implemented in a tool to assess cloud providers.

Elio Goettelmann has defended his PhD thesis entitled "Risk-aware Business Process Modeling and Trusted Deployment in the Cloud" on October 2015 [1] based on this result. This framework has been combined to an access control model for strengthening access controls in the context of a collaborative federation of components [9].

5.6. Secure Business Process Deployment in SaaS Contexts

Participants: Amina Ahmed Nacer, Claude Godart [contact], Elio Goettelmann, Samir Youcef.

Business process (BP) stakeholders want the benefits of the cloud, but they are also reluctant to expose their BP models which express the know-how of their companies. To prevent such a know-how exposure, we are developing a design-time approach for obfuscating a BP model by splitting its model into a collaboration of BP fragments semantically equivalent to the initial BP. This breaking down renders the discovery of the deep content of a critical fragment or of the whole process semantics, by cloud providers much harder when these fragments are deployed in a multi-cloud context. While existing contributions on this topic remain at the level of principles, we propose an algorithm supporting such a BP model transformation [11]. To validate this approach, we are developing a new metric of obfuscation. Complementary to obfuscation, we are developing techniques to reuse, at design time, business process fragments from the cloud, but with limited security risks [8].

5.7. Web Services Selection with QoS

Participants: Amina Ahmed Nacer, Kahina Bessai, Claude Godart [contact], Samir Youcef.

The development of the web technologies and the increase of available services raise the issue of the selection of the most appropriate service among a set of candidate web services. First of all, the services offering a given functionality are discovered. Then, the service section process assists users in choosing the services that better meets their preferences. These preferences are generally expressed as potentially objective functions often conflicting.

Most of existing works trying to select the best web services are based either on a single evaluation criterion or, at best, on the use of an aggregation function like weighted sum of several quantitative evaluation criteria, or the use of the Pareto optimality notion.

In this work, we address some shortcomings of existing approaches by introducing a new optimality notion based on two tests: (i) concordance and (ii) discordance tests. It presents an efficient algorithm to select only the best services using the introduced optimality notion. Moreover, the proposed algorithm exhibits encouraging results as supported by a series of experiments [7].

LARSEN Team

7. New Results

7.1. Lifelong Autonomy

7.1.1. Adaptation / Learning

Participant: Jean-Baptiste Mouret.

We collaborate on this subject with Jeff Clune (University of Wyoming, USA).

7.1.1.1. Adaptation to Unforeseen Damage Conditions

Whereas animals can quickly adapt to injuries, current robots cannot "think outside the box" to find a compensatory behaviour when they are damaged: they are limited to their pre-specified self-sensing abilities and can diagnose only anticipated failure modes, an impracticality for complex robots. A promising approach to reducing robot fragility involves having robots learn appropriate behaviours in response to damage, but current techniques are slow even with small, constrained search spaces. We introduced an intelligent trial-and-error algorithm that allows robots to adapt to damage in less than two minutes in large search spaces without requiring self-diagnosis or pre-specified contingency plans [11]. Before the robot is deployed, it uses a novel technique (based on evolutionary algorithms) to create a detailed map of the space of high-performing behaviours. This map represents the robot's prior knowledge about what behaviours it can perform and their value. When the robot is damaged, it uses this prior knowledge to guide a trial-and-error learning algorithm (based on Bayesian optimization) that conducts intelligent experiments to rapidly discover a behaviour that compensates for the damage. Experiments reveal successful adaptations for a legged robot injured in five different ways, including damaged, broken, and missing legs, and for a robotic arm with joints broken in 14 different ways. This new algorithm will enable more robust, effective, autonomous robots, and may shed light on the principles that animals use to adapt to injury.

This work was the cover of Nature on the 28th of May, 2015 (see the "highlights" section).

7.1.2. Robotics Perception

Participants: François Charpillet, Francis Colas, Abdallah Dib, Van Quan Nguyen.

We collaborate on this subject with Emmanuel Vincent from the Multispeech team (Inria Nancy - Grand Est).

7.1.2.1. Audio Source Localization

We considered, here, the task of audio source localization using a microphone array on a mobile robot. Active localization algorithms have been proposed in the literature that can estimate the 3D position of a source by fusing the measurements taken for different poses of the robot. However, the robot movements are typically fixed or they obey heuristic strategies, such as turning the head and moving towards the source, which may be suboptimal. This work proposes an approach to control the robot movements so as to locate the source as quickly as possible [17]. We represent the belief about the source position by a discrete grid and we introduce a dynamic programming algorithm to find the optimal robot motion minimizing the entropy of the grid. We report initial results in a real environment.

This work is carried on through the PhD Thesis of Van Quan Nguyen under the supervision of Emmanuel Vincent and Francis Colas.

7.1.2.2. State Estimation for Autonomous Surface Vessels

Autonomous Surface Vessels (ASVs) are increasingly proposed as tools to automatize environmental data collection, bathymetric mapping and shoreline monitoring. For many applications it can be assumed that the boat operates on a 2D plane. However, with the involvement of exteroceptive sensors like cameras or laser rangefinders, knowing the 3D pose of the boat becomes critical. We formulated three different algorithms based on 3D extended Kalman filter (EKF) state estimation for ASVs localization [12]. We compared them using field testing results with ground truth measurements, and demonstrated that the best performance is achieved with a model-based solution in combination with a complementary filter for attitude estimation. Furthermore, we presented a parameter identification methodology and showed that it also yielded accurate results when used with inexpensive sensors. Finally, we presented a long-term series (i.e., over a full year) of shoreline monitoring data sets and discussed the need for map maintenance routines based on a variant of the Iterative Closest Point (ICP) algorithm.

7.1.2.3. Geometric Registration

We proposed a review of geometric registration in robotics [16]. Registration algorithms associate sets of data into a common coordinate system. They have been used extensively in object reconstruction, inspection, medical application, and localization of mobile robotics. We focus on mobile robotics applications in which point clouds are to be registered. While the underlying principle of those algorithms is simple, many variations have been proposed for many different applications. In this work, we gave a historical perspective of the registration problem and showed that the plethora of solutions can be organized and differentiated according to a few elements. Accordingly, we presented a formalization of geometric registration and cast algorithms proposed in the literature into this framework. Finally, we reviewed a few applications of this framework in mobile robotics that cover different kinds of platforms, environments, and tasks. These examples allowed us to study the specific requirements of each use case and the necessary configuration choices leading to the registration implementation. Ultimately, the objective of this work is to provide guidelines for the choice of geometric registration configuration.

7.1.2.4. Robust Dense Visual Odometry for RGB-D Cameras in a Dynamic Environment

Visual odometry is a fundamental challenge in robotics and computer vision. The aim of our work is to estimate RGB-D camera motion (onboard a mobile robot) from RGB-D images in a dynamic scene with people moving in the scene. Most of the existing methods have a poor localization performance in such case, which makes them inapplicable in real world conditions. This year, we have proposed a new dense visual odometry method [27] that uses random sampling consensus (RANSAC) to cope with dynamic scenes. We show the efficiency and robustness of the proposed method on a large set of experiments in challenging situations and from publicly available benchmark datasets. Additionally, we compare our approach to another state-of-art method based on M-estimator that is used to deal with dynamic scenes. Our method gives similar results on benchmark sequences and better results on our own dataset.

7.1.3. Distributed Sensing and Acting

Participants: Mihai Andries, Amine Boumaza, François Charpillet, Iñaki Fernández Pérez, Nassim Kaldé.

We collaborate on this subject with Olivier Simonin from the Chroma team (Inria Grenoble - Rhône Alpes).

7.1.3.1. Localisation of Humans, Objects and Robots Interacting on Load-Sensing Floors

The use of floor sensors in ambient intelligence contexts began in the late 1990's. We designed such a sensing floor in Nancy in collaboration with Hikob company (http://www.hikob.com/) and Inria SED (*service d'expérimentation et de développement*). This is a load-sensing floor which is composed of square tiles, each equipped with two ARM processors (Cortex m3 and a8), 4 load cells, and a wired connection to the four neighboring cells. Ninety tiles cover the floor of our intelligent apartment experimental platform. This load-sensing floor includes as well a LED lighting system which sits flush with the floor surface. This provides people with a new way to interact with their environment at home. This year, we have focused on localisation, tracking and recognition of humans, objects and robots interacting on load-sensing floors [9]. Inspired by computer vision, the proposed technique processes the floor pressure-image by segmenting

the blobs containing objects, tracking them, and recognizing their contents through a mix of inference and combinatorial search. The result lists the probabilities of assignments of known objects to observed blobs. The concept was successfully evaluated in daily life activity scenarii, involving multi-object tracking and recognition on low resolution sensors, crossing of user trajectories, and weight ambiguity.

7.1.3.2. Online Distributed Learning for a Swarm of Robots

We propose a novel innovation marking method [22] for neuro-evolution of augmenting topologies in embodied evolutionary robotics. This method does not rely on a centralized clock, which makes it well suited for the decentralized nature of embodied evolution where no central evolutionary process governs the adaptation of a team of robots exchanging messages locally. This method is inspired from event dating algorithms, based on logical clocks, that are used in distributed systems, where clock synchronization is not possible. We compare our method to odNEAT, an algorithm in which agents use local time clocks as innovation numbers, on two multi-robot learning tasks: navigation and item collection. Our experiments showed that the proposed method performs as well as odNEAT, with the added benefit that it does not rely on synchronization of clocks and is not affected by time drifts.

The effect of selection pressure on evolution in centralized evolutionary algorithms (EA's) is relatively well understood. Selection pressure pushes evolution toward better performing individuals. However, distributed EA's in an Evolutionary Robotics (ER) context differ in that the population is distributed across the agents, and a global vision of all the individuals is not available. In this work, we analyze the influence of selection pressure in such a distributed context. We propose a version of mEDEA [22] that adds a selection pressure, and evaluate its effect on two multi-robot tasks: navigation and obstacle avoidance, and collective foraging. Experiments show that even small intensities of selection pressure lead to good performances, and that performance increases with selection pressure. This is opposed to the lower selection pressure that is usually preferred in centralized approaches to avoid stagnating in local optima.

7.1.3.3. Online Distributed Exploration of an Unknown Environment by a Swarm of Robots

This year, we have proposed a new taboo-list approach [18] for multi-robot exploration of unknown structured environments, in which robots are implicitly guided in their navigation on a globally shared map. Robots have a local view of their environment, inside which they navigate in a asynchronous manner. When the exploration is complete, robots gather at a rendezvous point. The novelty consists in using a distributed exploration algorithm which is not guided by frontiers to perform this task. Using the Brick and Mortar Improved ant-algorithm as a base, we add robot-perspective vision, variable vision range, and an optimization which prevents agents from going to the rendezvous point before exploration is complete. The algorithm was evaluated in simulation on a set of standard maps.

Another work [14] carried out within the PhD of Nassim Kaldé concerns exploration in populated environments. The difficulty here is that pedestrian flows can severely impact performances. However, humans have adaptive skills for taking advantage of these flows while moving. Therefore, in order to exploit these human abilities, we propose a novel exploration strategy that explicitly allows for human-robot interactions. Our model for exploration in populated environments combines the classical frontier-based strategy with our interactive approach. We implement interactions where robots can locally choose a human guide to follow and define a parametric heuristic to balance interaction and frontier assignments. Finally, we evaluate to which extent human presence impacts our exploration model in terms of coverage ratio, travelled distance and elapsed time to completion.

7.2. Natural Interaction with Robotics Systems

7.2.1. Human Characterization

Participants: François Charpillet, Abdallah Dib, Xuan Son Nguyen, Vincent Thomas.

We collaborate on this subject with Olivier Buffet and Alain Dutech from Inria Nancy - Grand Est, Arsène Fansi Tchango and Fabien Flacher from Thales ThereSIS, and Alain Filbois from SED Inria Nancy - Grand Est.

7.2.1.1. Multi-Camera Tracking in Partially Observable Environment

In collaboration with Thales ThereSIS - SE&SIM Team (Synthetic Environment & Simulation), we focus on the problem of following the trajectories of several persons with the help of several controllable cameras. This is a difficult problem since the set of cameras cannot simultaneously cover the whole environment, since some persons can be hidden by obstacles or by other persons, and since the behavior of each person is governed by internal variables which can only be inferred (such as his motivation or his hunger).

The approach we are working on is based on (1) the HMM (Hidden Markov Models) formalism to represent the state of the system (the persons and their internal states), (2) a simulator provided and developed by Thales ThereSIS, and (3) particle filtering approaches based on this simulator. Since activity and location depend on each other, we adopt a Simultaneous Tracking and Activity Recognition approach.

After having shown that it was possible to use a complex behavioral simulator to infer the behavior of complex individuals (motivation, possession, ...) even in case of long periods of occlusions [40], we investigated how to propose a factored particle filter (with one distribution per target) for efficiently tracking multiple targets simultaneously. To that end, we use a Joint Probabilistic Data Association Filter with a particular model of dynamics that largely decouples the evolution of several targets, and turns out to be very natural to apply. We proposed to use a small number of "representatives" of each target to determine and consider only effective interactions among targets.

This work has been published in Arsène Fansi Tchango's PhD thesis which has been defended in December [7].

7.2.1.2. Human Posture Recognition

Human pose estimation in realistic world conditions raises multiple challenges such as foreground extraction, background update and occlusion by scene objects. Most of existing approaches were demonstrated in controlled environments. In this work, we propose a framework to improve the performance of existing tracking methods to cope with these problems. To this end, a robust and scalable framework is provided composed of three main stages. In the first one, a probabilistic occupancy grid updated with a Hidden Markov Model used to maintain an up-to-date background and to extract moving persons. The second stage uses component labelling to identify and track persons in the scene. The last stage uses a hierarchical particle filter to estimate the body pose for each moving person. Occlusions are handled by querying the occupancy grid to identify hidden body parts so that they can be discarded from the pose estimation process. We provide a parallel implementation that runs on CPU and GPU at 4 frames per second. We also validate the approach on our own dataset that consists of synchronized motion capture with a single RGB-D camera data of a person performing actions in challenging situations with severe occlusions generated by scene objects. We make this dataset available online (http://www0.cs.ucl.ac.uk/staff/M.Firman/RGBDdatasets/).

7.2.2. Social Robotics

Participants: Amine Boumaza, Serena Ivaldi.

We collaborate on this subject with Yann Boniface from Loria, Alain Dutech from Inria Nancy - Grand Est and Nicolas Rougier from the Mnemosyne team (Inria Bordeaux - Sud-Ouest).

7.2.2.1. PsyPhINe: Cogito Ergo Es

PsyPhINe is an interdisciplinary and exploratory project (see 9.1.2) between philosophers, psychologists and computer scientists. The goal of the project is related to cognition and behavior. Cognition is a set of processes that are difficult to unite in a general definition. The project aims to explore the idea of assignments of intelligence or intentionality, assuming that our intersubjectivity and our natural tendency to anthropomorphize play a central role: we project onto others parts of our own cognition. To test these hypotheses, our aim is to design a "non-verbal" Turing Test, which satisfies the definitions of our various fields (psychology, philosophy, neuroscience and computer science), using a robotic prototype. Some of the questions that we aim to answer are: is it possible to give the illusion of cognition and/or intelligence through such a technical device? How elaborate must be the control algorithms or "behaviors" of such a device so as to fool test subjects? How many degrees of freedom must it have?

Preliminary experiments with human subjects conducted this past year on a simple device helped to design an experimental protocol and test simple hypotheses which set the ground for the full fledged non verbal Turing Test. This project was funded under a PEPS Mirabelle grant (see 9.1.2) which helped build a robotic device with many degrees of freedom to perform further experiments. We also organized an inter-disciplinary workshop gathering top researchers from philosophy, anthropology, psychology and computer science to discuss and exchange on our methodology (see 10.1.1.1).

7.2.2.2. Multimodal Object Learning During Human-Robot Interaction

Robots working in evolving human environments need the ability to continuously learn to recognize new objects. Ideally, they should act as humans do, by observing their environment and interacting with objects, without specific supervision. However, if object recognition simply relies on visual input, then it may fail during human-robot interaction, because of the superposition of human and body parts. A multimodal approach was then proposed in [15], where visual input from cameras was combined with the robot proprioceptive information, in order to classify objects, robot, and human body parts. We proposed a developmental learning approach that enables a robot to progressively learn appearances of objects in a social environment: first only through observation, then through active object manipulation. We focused on incremental, continuous, and unsupervised learning that does not require prior knowledge about the environment or the robot. In the first phase of the proposed method, we analyse the visual space and detect proto-objects as units of attention that are learned and recognized as possible physical entities. The appearance of each entity is represented as a multi-view model based on complementary visual features. In the second phase, entities are classified into three categories: parts of the body of the robot, parts of a human partner, and manipulable objects. The categorization approach is based on mutual information between the visual and proprioceptive data, and on motion behaviour of entities. The ability to categorize entities is then used during interactive object exploration to improve the previously acquired objects models. The proposed system was implemented and evaluated with an iCub and a Meka robot learning 20 objects. The system was able to recognize objects with 88.5% success rate and create coherent representation models that are further improved by learning during human-robot interaction.

7.2.2.3. Robot Functional and Social Acceptance

To investigate the functional and social acceptance of a humanoid robot, we carried out an experimental study with 56 adult participants and the iCub robot. Trust in the robot has been considered as a main indicator of acceptance in decision-making tasks characterized by perceptual uncertainty (e.g., evaluating the weight of two objects) and socio-cognitive uncertainty (e.g., evaluating which is the most suitable item in a specific context), and measured by the participants' conformation to the iCub's answers to specific questions. In particular, we were interested in understanding whether specific (i) user-related features (i.e., desire for control), (ii) robot-related features (i.e., attitude towards social influence of robots), and (iii) context-related features (i.e., collaborative vs. competitive scenario), may influence their trust towards the iCub robot. We found that participants conformed more to the iCub's answers when their decisions were about functional issues than when they were about social issues. Moreover, the few participants conforming to the iCub's answers for social issues also conformed less for functional issues. Trust in the robot's functional savvy does not thus seem to be a pre-requisite for trust in its social savvy. Finally, desire for control, attitude towards social influence of robots and type of interaction scenario did not influence the trust in iCub. Results are also discussed with relation to methodology of HRI research in a currently submitted paper (http://arxiv.org/abs/1510.03678 [cs.RO]). This work follows the research on engagement with social robots that was previously published [10].

7.2.2.4. Relation Between Extroversion and Negative Attitude Towards Robot

Estimating the engagement is critical for human - robot interaction. Engagement measures typically rely on the dynamics of the social signals exchanged by the partners, especially speech and gaze. However, the dynamics of these signals is likely to be influenced by individual and social factors, such as personality traits, as it is well documented that they critically influence how two humans interact with each other. We assess the influence of two factors, namely extroversion and negative attitude toward robots, on speech and gaze during a cooperative task, where a human must physically manipulate a robot to assemble an object [23]. We evaluate if the score of extroversion and negative attitude towards robots co-variate with the duration and frequency of gaze and

speech cues. The experiments were carried out with the humanoid robot iCub and 56 adult participants. We found that the more people are extrovert, the more and longer they tend to talk with the robot; and the more people have a negative attitude towards robots, the less they will look at the robot face and the more they will look at the robot hands where the assembly and the contacts occur. Our results confirm and provide evidence that the engagement models classically used in human-robot interaction should take into account attitudes and personality traits.

MADYNES Project-Team

7. New Results

7.1. Monitoring

7.1.1. Anonymous networks monitoring

Participants: Thibault Cholez [contact], Isabelle Chrisment, Olivier Festor.

In 2015, we pursued our collaboration with Juan Pablo Timpanaro a former team's PhD student and published a new paper [47] on the I2P anonymous network (http://i2p2.de). More precisely, we monitored I2P's decentralised directory, known as the netDB, and produced two contributions. On the one hand, we conducted arguably the first *churn* study of the I2P network, showing that I2P users are more stable than non-anonymous peer-to-peer users. On the other hand, we analysed the design of the netDB and compared it against the popular KAD design, demonstrating that the former is more vulnerable to different attacks, specially to Eclipse attacks, which can be mitigated by applying some safer design choices of the latter. We lately showed the positive impact on performance of including KAD's DHT configuration into the netDB in terms of bandwidth, storage and messages overhead.

7.1.2. Smartphone usage monitoring

Participants: Vassili Rivron [contact], Mohammad Irfan Khan, Simon Charneau [Inria], Isabelle Chrisment.

In [39] we presented some results from our study based on a combination of crowdsending and survey. We discussed some technical problems we faced and some lessons learned during our crowdsensing experiment. Furthermore we showed how information regarding social context can be used for better interpretation of crowdsensed data. Next we selected some questions from the multiple choice survey questionnaire and combined the responses with crowdsensed data to analyze users' perception about their smartphone usage and discussed cognitive factors associated with reporting information on questionnaires. Moreover we showed that combining sensing with survey can improve both the techniques and the combination has important use cases such as helping users to have a better understanding and control of their technology usage.

7.1.3. Active Monitoring

Participants: Abdelkader Lahmadi [contact], Jérôme François, Valentin Giannini, Frederic Beck [LHS], Bertrand Wallrich [LHS].

The main motivation of this work was to assess the exposition of industrial systems in the Internet, especially by measuring how many SCADA systems are accessible. To do so, we built an IPv4 methodology which is able to scan the entire IPv4 address space by maximizing the distance between consecutive IP addresses. It thus avoids colateral effect of overloading targeted networks and being blacklisted. We thus extend the Zmap tool (zmap.io) by also including other functionalities such as distributed scans, indexation and visualisation of the results [63]. First experiences have been performed and are under evaluation.

7.1.4. Sensor networks monitoring

Participants: Rémi Badonnel, Isabelle Chrisment, Olivier Festor, Abdelkader Lahmadi [contact], Anthea Mayzaud.

This year, our work on security-oriented monitoring has been centered on building a distributed architecture that supports passive monitoring in the Internet of Things using the RPL protocol [37]. A particular interest has been given to advanced metering infrastructure (AMI) networks, where higher order devices are expected to form the backbone infrastructure, to which more constrained nodes would connect. Our distributed architecture exploits the capabilities of these higher order devices to perform network monitoring tasks, and takes benefits from properties inherent to that protocol, such as DODAG building and multi-instance routing mechanisms, in order to passively monitor the environment with a minimal impact on constrained nodes.

We have also consolidated our taxonomy on security attacks in these networks [8]. In addition, we have pursued our work on topological inconsistency attacks [9]. It is evident from the experiments that we have conducted that mitigating such attacks is critical to avoid channel congestion and high resource usage. Our initial adaptive threshold (AT) strategy to mitigate the effects of such attacks has been further improved. The new strategy dynamically takes into account network characteristics in order to infer an appropriate threshold for counteracting these attacks.

7.2. Security

7.2.1. Security analytics

Participants: Jérôme François [contact], Abdelkader Lahmadi, Manobala Nirmala, Vincent Noyalet.

During the year 2015, we have extended our monitoring platform dedicated to Android environments [69] with more analytics features. The monitoring platform is dedicated to the collection, storage, analysis and visualization of logs and network flow data of mobile applications. The platform relies on a set of on-device probes to monitor network and system activities of these applications. The data are collected from these probes and parsed through generic and flexible collectors relying on Flume agents that we have adapted and extended. We are storing the collected data using a column oriented Hbase storage engine (Hadoop database). Finally, after being parsed, the data are made available within the Elasticsearch engine to search and visualize them using the Kibana tool. We have also presented the building blocks of the platform in a lab session within the conference AIMS 2015 [70].

We have also maintained an IETF draft [75] to promote a standardization effort towards the extension of IP Flow-based monitoring with geographic information. Associating Flow information with their measurement geographic locations will enable security applications to detect anomalous activities. In the case of mobile devices, the characterization of communication patterns using only time and volume is not enough to detect unusual location-related communication patterns.

Besides, we looked at aggregating flows collected at the High Security Lab since a single attack is represented by multiple flows. For example, a DDoS or a scan is a sequence of similar parallel flows coming from the same or distributed machines. As attacks occur very frequently and even at the same time, grouping flows occurring in a pre-defined time window is not a valid approach. Two approaches have been investigated and are actually dependent of the sources of collected flows. First, we analyzed collected Netflow data from the Darknet which is basically a sinkhole without any services running or announced. Hence, all traffic is considered as abnormal and is limited to a set of predefined attacks. Indeed, since no packets can be sent back, complex attacks with different steps cannot be caught. Therefore, scanning, flooding-based denial-of-service and backscatter are the main types of anomalies we can observe. Flows are thus grouped and labeled regarding certain criteria (common IP addresses/subnets, ports, co-occurrence) thanks to a pre-established decision process [58]. The final goal was to compare data collected in Nancy and in Tokyo. Secondly, we assume flow data without specific knowledge about the type of traffic it embeds. In such a case, the goal is to automatically extract recurrent patterns. The initial approach consisted in representing flows as nodes in a graph and linking them when sharing some properties (IP addresses, ports). Major subsequent problems have been faced like indexation, split flows in multiple files and visualization [59].

7.2.2. Management of HTTPS traffic

Participants: Thibault Cholez [contact], Shbair Wazen, Jérôme François, Isabelle Chrisment.

We previously investigated the latest technique for HTTPS traffic filtering that is based on the Server Name Indication (SNI) field of TLS and which has been recently implemented in many firewall solutions. We showed that SNI has two weaknesses, regarding (1) backward compatibility and (2) multiple services using a single certificate and we implemented a proof of concept of these vulnerabilities as a web browser extension (Escape). This work was published in the IFIP/IEEE IM'15 conference [44].

This led us to the development of new reliable methods to investigate the increasing number of HTTPS traffic that may hold security breaches but without relying on decryption at any step, in order to respect users' privacy (no HTTPS proxy). Many approaches already identify the main type of an application (Web, P2P, SSH,..) running in secure tunnels, and others identify a couple of specific encrypted web pages through website fingerprinting.

In this context, we developed a better technique to precisely identify the services run within HTTPS connections, i.e. to name the services, without relying on specific header fields that can be easily altered. We have defined dedicated features for HTTPS traffic that are used as input for a multi-level identification framework based on machine learning algorithms. Our evaluation based on real traffic shows that we can identify encrypted web services with a high accuracy. This work will be published next year in the IFIP/IEEE Network Operations and Management Symposium (NOMS 2016).

7.2.3. Configuration security automation

Participants: Rémi Badonnel [contact], Gaetan Hurel, Abdelkader Lahmadi, Olivier Festor.

Our work during year 2015 was mainly focused on the orchestration of security functions in the context of mobile smart environments [35]. Most of current security approaches for these environments are provided in the form of applications or packages to be directly installed on the devices themselves. Such approaches may be qualified as on-device. However, on-device approaches generally induce significant local resource consumption leading to the significant reduction of battery lifetime. In the meantime, current cloud-based approaches for mobile security attempt to deal with this issue by offloading most of the workload on a remote server, but may introduce significant additional latency. In that context, we have pursued the efforts on our strategy for dynamically outsourcing and composing security functions in the cloud, considering software-defined networking. The architecture relies on a set of security functions that are activated, configured and orchestrated according to the current contexts and risks, while a dedicated modelling has been introduced for supporting the evaluation of security compositions and their properties. The chaining of security functions is performed dynamically in order to fit with the security requirements of mobile devices at runtime. In particular, we have proposed in [35] to analyze and cluster applications running on the mobile devices based on their network behaviors, in order to drive the selection and deployment of adequate security compositions that may be fully outsourced or split between in-cloud and on-device.

We have also investigated in [23] to what extent security automation, more specifically in the context of vulnerability management, might be supported by conceptual knowledge discovery. The intended extension might be a mean to cope with the increasing dynamics and complexity of networked environments. Most current security solutions still seem to work under certain boundaries that prevent them to act intelligently and flexibly, i.e. strictly sticked to the available security information in order to analyze, report and eventually remediate found problems. Our purpose is to exploit methods and techniques coming from formal concept and knowledge discovery in databases, in order to provide high-level automation based on mechanisms capable of understanding, reasoning about, and anticipating the surrounding environment and its vulnerabilities.

7.3. Experimentation, Emulation, Reproducible Research

This section covers our work on experimentation on testbeds (mainly Grid'5000), on emulation (mainly on Distem), and on Reproducible Research.

7.3.1. Grid'5000 design and evolutions

Participants: Jérémie Gaidamour, Arthur Garnier, Lucas Nussbaum [contact], Clément Parisot.

The team was again heavily involved in the evolutions and the governance of the Grid'5000 testbed.

In the context of ADT LAPLACE, Jérémie Gaidamour adapted and configured the CiGri middleware on Grid'5000. CiGri enables the execution of large campaigns of *best-effort* jobs (low priority, interruptible jobs). It is expected that this work will allow the remaining free time slots to be filled by tasks from other research communities such as natural language processing.

Jérémie Gaidamour also greatly improved *stats5k*, our tool to generate metrics about the testbed (usage, resources availability, etc.), available at https://intranet.grid5000.fr/stats/.

Arthur Garnier added the testing of Grid'5000 tutorials to our continuous integration installation, enabling the earlier detection of problems on the testbed. He then led the migration to PostgreSQL as the backend for the OAR batch scheduler – a behind-the-scenes but major migration.

In addition to daily administratrive duties and to his work on Kwapi described below (section 7.3.2), Clément Parisot added support for *production* workloads to Grid'5000, extending the scope of the testbed to make it more suitable for additional user communities. He then managed the installation of the new clusters at Nancy, purchased in the context of OIP Grid'5000 and CPER CyberEntreprises.

Finally, in addition to his roles in the *bureau*, *comité d'architectes* and *comité des responsables de sites* of Grid'5000, Lucas Nussbaum managed the purchase of the new clusters at Nancy mentioned above, and gave several presentations about the testbed, at *Journées SUCCES* [14], at *Retour d'expéRiences sur la Recherche Reproductible* [15], and at *École Cumulo Numbio*.

7.3.2. A unified monitoring framework for energy consumption and network traffic

Participants: Lucas Nussbaum [contact], Clément Parisot.

Providing experimenters with deep insight about the effects of their experiments is a central feature of testbeds, that Grid'5000 was only partially addressing. We designed Kwapi, a framework that unifies measurements for both energy consumption and network traffic. Because all measurements are taken at the infrastructure level (using sensors in power and network equipment), using this framework has no dependencies on the experiments themselves. Initially designed for OpenStack infrastructures, the Kwapi framework allows monitoring and reporting of energy consumption of distributed platforms. In this work, we extended Kwapi to network monitoring, and overcame several challenges: scaling to a testbed as large as Grid'5000 while still providing high-frequency measurements; providing long-term loss-less storage of measurements; handling operational issues when deploying such a tool on a real infrastructure.

This work was published at Tridentcom [31] and presented in a GENI/FIRE collaboration workshop [12]. It is now in production as the default monitoring framework on Grid'5000.

7.3.3. Comparison of HPC and Clouds testbeds

Participant: Lucas Nussbaum [contact].

Given the recent launch of two large NSF-funded projects that share similar goals as Grid'5000 (CloudLab and ChameleonCloud), we worked on analyzing the design choices made so far by those projects, comparing them with Grid'5000. Preliminary results were presented at REPPAR [17] and at a GENI/FIRE collaboration workshop [13].

7.3.4. Emulation with Distem

Participants: Emmanuel Jeanvoine, Lucas Nussbaum [contact], Cristian Ruiz.

Several improvements have been made around Distem, mostly in the context of ADT COSETTE.

During the internship of Arthur Carcano, we tried to use Distem to experiment on NDN infrastructures. We obtained promising results, especially in terms of scale. We plan to continue this work and publish it in 2016.

We also submitted, to CCGRID, a paper demonstrating the use of Distem to evaluate fault tolerance and load balancing strategies implemented in Charm++. This submission is still pending evaluation.

Finally, in an effort to validate Distem performance, we studied the performance of Container-based virtualization technologies such as LXC or Docker, as most of the underlying technology is also shared with Distem. We studied their performance in the context of HPC, and showed that containers technology has matured over the years, and that performance issues are being solved. This work has been published at VHPC [43].

7.3.5. Management of large-scale experiments

Participants: Emmanuel Jeanvoine, Lucas Nussbaum [contact], Cristian Ruiz.

Following our survey of experiment management tools [7] accepted at FGCS at the end of 2014 and published early this year, we worked on Ruby-Cute, a library that aggregates various useful functionality in the context of such tools. We hope that it will be useful as a basis for future tools, and ease testing of new ideas in that field. The library is available on http://ruby-cute.github.io/.

7.3.6. Tracking provenance in experiment control tools

Participants: Tomasz Buchert, Lucas Nussbaum [contact].

In the context of our work on XPFlow, we worked on the collection of provenance during experiments. We surveyed provenance collection in various domains of computer science, introduced a new classification of provenance types suited to distributed systems experiments, and proposed a design of a provenance system inspired by this classification. This work has been published at REPPAR [29].

7.3.7. Reproducible Research

Participant: Lucas Nussbaum [contact].

Lucas Nussbaum gave a presentation on Reproducible Research[16] at the ICube laboratory seminar (Strasbourg). A shorter version of the talk was given to the Inria *Comité des projets* in Nancy.

Lucas Nussbaum also co-organized the second edition of REPPAR, a workshop on Reproducibility in Parallel Computing, held in conjunction with Euro-Par'2015.

7.4. Routing

7.4.1. Routing in Wireless Sensor Networks

Participants: Emmanuel Nataf [contact], Patrick-Olivier Kamgueu, Nesrine Khelifi.

We have formalized our previous work on the routing protocol for wireless sensor network by fuzzy logic specifications. The rules of routing metric composition are now valid for any network depth and we demonstrated its quality by real experimentation [36]. This work is done in the context of the associated team we build with the Cameroun and the Inria international lab LIRIMA.

For potentially very large wireless sensor network, our routing or any other routing, can not limit traffic bottleneck near the network root. Network depth should also be reduced as hop by hop communication is a factor which strongly increases data loss rate. Considering theses problems Nesrine Khelifi PhD student of the Manouba University in Tunisia spent 3 months within the Madynes team trying to limit the depth of the network by splitting it under the supervision of network quality observers we had to define.

7.4.2. Operator calculus based routing in Wireless Sensor Networks

Participants: Evangelia Tsiontsiou, René Schott, Stacey Staples [Southern Illinois University Edwardsville], Jamilla Benslimane, Bilel Nefzi, Ye-Qiong Song [contact].

Recently, Operator calculus (OC) has been developed by Schott and Staples with whom we collaborate. We make use of OC methods on graphs to solve path selection in the presence of multiple constraints. Based on OC, we developed a distributed algorithm for path selection in a graph. This approach has been applied to efficiently solve a joint routing, channel and time slot scheduling optimization problem in UWB wireless sensor networks [6]. We also designed a new routing protocol which makes use of this algorithm: the Operator Calculus based Routing Protocol (OCRP). In OCRP, a node selects the set of eligible next hops based on the given constraints and the distance to the destination. It then sends the packet to all eligible next hops. The protocol is implemented in Contiki OS (Rime profile) and emulated for TelosB motes using Cooja. We compared its performance against tree and directional flooding routing and showed the advantages of our technique [28]. Our ongoing work consists in its comparison with RPL to show its practical contribution to handle simultaneously several IETF ROLL routing metrics. This work is part of Lorraine AME Satelor project granted by Lorraine Region.

7.4.3. Probabilistic Energy-Aware Routing for Wireless Sensor Networks

Participants: Evangelia Tsiontsiou, Bernardetta Addis, Alberto Ceselli [Universita degli Studi di Milano], Ye-Qiong Song [contact].

Healthcare applications are considered as promising fields for Wireless Sensor Networks (WSNs). Thanks to WSNs, patients can be monitored in hospitals or smart home environments, providing health improvement, or emergency care. A key issue is the limited battery of sensors; indeed, current WSN research trends for healthcare applications include energy efficient routing and network lifetime guarantee mechanisms, among others. One of our ongoing work consists in designing a Smart Probabilistic Energy-Aware Routing Protocol (SPEAR) for WSNs which aims at maximizing the network lifetime by keeping low energy consumption and balancing network traffic between nodes. Our experimental campaign reveals that our SPEAR protocol outperforms the popular Energy Aware Routing Protocol (EAR) from the literature, proving to be more effective in extending the network lifetime. This work has resulted in a conference submission. It is part of Lorraine AME Satelor project granted by Lorraine Region.

7.4.4. Energy-aware IP networks management

Participants: Bernardetta Addis [contact], Giuliana Carello [DEIB, Politecnico di Milano, Italy], Antonio Capone [DEIB, Politecnico di Milano, Italy], Luca Gianoli [Polytechnique de Montreal, Canada], Sara Mattia [IASI, CNR, Roma, Italy], Brunide Sansò [Polytechnique de Montreal, Canada].

The focus of our research is to minimize the energy consumption of the network through a management strategy that selectively switches off devices according to the traffic level. We consider a set of traffic scenarios and jointly optimize their energy consumption assuming a per-flow routing. We propose a traffic engineering mathematical programming formulation based on integer linear programming that includes constraints on the changes of the device states and routing paths to limit the impact on quality of service and the signaling overhead.

A very important issue that may be affected by green networking techniques is resilience to node and link failures. We thus extended the optimization models to guarantee network survivability. Results show that significant savings, up to 30%, may be achieved even when both survivability and robustness are fully guaranteed.

Computational cost of proposed models can be very high when dealing with large size instances (network size and/or number of demands). For this reason, we proposed and tested different problem formulations with the aim of solving larger size instances at optimality. We focus on a particular form of shared protection mechanism, where energy consumption is associated only to active devices during normal functioning. We propose a standard and a projected formulation, with additions of cuts and valid inequalities. Computational results show that the projected formulation is very effective [20]. We plan to extend the work to consider multiperiod scenarios.

7.4.5. Virtual Network Functions Placement and Routing Optimization

Participants: Bernardetta Addis [contact], Dallal Belabed [LIP6, Univ Paris 06, France], Mathieu Bouet [Thales Communications & Security, France], Stefano Secci [LIP6, Univ Paris 06, France].

Network Functions Virtualization (NFV) is incrementally deployed by Internet Service Providers (ISPs) in their carrier networks, by means of Virtual Network Function (VNF) chains, to address customers' demands. The motivation is the increasing manageability, reliability and performance of NFV systems, the gains in energy and space granted by virtualization, at a cost that becomes competitive with respect to legacy physical network function nodes. From a network optimization perspective, the routing of VNF chains across a carrier network implies key novelties making the VNF chain routing problem unique with respect to the state of the art: the bitrate of each demand flow can change along a VNF chain, the VNF processing latency and computing load can be a function of the demands traffic, VNFs can be shared among demands, etc. We started our work providing an integer linear programming model for Virtual Network Functions Placement and demand rerouting. By extensive simulation on realistic ISP topologies, we draw conclusions on the trade-offs achievable between legacy Traffic Engineering (TE) ISP goals and novel combined TE-NFV goals [19].

7.4.6. Composing IoT protocols with Named-Data Networking

Participants: Salvatore Signorello [University of Luxembourg], Olivier Festor [contact], Radu State [University of Luxembourg].

With the emergence of IoT, many layer 2 protocols have been proposed with each of them its own characteristics, advantages and drawbacks. Choosing a protocol often depends on the global context, as for example number of users, time of the day... Although devices can now be fitted with multiple interfaces, using always the same specific layer 2 protocol is not efficient, in particular if we assume that connected devices are retrieving or exchanging similar contents. For example, assuming that WiFi is the most usable interface to download some files in Internet through an access point may not be ideal if a close-by device accessible by Bluetooth already has it. To accommodate so multiple layer 2 protocols, we propose to leverage the Named-Data Networking (NDN) paradigm which allows to explore in parallel multiple paths for retrieving content independently of the underlying protocol. Our first results [46] show that such a theoretical solution cannot work practically. Indeed, applying NDN in a blind mode over multiple layer 2 protocols does not assume the corresponding specificities like for example various collision rates depending on the underlying protocols, which have to be taken into account.

7.5. Multi-modeling and co-simulation

Participants: Laurent Ciarletta [contact], Olivier Festor, Ye-Qiong Song, Yannick Presse, Victorien Elvinger, Julien Vaubourg, Alexandre Tan, Benjamin Segault, Emmanuel Nataf.

Vincent Chevrier (former Maia team, Dep 5, LORIA) is a collaborator and the correspondant for the MS4SG project, Benjamin Camus, and Christine Bourjot (former MAIA team, Dep 5, LORIA) are collaborators for AA4MM/MECSYCO. Julien Vaubourg's PhD is under the co-direction of V. Chevrier and L. Ciarletta.

In Pervasive or Ubiquitous Computing, a growing number of communicating/computing devices are collaborating to provide users with enhanced and ubiquitous services in a seamless way.

These systems, embedded in the fabric of our daily lives, are complex: numerous interconnected and heterogeneous entities are exhibiting a global behavior impossible to forecast by merely observing individual properties. Firstly, users physical interactions and behaviors have to be considered. They are influenced and influence the environment. Secondly, the potential multiplicity and heterogeneity of devices, services, communication protocols, and the constant mobility and reorganization also need to be addressed. Our research on this field is going towards both closing the loop between humans and systems, physical and computing systems, and taming the complexity, using multi-modeling (to combine the best of each domain specific model) and co-simulation (to design, develop and evaluate) as part of a global conceptual and practical toolbox.

We proposed the AA4MM meta-model [76] that solves the core challenges of multimodeling and simulation coupling in an homogeneous perspective. In AA4MM, we chose a multi-agent point of view: a multi-model is a society of models; each model corresponds to an agent and coupling relationships correspond to interaction between agents. In the MS4SG (Multi Simulation for Smart Grids) projet which involves some members of the former MAIA team, Madynes and EDF R&D on smart-grid simulation, we developed a proof of concepts for a smart-appartment case that serves as a basis for building up use cases.

In 2015 we worked on the following research topics:

- Assessment and evaluation of complex systems.
- Cyber Physical Systems

  We have led the design and implementation of the Aetournos platform at Loria. The collective movements of a flock of flying communicating robots / UAVs, evolving in potentially perturbed environment constitute a good example of a Cyber Physical System. Applying co-simulation technique we plan to develop a hybrid "network-aware flocking behavior" / "behavior aware routing protocol".

We have provided a working set of tools: multi-simulation behavior / network / physics and generic software development using ROS (Robot Operating System). The UAVs carry a set of sensors for location awareness, their own computing capabilities and several wireless networks.

The effort put in the UAVs gathers academic and research ressources from the Aetournos platform, the R2D2 ADT and the 6PO project, while applied, industrial and more R&D projects have been pursued this year (Outback Joe Search and Rescue Challenge, Alerion, Hydradrone) .

- MS4SG to link multi-simulations tools such as HLA (High Level Architecture) and FMI (Functional Mockup Interface) thanks to our AA4MM framework. We have so far successfully applied our solution to the simulation of smart apartment complex and to combine the electrical and networking part of a Smart Grid. The AA4MM software has seen major improvements in 2015 thanks to the ressources provided by the MS4SG project and a Carnot engineer financed thanks to Inria. It has been renamed as MECSYCO (http://www.mecsyco.com).

  Starting from domain specific and heterogenous models and simulators, the MECSYCO suite allows for multi *systems* integration at several levels: conceptual, formal and software. A couple of visualization tools have been developed as proof of concepts both at run-time and post-mortem.

  We have developed software components and plugins that interconnects within MECSYCO heterogeneous simulators from different domains: FMU (working with the 1.0 and 2.0 FMI standard for CoSimulation) ou non-FMU such as NS3 or Omnet++.

  Several EDF oriented use cases have demonstrated multi-simulations.

In additon to technical reports, several publications have been accepted in 2015 on these subjects [51], [49] and [48].

7.6. Pervasive or Ubiquious Computing

Participants: Laurent Ciarletta [contact], Olivier Festor, Ye-Qiong Song, Emmanuel Nataf, Thomas Paris, Quentin Houbre, Benjamin Segault, Jonathan Arnault, Eric Perlinski, Antoine Richard.

In Pervasive or Ubiquitous Computing, a growing number of communicating/computing devices are collaborating to provide users with enhanced and ubiquitous services in a seamless way.

These systems, incresingly numerous and heterogeneous, are embedded in the fabric of our daily lives. Our initial subject of interest is to study them with regards to their complexity: Those numerous interconnected and heterogeneous entities are exhibiting a global behavior impossible to forecast by merely observing individual properties.

Firstly, users physical interactions and behaviors have to be considered. They are influenced and influence their surroundings and the environment. Secondly, the potential multiplicity and heterogeneity of devices, services, communication protocols, and the constant mobility and reorganization also need to be addressed.

Our research on this field is going towards both closing the loop between humans and systems, physical and computing systems, and taming the complexity, using multi-modeling (to combine the best of each domain specific model) and co-simulation (to design, develop and evaluate) as part of a global conceptual and practical toolbox.

During some exploratory work, we have seen the potential of these Pervasive Computing ressources in the (Very Serious) Gaming area.

In 2015 we worked on the following topics:

- Cyber Physical Systems

  We pursuied the design and implementation of the Aetournos platform at Loria. The collective movements of a flock of flying communicating robots / UAVs, evolving in potentially perturbed environment constitute a good example of a Cyber Physical System. Eventually, we applied co-simulation technique and plan to develop a hybrid "network-aware flocking behavior" / "behavior aware routing protocol".

We developed a working set of tools: multi-simulation behavior / network / physics and generic software development using ROS (Robot Operating System). The UAVs carry a set of sensor for location awareness, their own computing capabilities and several wireless networks.

The effort put in the UAVs gathers academic and research ressources from the Aetournos platform, the Inria ADT R2D2 and the 6PO project, while applied, industrial and more R&D projects have been pursued this year (Medical Express / Outback Joe Search and Rescue Challenge, Alerion, Hydradrone, and a CIFRE PhD with Thales) .

- Smart * (MS4SG)

  We have studied scientific problems around model and simulator composition. We have also looked into practical and implementation issues in the frame of our MECSYCO /AA4MM solutions. We have added to our Smart Grid scenarios a smart appartment complex use case.

- (Very Serious) Gaming: Starburst Gaming

7.7. Quality-of-Service

7.7.1. Self-adaptive MAC protocol for both QoS and energy efficiency

Participants: Kévin Roussel, Shuguo Zhuo, Olivier Zendra, Ye-Qiong Song [contact].

Three main contributions have been made this year. Firstly iQueue-MAC has been extended to work on both single channel mode and multi-channel mode, improving its throughput performance [11]. Secondly, S-CoSenS and iQueue-MAC our previously designed protocols have been implemented on RIOT OS over MSP430-based motes. Our contribution consists in developing a port of RIOT OS on the MSP430 micro-controller and demonstrating that RIOT OS offers rich and advanced real-time features, especially the simultaneous use of as many hardware timers as the underlying platform (micro-controller), which are fundamental features to implement high performance MAC protocols [41]. The Cooja/MSPSim network simulation framework is widely used for developing and debugging, but also for performance evaluation of WSN projects. Our third contribution shows that Cooja is not limited only to the simulation of the Contiki OS based systems and networks, but can also be extended to perform simulation experiments of other OS based platforms, especially that with RIOT OS. Moreover, when performing our own simulations with Cooja and MSPSim, we observed timing inconsistencies with identical experimentations made on actual hardware. Such inaccuracies clearly impair the use of the Cooja/MSPSim framework as a performance evaluation tool, at least for time-related performance parameters. The detailed results of our investigations on the inaccuracy problems, as well as the consequences of this issue, and possible ways to fix or avoid it are available in [42]. Part of this work has been supported by PIA LAR project.

7.7.2. End-to-end delay modeling and evaluation in wireless sensor networks

Participants: François Despaux, Abdelkader Lahmadi, Ye-Qiong Song [contact].

Probabilistic end-to-end performance guarantee may be required when dealing with real-time applications. As part of ANR QUASIMODO project, we are dealing with Markov modeling of multi-hop networks running duty-cycled MAC protocols. One of the problems of the existing Markovian models resides in their strong assumptions that may not be directly used to assess the end-to-end delay in practice. In particular, realistic radio channel, capture effect and OS-related implementation factors are not taken into account. We proposed to explore a new approach combining code instrumentation and Markov chain analysis. In [32] we propose a novel approach to obtain the Markov chain model of sensor nodes by means of Process Mining techniques through the analysis of MAC protocol execution traces for a given traffic scenario. End to end delay is then computed based on this Markov chain. Experimentations were done using IoT-LAB testbed platform. Comparisons in terms of delay have been presented for two different metrics of the RPL protocol (hop count and ETX). The overall approach and its generalization using non-linear regression techniques in terms of traffic rate are detailed in the PhD thesis of François Despaux defended in September 2015 [1].

7.7.3. Dynamic resource allocation in network virtualization

Participants: Mohamed Said Seddiki, Mounir Frikha [SupCom, Tunis, Tunisie], Ye-Qiong Song [contact].

This work has been carried out as part of a co-supervised PhD thesis between University of Lorraine and SupCom Tunis.

The objective of this research topic is to develop different resource allocation mechanisms in Network Virtualization, for increasing the QoS guarantee. Firstly, we demonstrated the potential of SDN in the QoS management of a virtualized home network (VN). We proposed and implemented "FlowQoS", a mechanism that can be deployed by an Internet Service Provider in the last-mile hop or in the home gateway. Performance measurements show that this solution can share bandwidth between applications according to user-defined configuration to guarantee QoS for each active traffic. The second contribution is the modeling of the interaction between service providers and infrastructure providers using game theory framework to offer dynamic sharing of physical infrastructure across multiple VN with different QoS requirements. We presented a set of non-cooperative games to model the negotiation phase and the dynamic allocation of nodes and physical links for each deployed VN[10]. Finally we proposed a predictive approach that allows an adaptive control of bandwidth allocation in order to reduce the packet delays for a given VN on each physical link. The last two contributions offer dynamic sharing models of physical infrastructure resources while guaranteeing the QoS for each VN.

The overall approach is detailed in the PhD thesis of Said Seddiki defended in April 2015 [2].

7.7.4. QoS and fault-tolerance in distributed real-time systems

Participants: Florian Greff, Laurent Ciarletta, Arnauld Samama [Thales TRT], Eric Dujardin [Thales TRT], Ye-Qiong Song [contact].

The QoS must be guaranteed when dealing with real-time distributed systems interconnected by a network. Not only task schedulability in processors, but also message schedulability in networks should be analysed for validating the system design. Fault-tolerance is another critical issue that one must take into account. In collaboration with Thales TRT industrial partner as part of a CIFRE PhD work, we started a study on the real-time dependability of distributed multi-criticity systems interconnected by an embedded mesh network (RapidIO). For easing the QoS specification at the higher level, DDS middleware is used. We postulate that enhancing QoS for real-time applications entails the development of a cross-layer support of high-level requirements, thus requiring a deep knowledge of the underlying networks. This year, we proposed and implemented a new simulation/emulation/experimentation framework called ERICA, for designing such a feature. ERICA integrates both a network simulator (Ptolemy) and an actual hardware network to allow implementation and evaluation of different QoS-guaranteeing mechanisms. It also supports real-software-in-the-loop, i.e. running of real applications and middleware over these networks. Each component can evolve separately or together in a symbiotic manner, also making teamwork more flexible [68], [33].

7.7.5. Wireless sensor and actuator networks

Participants: Lei Mo, Xiufang Shi [Zhejiang University], Jiming Chen [Zhejiang University], Ye-Qiong Song [contact].

Wireless sensor and actuator networks provide a key technology for fully interacting within a CPS (Cyber-Physical System). However, the introduction of the mobile actuator nodes in a network rises some new challenging issues. In this context, we addressed two important issues: the multiple target tracking using both fixed and mobile sensors and the optimal scheduling of mobile wireless energy chargers (actuators) for fixed sensor nodes.

In our work, the data association problem in multiple target tracking is investigated. To reduce the computational complexity of traditional Joint Probabilistic Data Association (JPDA) algorithm, a modified JPDA algorithm is proposed to execute data association in multiple target tracking by utilizing the information of occlusion conditions, which is identified by a three-step algorithm. Simulation results show that the proposed algorithm has good tracking performance but low computational complexity [45].

We also investigated the multiple mobile chargers coordination problem that is minimizing the energy expenditure of the mobile chargers while guaranteeing the perpetual operation of the wireless sensor network. We formulated this problem as a mixed-integer linear program (MILP). To solve this problem efficiently, we proposed a novel decentralized method which is based on Benders decomposition. The multiple mobile chargers coordination problem is then decomposed into a master problem (MP) and a slave problem (SP), with the MP for mobile chargers scheduling and the SP for mobile chargers moving and charging time allocation. The convergence of proposed method is analyzed theoretically. Simulation results demonstrated the effectiveness and scalability of the proposed method [38].

7.7.6. Big Data-oriented networking

Participants: Jérôme François [contact], Lautaro Dolberg [University of Luxembourg], Thomas Engel [University of Luxembourg], Raouf Boutaba [University of Waterloo], Reaz Ahmed [University of Waterloo], Shihabur Rahman Chowdhury [University of Waterloo].

Performances of Big Data applications are tightly coupled with the performance of the network in supporting large data transfers. Deploying high-performance networks in data centers is thus vital but configuration and performance management as well as the usage of the network are of paramount importance. We thus surveyed helpful approaches in a book chapter [55]. This chapter starts by discussing the problem of virtual machine placement and its solutions considering the underlying network topology. It then provides an analysis of alternative topologies highlighting their advantages from the perspective of Big Data applications needs. In this context, different routing and flow scheduling algorithms are discussed in terms of their potential for using the network most efficiently. In particular, Software-Defined Networking relying on centralized control and the ability to leverage global knowledge about the network state is propounded as a promising approach for efficient support of Big Data applications.

7.8. Advanced Cache Management in Content-centric Networks

Participants: Thomas Silverston [contact], Cholez Thibault, Bernardini César, Aubry Elian, Chrisment Isabelle, Olivier Festor.

Information Centric Networking (ICN) has become a promising new paradigm for the future Internet architecture. It is based on named data, where content address, content retrieval and the content identification is led by its name instead of its physical location. One of the ICN key concepts relies on in-network caching to store multiple copies of data in the network and serve future requests, which helps reducing the load on servers, congestion in the network and enhances end-users delivery performances. Thus, the efficiency of the CCN architecture depends drastically on performances of caching strategies at each node. To date, there has been a lot of studies proposing new caching strategies to improve the performances of CCN. However, among all these strategies, it is still unclear which one performs better as there is a lack of common environment to compare these strategies. To this end, we compared the performances of CCN caching strategies within the same simulation environment. We build a common evaluation scenario and we compare via simulation five relevant caching strategies: Leave Copy Everywhere (LCE), Leave Copy Down (LCD), ProbCache, Cache “Less” For More and MAGIC. We analyze the performances of all the strategies in terms of Cache Hit, Stretch, Diversity and Complexity, and determine the cache strategy that fits the best with every scenario. This work has been published in IEEE Globecom 2015 [26].

At the meantime, CCN architecture uses *Interest* and *Data* messages to request and receive the data, and there has been no routing scheme to match a request to a specific content, as it is currently the case in the Internet. Indeed, CCN relies on flooding, which is a limitation for a future deployment at the Internet-scale. To this end, we proposed a Routing Scheme for CCN based on the softwarization (SDN). In our scheme SRSC, a controller gets knowledge of the network it administers as well as the content, and each node request the next hop to forward the Interest to their controller, until it reaches the closer Content Stores with the requested content. Nodes use a communication channel with the controller that relies only CCN messages and does not use the traditional SDN communication channel protocol Openflow over IP. The rationale is to help having CCN as a stand-alone new networking stack and to enforce its deployement without the IP infrastructure. This research work has been published in IEEE Netsoft 2015 [22] and Algotel 2015 [21].

MAGRIT Project-Team

6. New Results

6.1. Matching and 3D tracking

Participants: Marie-Odile Berger, Jaime Garcia Guevara, Nazim Haouchine, Gilles Simon, Frédéric Sur.

Pose initialization Automating the camera pose initialization is still a problem in non instrumented environments. Difficulties originate in the possibly large viewpoint changes between the data stored in the model and the current view. In this context, Pierre Rolin's PhD work concerns viewpoint simulation techniques for localization. The idea is to generate keypoint descriptors from simulated views in order to enrich the model and to ease the matching of the current view to the model. We have demonstrated the effectiveness of this technique in several situations, either under an affine or a perspective camera model [17], [21]. The computed pose is more stable when it is difficult to obtain reliable correspondences between the model and the current view. In addition, several examples show that our method successfully computes the camera pose whereas the traditional methods fail. Our recent work concerns a progressive sampling strategy to speed up the search of correspondences when confronted to a large outlier rate, which is inherent to viewpoint simulation. We also currently investigate the localization of the virtual camera from which viewpoints should be simulated.

AR in urban environments
Pose initialization is especially difficult in urban scenes due to the presence of repeated patterns. Another difficulty originates in the fact that a pedestrian is free of his motion in the scene and can therefore adopt uncontrolled viewpoints (close or distant views) with respect to the model. As a result, the set of 2D/3D correspondence hypotheses may contain a high ratio of outliers which may lead to erroneous pose computation. In order to improve the matching / recognition stage, we investigated how facades in calibrated images can be orthorectified and delimited by considering prior information about the scene and the camera relevant to AR in urban context [20]. We provide a Bayesian framework to detect vanishing points in Manhattan worlds, which incorporate priors about the Manhattan frame by imposing a near-vertical direction as well as orthogonality constraints. Second, we propose to detect right-angle corners due to windows or doors using a SVM-based machine learning technique. Rectangular facade hypotheses are then generated through min-cuts techniques with the idea to identify rectangles with high density of right-angle corners. Our algorithm performs better or as well as state-of-the-art techniques and is much faster, mainly as a result of using a suitable prior.

Tracking 3D deformable objets
3D augmentation of deformable objects is a challenging problem with many potential applications in computer graphics, augmented reality and medical imaging. Most existing approaches are dedicated to surface augmentation and are based on the inextensibility constraint, for sheet-like materials, or on the use of a model built from representative samples. However, few of them consider in-depth augmentation which is of utmost importance for medical applications. Since the beginning of N. Haouchine's PhD thesis, we have addressed several important limitations that currently hinder the use of augmented reality in the clinical routine of minimally invasive procedures. In collaboration with the MIMESIS team, our main contribution is the design and the validation of an augmented reality framework based on a mechanical model of the organ and guided by features extracted and tracked on the video at the surface of the organ [12]. Specific models which best suit the considered organs, such as a vascularized model of the liver, have been introduced in this framework. Experiments conducted on ex-vivo data of a porcine liver show that the localization error of a virtual tumor were less than 6mm, and thus below the safety margin required by surgery. To our knowledge, we were the first to produce such evaluation for deformable objects.

This work has been extended to augment highly elastic objects in a monocular context. Shape recovery from a monocular video sequence is an underconstrained problem. State-of-the art solutions enforce smoothness or geometric constraints, consider specific deformation properties such as inextensibility or resort to shading constraints. However, few of them can handle properly large elastic deformations. We have proposed [13] a real-time method that uses a mechanical model and is able to handle highly elastic objects. The problem is formulated as an energy minimization problem accounting for a non-linear elastic model constrained by external image points acquired from a monocular camera. This method prevents us from formulating restrictive assumptions and specific constraint terms in the minimization. In addition, we propose to handle self-occluded regions thanks to the ability of mechanical models to provide appropriate predictions of the shape.

The work conducted during N. Haouchine's PhD thesis allowed us to build a complete framework for the use of AR in liver surgery. We now want to focus on specific points to improve the accuracy and the robustness of the augmented process and to facilitate the clinical use of such AR systems. The PhD thesis of Jaime Garcia Guevara started in October on this topic with the aim to build more realistic mechanical models of organs during the surgery (taking into account liver deformation due to insuflation of air during surgery) and to improve the robustness of visual tracking through the use of multiple visual cues and improved methods for outlier detection.

6.2. Image-based modeling

Participants: Marie-Odile Berger, Charlotte Delmas, Antoine Fond, Erwan Kerrien, Gilles Simon, Pierre-Frédéric Villard, Brigitte Wrobel-Dautcourt.

Finding Manhattan directions in uncalibrated images

Finding orthogonal vanishing points (VPs) in a photography has many potential applications in computer vision and computer graphics, including perspective correction, image-based reconstruction and texture extraction. Surprisingly, while this problem has been extensively studied in the literature, manual solutions are still used in most software. Existing algorithms generally follow two steps. First, lines are grouped into pencils, whose centers are considered as potential VPs. Then, an orthogonality measure is evaluated for every triplet of VPs and the most plausible triplet is used for camera calibration. A drawback of this approach is that complex and time-consuming techniques have to be used to solve the general problem of VP detection, while only three particular VPs are finally used. By contrast, we propose an effective and easy-to-implement algorithm that estimates the zenith and the horizon line before detecting the VPs, using simple properties of the central projection and exploiting accumulations of oriented segments around the horizon. Our method is fast and yields an accuracy comparable, and even better in some cases, to that of state-of-the-art algorithms. This work was submitted to Eurographics 2016.

Tools reconstruction for interventional neuro-radiology

Minimally invasive techniques impact surgery in such ways that, in particular, an imaging modality is required to maintain a visual feedback. Live X-ray imaging, called fluoroscopy, is used in interventional neuroradiology. Such images are very noisy, and cannot show but the vasculature and no other brain tissue. In particular, since at most only two projective fluoroscopic views are available, containing absolutely no depth hint, the 3D shape of the micro-tool (guidewire, micro-catheter or micro-coil) can be very difficult, if not impossible to infer, which may have an impact on the clinical outcome of the procedure.

In collaboration with GE Healthcare, we aim at devising ways to reconstruct the micro-tools in 3D from fluoroscopy images. Charlotte Delmas has been working as a PhD CIFRE student on this subject since April 2013. She first devised a solution in a two-view reconstruction context. A paper reporting on this work was published and an oral presentation was made at SPIE Medical Imaging 2015 [19]. The focus was made this year on reconstructing the first coil, a single wire that tangles into a complex pattern when deployed in an aneurysm. The challenge is to produce a 3D reconstruction with as little X-ray dose and acquisition time as possible. Two paths were simultaneously followed this year: 1) design and compare various configurations to rapidly shoot 6 X-ray views from different viewpoints with a biplane (stereo) system; 2) compensate the lack of information (small number of images) with a prior, incorporated in the tomographic reconstruction algorithm, to express the sparsity in space of the shape to be reconstructed. Preliminary work sets forward a

simultaneous fast rotation of both planes around the patient's head. A database is currently being acquired for a full validation in a view to submitting this work for publication early next year.

Individual-specific heart valve modeling
Mitral valve surgical repair is a complex procedure where the outcome largely depends on the surgeons' experience. Predicting a good coaptation of the two leaflets post-operatively is one of the most difficult sub-tasks in the procedure. We worked on a biomechanical simulation framework [25] that computes the leaflet deformation and coaptation based on individual-specific microtomography data as an initial step toward patient-based mitral valve surgical repair assistance through simulation. Results from FEM analysis on three explanted porcine hearts showed that it is possible to obtain the real shape of the leaflets during systolic peak. We also measured the influence of the positions of chordae tendineae on the simulation and showed that marginal chordae have a greater influence on the final shape than intermediary chordae.

Quasi-periodic noise removal
Images may be affected by quasi-periodic noise. This undesirable feature manifests itself by spurious repetitive patterns covering the whole image, well localized in the Fourier domain. While notch filtering permits to get rid of this phenomenon, this however requires to first detect the resulting Fourier spikes, and, in particular, to discriminate between noise spikes and spectrum patterns caused by spatially localized textures or repetitive structures. Several approaches have been investigated. First, we have reviewed the available methods, most of them requiring expert tuning, with applications to experimental mechanics in view [11]. We have also proposed two automated patch-based approaches. A parametric approach has been investigated in [14] (more information available in [26]), based on the detection of noise spikes as statistical outliers to the distribution expected from natural non-noisy patches, which is known to follow the inverse of a power of the frequency. A non-parametric approach, based on a-contrario detection, was also proposed in [22].

6.3. Parameter estimation

Participants: Frédéric Sur, Erwan Kerrien, Raffaella Trivisonne.

Metrological performance assessment in experimental mechanics
A problem of interest in experimental solid mechanics is to estimate displacement and strain maps on the surface of a specimen subjected to a load or a tensile test. Two contactless approaches are available in the literature, based on depositing on the surface of the specimen either a pseudo-periodic grid or a random speckle. Analyzing images taken before and after deformation permits to estimate strain maps. While periodicity permits to make use of Fourier analysis in the first case, digital image correlation (DIC) is used in the second case. Concerning pseudo-periodic grids, we have investigated noise reduction techniques. While averaging a series of images is certainly the most basic option to reduce the noise, it is impossible to get rid of residual vibrations carried for instance by concrete floor slabs. We have shown in [16] that, while these vibrations indeed blur grid images, they still permit to reduce the noise amplitude in the displacement and strain maps. Concerning DIC-based techniques, we have investigated the effect of sensor noise on the measurement resolution. Since displacement of interest are most of the time below one pixel, interpolation plays a major role. We have proposed a new resolution formula which takes interpolation into account. Besides, this formula has been the subject of an experimental assessment on real data, in the presence of signal-dependent noise [24], [18].

Sensor noise measurement
We have investigated in [15] (additional information available in [27]) the problem of sensor parameter estimation from a series of images, under illumination flickering and vibrations. Illumination flickering is indeed a natural assumption for indoor artificial lights. It is also involved by slight variations in the opening time of a mechanical shutter. We have proposed a model of the pixel intensity based on a Cox process, together with an algorithm which, taking benefit of flickering, gives an estimation of every sensor parameter, namely the gain, the readout noise, and the offset.

Image driven simulation

The IDeaS ANR project targets image-driven simulation, applied to interventional neuroradiology: a coupled system of interactive computer-based simulation (interventional devices in blood vessels) and on-line medical image acquisitions (X-ray fluoroscopy). The main idea is to use the live X-ray images as references to continuously refine the parameters used to simulate the blood vessels and the interventional devices (micro-guide, micro-catheter, coil). Our guideline is to follow a sequential statistical filtering approach to fuse such heterogeneous data.

Christo Gnonnou was hired as an engineer (located at Inria Lille in Defrost team (ex-Shacra), contract started on January 1st and ended on October 31st). He continued the work to specify which parameters the simulation is sensitive to, in a view to design a reduced parametric model of the device, and associate covariances to its parameters. He also worked on inverting the mechanical parameters of any device, using our experimental setup based on a high speed stereo rig.

Maxime Malgras worked on his Master's thesis in the team. His investigations aimed at designing a particle filter where the transition function is approximated by a polynomial chaos (PC) instead of particles. It appeared that PC is adapted to compute the posterior in a particle filter but it is not clear whether the number of samples required to estimate the PC coefficients is smaller than the number of particles required for the filter to be accurate, which questions the capacity of PC to reduce the computational burden of particle filters in high dimensions. Raffaella Trivisonne started her PhD thesis in November (co-supervised by Stéphane Cotin, from MIMESIS team in Strasbourg) to investigate deeper on this subject and apply data assimilation to image driven simulation of endovascular interventions.

MIMESIS Team

6. New Results

6.1. Augmented reality for surgery

We have developed a method for real-time augmented reality of internal liver structures during minimally invasive hepatic surgery. Vessels and tumors computed from pre-operative Computed Tomography Angiograms (CTA) scans can be overlaid onto the laparoscopic view for surgery guidance. Compared to current methods, our method is able to locate the in-depth positions of the tumors based on partial three-dimensional liver tissue motion using a real-time biomechanical model. We are pursuing the development of this augmented reality system by using a better biomechanical model, and by relying on parameter optimization and additional per-operative information to further improve accuracy and robustness. In addition, more experiments, and also clinical studies are being performed to precisely measure the benefits and limitations of our approach. This work is strongly related to our involvement in the IHU Strasbourg and is tightly linked to the SOFA-OR project. Many articles were published on this topic [28], [16], [17].

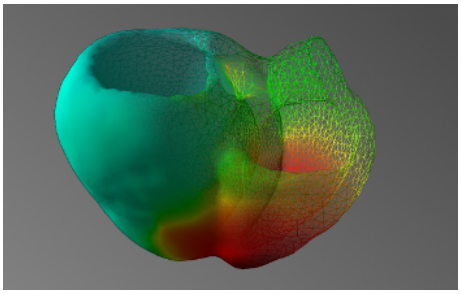

Figure 9. Electrophysiology model of the human heart

6.2. Cardiac electrophysiology

Cardiac arrhythmia is a very frequent pathology that comes from an abnormal electrical activity in the myocardium. This pathology can be treated by catheterization and ablation of the malfunctionning cardiac tissue. The skills required for such interventions are still very challenging to learn, and typically acquired over several years. We first developed a training simulator for interventional electrocardiology and thermo-ablation of these arrhythmias. Based on physical models 9 , this training system reproduces the different steps of the procedure, including endovascular navigation, electrophysiological mapping, pacing and cardiac ablation. Based on a scenario of cardiac arrhythmia, cardiologists assessed the interactivity and the realism of our simulation. This work has been submitted in a journal and is currently under review.

Beyond electrophysiology training, our work around the cardiac electrophysiology also consisted in personalizing our mathematical models. Using the dense electrograms recorded intra-operatively, we presented an accurate and innovative approach to personalize our model, i.e. estimate patient-specific parameters. The modeling in silico of a patient electrophysiology is needed to better understand the mechanism of cardiac arrhythmia.

6.3. Cryoablation

In 2015, we carried on the work around thermal ablation and pre-operative planning based on a thermal Finite Element Model (FEM). The cryoablation technique consists in inserting needles that freeze the surrounding tissues, thus immediately leading to cellular death of the tissues. Cryoablation procedure is used in many medical fields for tumor ablation, and even starts being used in cardiology. In this scope, we built a simulator 10 able to place the cryoprobes and run a simulation representing the evolution of iceballs in living tissues.

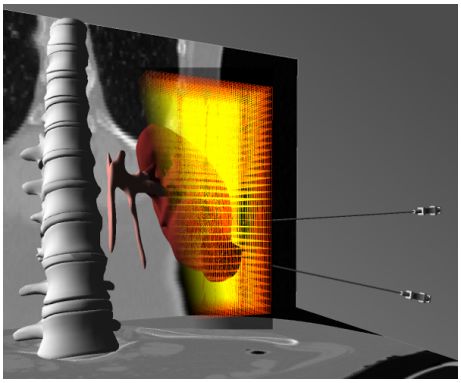

Figure 10. Cryosurgery simulation with the creation of an iceball in the kidney

6.4. Lipofilling reconstructive surgery

We have developed a method to simulate the outcome of reconstructive facial surgery based on fat-filling. Facial anatomy is complex: the fat is constrained between layers of tissues which behave as walls along the face; in addition, connective tissues that are present between these different layers also influence the fat-filling procedure. To simulate the end result, we have proposed a method which couples a 2.5D Eulerian fluid model for the fat and a finite element model for the soft tissues. Both models are coupled using the computation of the mechanical compliance matrix. We had two contributions: a solver for fluids which couples properties of solid tissues and fluid pressure, and an application of this solver to fat-filling surgery procedure simulation. Vincent Majorczyk defended his PhD [14] on this topic in 2015.

6.5. Neurosurgery

Based on an intra-operative registration method, we developed a simulation of a DBS (Deep Brain Stimulation) surgery which can help the surgeon to locate anatomical structures for a safer and a more efficient treatment. The method relies on the biomechanical model of brain shift we developed during the last years. Because some parameters of the model are unknown, we propose to estimate them with an optimization process. The cost function evaluates the distance between the model and the segmentation of pneumocephalus, the only indicator of brain shift visible on an intra-operative CT scan. In 2015, an article about the rest shape of the brain was accepted [19].

6.6. Physics-based registration algorithms

Before targeting the augmented reality for laparoscopic operations, an important step consists in solving the initial alignment problem. Given a pre-operative image of the organ (usually a CT scan) a detailed mesh is constructed. To make the information stored in this mesh available during the operation, the mesh must be registered onto the intra operative view. However, mainly due to the pneumoperitoneum, the organ has

undergone important deformation between the pre-operative images acquisition and the operation. The pre-operative shape and the intra-operative shape of the organ do not correspond. Therefore a non rigid registration is required to align the mesh and the real organ. Our registration algorithms also allowed us to work on means to automatically recover boundary conditions of a patient specific liver.

We created a statistical atlas of the human liver to store the positions of the boundary conditions: the veina cava and the anchor point of the falciform ligament positions. This method was accepted at ABME in 2015 [21]. We also developed a new registration method that evolves automatically from a rigid registration to a non rigid registration to solve the initial alignment problem. The method uses some anatomical features of the liver such as the anchor point position of the falciform ligament.

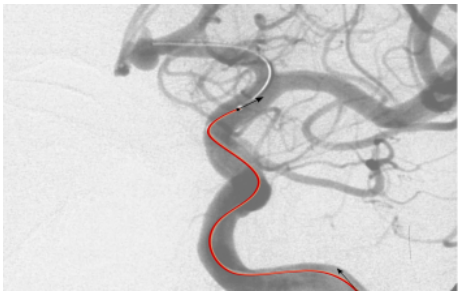

Figure 11. Detecting a catheter in interventional medical images

6.7. Radiation-less guidance during interventional radiology procedures

Significant changes have taken place over the past 20 years in medicine with the development of minimally-invasive procedures. While surgery evolved towards laparoscopy for instance, interventional radiology has become another alternative for many pathologies. Yet, some limitations remain: for percutaneous procedures, soft tissue motion, either due to breathing or deformation induced by the needle, changes the location of the target. When using image guidance, or robotic control, this remains a major obstacle. Regarding catheter-based interventions, the lack of 3D information, and extensive use of X-ray imaging to visualize the path to be followed, are among the main issues. We propose to address these different problems by developing an advanced navigation system which relies on a combination of real-time simulation and information extracted from intra-operative images to assess the current position of the needle. Such a method would have direct applications in pre-operative planning, per-operative guidance, and control for robotics. Our approach will combine advanced modeling of the device, soft tissue deformation, tissue-tool interactions, and planning algorithms 11 .

6.8. Regional anaesthesia

The RASimAs project (Regional Anaesthesia Simulator and Assistant) is a European research project funded by the European Union's 7th Framework Program. It aims at providing a virtual reality simulator and assistant to doctors performing regional anaesthesia by developing the patient-specific Virtual Physiological Human models. In this project, we are in charge of developing a simulation of a needle inserted into a leg using the SOFA framework 12 . We especially focused on the integration of the needle simulation into SOFA. We planned to release the first version of the simulator by January 2016.

In the context of RASimAs, we organized a coding sprint in Strasbourg in April 2015.

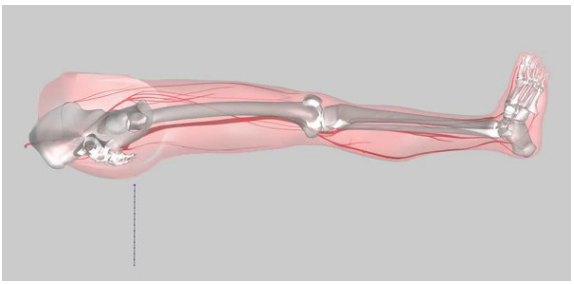

Figure 12. Needle insertion in a muscle in the context of local anaesthesia

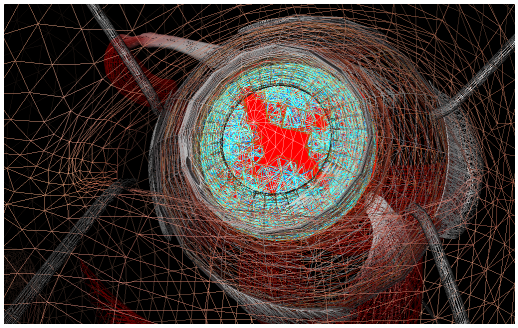

Figure 13. FEM model of the eye used in our simulation of retina surgery

6.9. Training for retina surgery

Retina surgery is an increasingly performed procedure for the treatment of a wide spectrum of retinal pathologies. Yet, as most micro-surgical techniques, it requires long training periods before being mastered. To properly answer requests from clinicians for highly realistic training on one hand, and new requirements for accreditation or recertification from surgical societies on the other hand, we are developing a high-fidelity training system for retinal surgery. This simulator will be built upon our strong scientific expertise in the field of real-time simulation, and a success story for technology transfer in the field of cataract surgery simulation. Members of the consortium have a long expertise in the development of prototypes, as well as collaborations with clinical partners. The simulation system that we propose to develop is based on the Open Source simulation platform SOFA, and relies on expertise from our partners to ensure clinical and industrial relevance. This work is initially funded through the ANR project RESET which started in March 2015. A first version 13 of the training system has been delivered and we made a live demonstration at the Journée Alsacienne d'Ophtalmologie.

6.10. Virtual Cutting

The simulation of cutting is a central interest in the team. We especially work on the simulation of surgical cuts 14 , tearing and other separations of materials induced by surgical tools. On the one hand, we investigated the theoretical aspect: using the standard finite element method (FEM) combined with a re-meshing approach, we replace locally the current structure of the mesh in order to allow for a separation. On the other hand, we detected a separation in the motion of an object provided by a monocular video stream. With that detection, we can provide an augmented reality during the cutting and tearing of a deformable object.

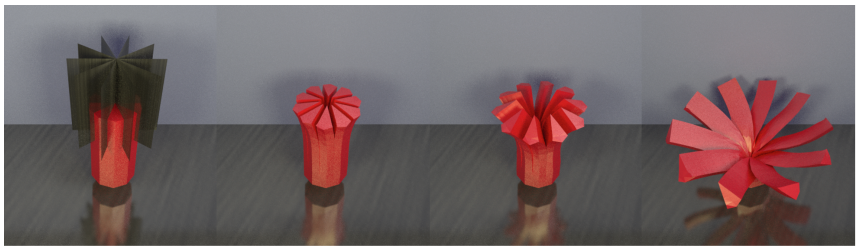

Figure 14. Our cutting algorithm in SOFA

The theoretical aspect of our work has been published in an article both at the conference Computer Graphics International CGI [24] and in the journal "The visual computer" [20]. The application in augmented reality 15 has been published at two conferences: "Augmented Reality during Cutting and Tearing of Deformable Objects", International Symposium on Mixed and Augmented Reality (ISMAR) [30].

To read more about our projects and results, please visit our website: http://mimesis.inria.fr.

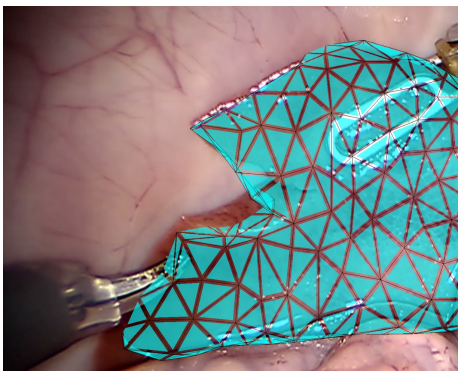

Figure 15. Augmented reality on a liver involving large deformation and cutting, i.e. topological changes

MULTISPEECH Project-Team

7. New Results

7.1. Explicit Modeling of Speech Production and Perception

Participants: Yves Laprie, Slim Ouni, Vincent Colotte, Anne Bonneau, Agnès Piquard-Kipffer, Emmanuel Vincent, Denis Jouvet, Julie Busset, Benjamin Elie, Andrea Bandini, Ilef Ben Farhat, Sara Dahmani, Valérian Girard.

7.1.1. Articulatory modeling

7.1.1.1. Acoustic simulations

The acoustic simulation plays a key role in articulatory synthesis since it generates the acoustic signal from the instantaneous geometry of the vocal tract. This year we extended the single-matrix formulation to enable self-oscillation models of vocal folds, including glottal chinks, to be connected to the vocal tract. It also integrates the case of a local division of the main air path into two lateral channels, as it may occur during the production of lateral approximants. Extensions give rise to a reformulation of the acoustic conditions at the glottis, and at the upstream connection of bilateral channels. Numerical simulations validate the simulation framework. In particular the presence of a zero around 4 kHz due to the presence of bilateral channels around both sides of the tongue for the sound /l/ is confirmed by the simulations. These results agree with those obtained via independent techniques. Simulations of static vowels reveal that the behavior of the vocal folds is qualitatively similar whether they are connected to the single-matrix formulation or to the classic reflection type line analog model.

7.1.1.2. Acquisition of articulatory data

Magnetic resonance imaging (MRI) is a technique which provides very good static images of the vocal tract. However, it cannot be used directly to acquire dynamic images of the vocal tract which would enable a better comprehension of articulatory phenomena and the development of better coarticulation models. We thus have a cooperation with the IADI (Imagerie Adaptatative Diagnostique et Interventionnelle) INSERM laboratory in Nancy Hospital intended to develop cineMRI [86], [87] (see. 6.8).

7.1.1.3. Articulatory models

An articulatory model of the velum [66], [65] was developed in order to complete an articulatory model already comprising other articulators. The velum contour was delineated and extracted from a thousand of X-ray images corresponding to short sentences in French. A principal component analysis was applied in order to derive the main deformation modes. The first component corresponds to the opening and comes with a shape modification linked to the apparition of a bulb in the upper part of the velum when it rises. The area function of the oral tract is modified so as to incorporate the velum movements. This model was connected with acoustic simulations in order to synthesize sentences containing French nasal vowels and consonants.

7.1.2. Expressive acoustic-visual synthesis

During this year, we have focused on the development of the acquisition infrastructure necessary to acquire audiovisual data. Mainly, we have developed several methods that allow acquiring acoustic and visual data synchronously. The visual data can originate from the Articulograph, Vicon or Intel RealSense devices. This heterogeneity of the data needs developing techniques to merge precisely the data in one unique reference. Synchronization techniques have also been developed for this purpose. We have evaluated the precision of the acquisition of such systems [61]. The combination of more than one motion capture technique aims to use the best quality data for each part of the face: (1) EMA (articulograph) for the lips, to have high precise measurement of the shape of the mouth that is related to speech and (2) kinect-like or Vicon system for the upper part of the face, that model mainly expressions.

We have acquired a small expressive audiovisual speech corpus of two actors: based on motion capture data (Vicon) and acoustic data. The content of the corpus is composed of six basic emotions (joy, sadness, anger, surprise, disgust and fear). This corpus will be used to investigate the characterization of emotions in audiovisual speech in the visual space and in the acoustic space.

We have also developed an algorithm to animate the 3D model of human face from a limited number of markers. The animation is very efficient and provides realistic animation results [82]. The 3D face will be used with the audiovisual system.

7.1.3. Categorization of sounds and prosody for native and non-native speech

7.1.3.1. Categorization of sounds for native speech

We investigated the schooling of a population of 166 students from primary to intermediate and secondary schools. These children and teenagers had specific language impairment: SLI (severe language impairment), dyslexia, dysorthographia. Since their childhood, they faced phonemic discrimination, phonological and phonemic analysis difficulties. We observed that they had trouble learning to read and more generally they experienced learning difficulties. Consequently, this lead them to repeat one or more grades, whereas in France, repetition is prohibited within each cycle and very limited between cycles.

7.1.3.2. Analysis of non-native pronunciations

Thanks to the detailed manual annotation of the French-German learner corpus that was carried out at the phonetic level in the IFCASL project (cf. 9.1.2), it was possible to investigate non-native pronunciation variants. The analysis revealed that German learners of French have most problems with obstruents in word-final position, whereas French learners of German show complex interferences with the vowel contrasts for length and quality [41]. Also, the correct pronunciation rate of the sounds, for several phonetic classes, was analyzed with respect to the learner's level, and compared to native pronunciations. One outcome is that different sound classes show different correct rates over the proficiency levels; and, for the German data, the frequently occurring syllabic [=n] is a prime indicator of the proficiency level.

We analyzed the realizations of French voiced fricatives by German non-native and French native speakers, in final position of an accentual group, a position where German fricatives are devoiced [27], [28]. Three speaker levels (from beginners to advanced) and different boundary types (depending on whether the fricative is followed by a pause, a schwa, or is directly followed by the first phoneme of the subsequent group) were considered. A set of cues, among which periodicity and fricative duration, have been analyzed. Results argue in favor of an influence of L1 (German) final devoicing on non-native realizations and show a strong interdependence between voicing, speakers' level, prosodic boundaries. The influence of orthography also strongly influenced voicing results.

We also investigated the realization of the short/long German contrast by French learners through three methods [60]. All these methods - phonetic annotation, perceptual experiment and acoustic analysis - used the same database (the IFCASL corpus). Depending on the method the results shed light on slightly different aspects of the same process, the interference of the French phonetic and phonological systems on the production of the German L2 vowels. Whereas the first method (phonetic annotation) revealed that especially rounded vowels are problematic in the long/short distinction, we could show with the second method (a perceptual experiment) that particularly the [o:]/[O] distinction seems to be hard to produce for French learners. The third method (an acoustical analysis) corroborated this finding and added acoustic details on duration and formants. The results of the studies can be used to create individualized training and feedback for foreign language learners, aimed at reducing their accent in L2.

7.2. Statistical Modeling of Speech

Participants: Antoine Liutkus, Emmanuel Vincent, Irina Illina, Dominique Fohr, Denis Jouvet, Joseph Di Martino, Vincent Colotte, Ken Deguernel, Amal Houidhek, Xabier Jaureguiberry, Aditya Nugraha, Luiza Orosanu, Imran Sheikh, Nathan Souviraà-Labastie, Dung Tran, Imene Zangar, Mohamed Bouallegue, Thibaut Fux, Emad Girgis, Juan Andres Morales Cordovilla, Sunit Sivasankaran, Freha Boumazza.

7.2.1. Source separation

Audio source separation is an inverse problem, which requires the user to guide the separation process using prior models for the source spectra and their spatial covariance matrices. We studied the impact of deterministic subspace constraints [14] over the spatial covariance matrices and pursued our work on the separation of multichannel mixtures guided by multiple, deformed reference signals such as repeated excerpts of the same music or repeated versions of the same sentence uttered by different speakers [17], [56]. Other models we have been working on include those based on local regularities of the spectral representations of musical sources (KAM, [52], [43], [51]). We also validated the positive impact of speech enhancement based on the FASST toolbox on speaker recognition [53].

As a new research direction, we extended the Gaussian framework for source separation to the family of α-stable stochastic processes [42]. This extension notably opens the path to new and robust parameters estimation algorithms for source separation [16], [67], that should be less prone to local minima. Current research notably comprises multichannel stable processes.

In parallel, we started yet another research track on the use of deep learning for source separation [24]. We proposed a new multichannel enhancement technique that exploits both the spatial properties of the sources as modeled by their spatial covariance matrices and their spectral properties as modeled by a deep neural network [75]. The model parameters are alternately estimated in an expectation-maximization (EM) fashion. We used this technique for music separation and speech enhancement in the context of the 2015 Signal Separation Evaluation Campaign (SiSEC) and the 3rd CHiME Speech Separation and Recognition Challenge, respectively [55]. We also used deep learning to address the fusion of multiple source separation techniques and found it to perform much better than the variational Bayesian model averaging techniques previously investigated [81].

Finally, we pursued our long-lasting efforts on the evaluation of audio source separation by co-organizing the 2015 Signal Separation Evaluation Campaign (SiSEC) [69] and writing a position paper about the scaling up of dataset sizes [21].

The ANR young researcher project KAMoulox (2016-2019 - cf. 9.1.5), that has just been accepted will deal with large audio archives, and more precisely with the "Archives du CNRS — Musée de l'homme" that gather a large set of old and noisy audio recordings (cf. 4.4). The work on source separation can lead to the design of semi automatic denoising and enhancement features, that would allow these researchers to significantly enhance their investigation capabilities, even without expert knowledge in sound engineering.

7.2.2. Acoustic modeling

We explored the use of an auxiliary function technique for fast training of neural networks [58]. We did not apply this technique to deep neural network acoustic models yet.

In the framework of using speech recognition for helping communication with deaf or hard-of-hearing people, robustness of the acoustic modeling was investigated. Studies were related to improving robustness with respect to speech signal level and environment noise through multicondition training and enhanced set of acoustic features (noise robust features or standard features after spectral noise subtraction) [37].

7.2.3. Linguistic modeling

7.2.3.1. Out-of-vocabulary proper name retrieval

Recognition of proper names (PN) is a challenging task in information retrieval in large audio/video databases. Proper names are semantically rich and are usually key to understanding the information contained in a document. Within the ContNomina project (cf. 9.1.3), we focus on increasing the vocabulary coverage of a speech transcription system by automatically retrieving proper names from contemporary text documents. We proposed methods that dynamically augment the automatic speech recognition system vocabulary, using lexical and temporal features in diachronic documents (documents that evolve over the time). Our work uses temporal context modeling to capture the lexical information surrounding proper names so as to retrieve out-of-vocabulary (OOV) proper names and increase the automatic speech recognition vocabulary.

We proposed new methods to retrieve OOV PNs relevant to an audio news document by using probabilistic topic models. We addressed retrieval of rare OOV PNs, which further improves the recall. Our proposed lexical context model improves the mean average precision of OOV PN retrieval [62]. We also proposed a two step approach for recognition of OOV PNs in an audio document. The first step retrieves OOV PNs relevant to an audio document using probabilistic topic models; and the second step uses a phonetic search for the target OOV PNs using a k-differences approximate string matching algorithm [63]. In [64], we discuss two specific phenomena, word frequency bias and loss of specificity, which affect the retrieval of OOV PNs using Latent Dirichlet Allocation (LDA) topic models. We studied different entity-topic models, which are extensions of LDA designed to learn relations between words, topics and PNs. We showed that our proposed methods of rare OOV PN and lexical context re-ranking improve the recall and the mean average precision for the LDA and the entity-topic models.

For OOV retrieval, we proposed the continuous space word representation using neural networks. This continuous vector representation (word embeddings) is learned from large amounts of unstructured text data. To model semantic and lexical context of proper names, different strategies of local context modeling were proposed [34], [33]. We studied OOV PN retrieval using temporal versus topic context modeling, different word representation spaces for word-level and document-level context modeling, and combinations of retrieval results [38]. We extended the previously proposed neural networks for word embedding models: the word vector representation proposed by Mikolov is enriched by an additional non-linear transformation. This model allows to better take into account lexical and semantic word relationships [39].

7.2.3.2. Adding words in a language model

A novel approach was proposed to add some new words in an existing n-gram language model, based on a similarity measure between the new words to be added and words already present in the language model [47]. Based on a small set of sentences containing the new words and on a set of n-gram counts containing the known words (known for the current language model), we search for known words which have the most similar neighbor distribution (of the few preceding and few following neighbor words) to the new words. The similar words are determined through the computation of KL divergences on the distribution of neighbor words. The n-gram parameter values associated to the similar words are then used to define the n-gram parameter values of the new words.

7.2.3.3. Selecting data for training a language model

Large vocabulary language models for speech transcription are usually trained from large amounts of textual data collected from various sources, which are more or less related to the target task. Selecting data that matches the target task was investigated in this context [46], this leads to a small reduction of the perplexity, and a smaller size of the resulting language model.

7.2.3.4. Music language modeling

Similarly to speech, music involves several levels of information, from the acoustic signal up to cognitive quantities such as composer style or key, through mid-level quantities such as a musical score or a sequence of chords. The dependencies between mid-level and lower- or higher-level information can be represented through acoustic models and language models, respectively. We pursued our pioneering work on music language modeling, with a particular focus on the modeling of long-term structure [12]. We also assessed the applicability of our prior work on joint modeling of note and chord sequences to new corpora of improvised jazz music, with the difficulty that these corpora are very small.

7.2.4. Speech generation by statistical methods

7.2.4.1. Pathological voice transformation

With respect to pathological voice processing, a competing approach to signal processing techniques consists in recognizing the pathological voice in order to transform it in a text version that can be re-synthesized. Such an approach is currently being experimented, and preliminary results are quite encouraging [15].

7.2.4.2. HMM-based synthesis

This year, we started working on HMM-based synthesis in the framework of a CMCU PHC project with ENIT (Engineer school at Tunis-Tunisia; cf. 9.3.2.2). Two topics will be explored by two PhD students. The first topic deals with the building of an Arabic corpora along with the analysis of linguistic features which are relevant for the HMM-based synthesis of the Arabic language. The second topic deals with improving the quality of the HMM-based synthesis system. In parallel, we started applying the HTS system (HMM-based Speech Synthesis System) to the French language.

7.3. Uncertainty Estimation and Exploitation in Speech Processing

Participants: Emmanuel Vincent, Odile Mella, Dominique Fohr, Denis Jouvet, Agnès Piquard-Kipffer, Baldwin Dumortier, Luiza Orosanu, Dung Tran, Sucheta Ghosh, Antoine Chemardin, Aghilas Sini.

7.3.1. Uncertainty and acoustic modeling

7.3.1.1. Noise-robust speech recognition

In many real-world conditions, the target speech signal overlaps with noise and some distortion remains after speech enhancement. In order to motivate further work by the community, we created an international evaluation campaign on that topic in 2011: the CHiME Speech Separation and Recognition Challenge. After two successful editions in 2011 and 2013, we organized the third edition in 2015 [25].

The framework of uncertainty decoding assumes that this distortion has a Gaussian distribution and seeks to estimate its covariance matrix in order to exploit it for subsequent feature extraction and decoding. A number of uncertainty estimators have been proposed in the literature, which are typically based on fixed mathematical approximations or heuristics. We made a conceptual breakthrough by proposing to learn the estimator from data using a non-parametric estimator and discriminative training [18], [59]. With GMM-HMM acoustic models, we obtained on the order of 30% relative word error rate reduction with respect to conventional decoding (without uncertainty), that is about twice as much as the reduction achieved by the best single uncertainty estimator. We also started working on the propagation of uncertainty in deep neural network acoustic models [19] and on its use for noise-robust speaker recognition [54].

7.3.1.2. Other applications

Besides the above applications, we started exploring applications of uncertainty modeling to robot audition [23] and control of wind turbines [31]. In the first context, uncertainty arises about the location of acoustic sources and the robot is controlled to locate the sources as quickly as possible. In the second context, uncertainty arises about the noise intensity of each wind turbine and the turbines are controlled to maximize electrical production under a maximum noise threshold.

7.3.2. Uncertainty and speech recognition

In the framework of using speech recognition for helping communication with deaf or hard-of-hearing people in the FUI project RAPSODIE (cf. 9.1.7), the best way for displaying the speech transcription results has been investigated. To our knowledge there is no suitable, validated and currently available display of the output of automatic speech recognizer for hard-of-hearing persons, in terms of size, colors and choice of the written symbols. The difficulty comes from the fact that speech transcription results contain recognition errors, which may impact the understanding process. Although the speech recognition system does not know the errors it makes, through the computation of confidence measures, the speech recognizer estimates if a word or a syllable is rather correctly recognized or not; hence such information can be used to adjust the display of the transcription results. Different ways were investigated for displaying the speech recognition results which take also into account the reliability of the recognized items. In this qualitative study, 10 persons have been interviewed to find the best way of displaying the speech transcription results. All the participants are deaf with different levels of hearing loss and various modes of communication [50].

7.3.3. Uncertainty and phonetic segmentation

Within the framework of the IFCASL project (cf. 9.1.2), a speech corpus of native and non-native speech for the French-German language pair was designed and recorded. Besides beeing used for analyzing non-native phenomena (cf. 7.1.3.2), this corpus will be used for developing and assessing automatic algorithms that will provide diagnosis on the learner mispronunciations [78]. Therefore, the automatic alignments of the audio files corresponding to the French and German speakers uttering French sentences (4100 audio files) were manually checked and corrected by a group of seven French annotators (the German data were handled by the German partner). We analyzed with CoALT the inter-annotator agreement with respect to an expert annotator for boundary shifts, insertions and deletions as well as devoicing diacritic [45]. The accuracy of the phone boundaries on non-native speech were investigated with respect to the HMM acoustic models used. The best performance (smallest amount of non-native phone segments whose boundaries are shifted by more than 20 ms compared to the manual boundaries) was obtained by combining each French native HMM model with an automatically selected German native HMM model [35].

Within the ANR ORFEO project (cf. 9.1.6), we addressed the problem of the alignment of spontaneous speech. The audio files processed in the ORFEO project were recorded under various conditions with a large SNR range and contain extra speech phenomena and overlapping speech. We trained several sets of acoustic models and tested different methods to adapt them to the various audio files [36]. Moreover in the framework of the EQUIPEX ORTOLANG (cf. 9.1.1), a web application, ASTALI (cf. 6.2), was developped in order to align a speech signal with its corresponding orthographic transcription (given in simple text file for short audio signals or in .trs files as generated by transcriber for longer speech signals).

In conventional speech-text alignments, a 10 ms frame shift is usually used for the acoustic analysis which leads to a minimum duration of 30 ms for each phone segment. Such duration constraint may not fit with actual sound duration in fast speaking rate. To overcome such contraint, a 5 ms frame shift can be used. Statistics on pronunciations variants estimated on large speech corpora have shown that when the conventional 10 ms frame shift is used, the frequency of the longest pronunciation variants gets underestimated [26]. Moreover, the analysis of some pronunciation variant frequencies have shown that some final consonantal cluster completely disappear at high speaking rates [40].

7.3.4. Uncertainty and prosody

Detection of sentence modality (question vs. affirmation) has been investigated using linguistic and prosodic features. Best results are achieved when the classifier uses all the available information [48], that is both linguistic and prosodic features. A detailed analysis has also shown that small errors in the determination of the sentence boundaries are not critical [49].

Speech-text alignments have been used to extract speech segments containing words and expressions that can be used either as normal lexical words or as discourse particles (as for example *quoi*, *voilà*, ...). The prosodic features for these words and expressions were extracted and analyzed [30]; automatic identification of the word function (discourse particle or not) from these prosodic features was also investigated.

In the context of the EQUIPEX ORTOLANG (cf. 9.1.1), several algorithms for computing the fundamental frequency have been implemented in the JSnoori software. These features can be computed directly from the GUI interface or through Python scripts. Future work will focus on improving the quality and robustness of the fundamental frequency estimation, and on determining the reliability of the estimations.

NEUROSYS Project-Team

7. New Results

7.1. From the Microscopic to the Mesoscopic Scale

Participants: Laure Buhry, Axel Hutt, Francesco Giovannini, Jean-Baptiste Schneider
In collaboration with LieJune Shiau (University of Houston)

7.1.1. Memory and Anaesthesia

7.1.1.1. Hippocampal Memory Networks

To improve our understanding of the effects of anaesthesia on the neural correlates of memory, we focussed on how anaesthetics disrupt the interaction between the hippocampus and the cerebral cortex. As a first step towards this objective Francesco Giovannini modelled a hippocampal pyramidal neuron using the Hodgkin-Huxley model capable of exhibiting long-lasting persistent firing activity when subject to a strong transient stimulus [16]. This behaviour is underlay by an intrinsic membrane current activated by the increase of intracellular calcium ions, following the discharge of an action potential by the neuron, in accord with that displayed in neural recordings of hippocampal slice preparations. Connecting these persistent firing neurons in a network comprising strong local excitation yields a wide range of behaviours depending on the interaction between CAN and synaptic currents. Indeed, the network model is capable of displaying rhythmic behaviour in the form of short synchronised bursts with intra-burst frequencies of $20 - 40\,Hz$ and inter-burst frequencies of $3\,Hz$. Furthermore, coupling CAN-equipped pyramidal neurons with a population of fast-spiking inhibitory interneurons yields emerging synchronous activity whose frequency is modulated by the strength of this coupling. These results hint towards a possible mechanism for the generation of memory-related oscillatory activity in the hippocampus.

7.1.1.2. Anaesthetic Effects on Hippocampal Oscillations

We investigated the effects of propofol-mediated tonic inhibition on the synchronous activity elicited in a network of hippocampal inhibitory interneurons. This work was conducted in collaboration with Jean-Baptiste Schneider, as part of hist 2-month internship. We studied the effect of propofol-induced tonic inhibition on the oscillations elicited in a network of hippocampal Hodgkin-Huxley gamma-aminobutyric acid ($GABA$) interneurons by studying the action of propofol on extrasynaptic GABAergic receptors. Our results [15] show that increasing doses of propofol reduce the overall network activity and slow down its oscillations until a critical value at which the synchronisation increases abruptly at values of twice the synchronisation displayed in the absence of tonic inhibition, and the mean firing rate increases. This emergence of synchronous activity mediated by anaesthetic perfusion point towards a possible mechanism for the emergence of paradoxical excitation under general anaesthesia.

In this context, Laure Buhry works with LieJune Shiau (University of Houston) on a better understanding of the models used by the community of computational neuroscientists. The goal is to show in which extent models are comparable or interchangeable. We focus on the comparison of oscillatory mechanisms of neuronal populations in different spiking models, especially in the Hodgkin-Huxley and the adaptive exponential integrate-and-fire model (AdEx). Especially, we have shown that a same number of synaptic connection per neuron is necessary to elicit synchronization in inhibitory neural networks of adaptive exponential integrate and fire neurons as in networks of Hodgkin-Huxley neurons. We have also conducted an extensive study regarding the effects of the different parameters of the AdEx model on the synchronization mechanisms in inhibitory neural networks, particularly in the context of gamma oscillations. A manuscript will be submitted soon to the Journal of Computational Neuroscience.

7.1.1.3. Noise Effects on Neural Rhythms

We have continued working on the effect of additive noise on neural oscillations and have shown that additive noise modulates the frequency of self-sustained neural rhythms [3].

7.2. From the Mesoscopic to the Macroscopic Scale

Participants: Laurent Bougrain, Axel Hutt, Pedro Garcia-Rodriguez, Eric Nichols, Guillaume Serrière, Tamara Tosic, Mariia Fedotenkova, Meysam Hashemi, Benjamin le Golvan, Cecilia Lindig-Leon, Sébastian Rimbert.

7.2.1. Level of Consciousness

7.2.1.1. Spatio-temporal Dynamics in Neural Fields

Neural fields serve as a model for experimental macroscopic activity. We have developed a numerical simulator NeuralFieldSimulator [21]. In addition, we have worked out a neural neural field model that exhibits a sequence of metastable activity states as observed in experimental data [4].

7.2.1.2. Synchronisation in Local Field Potentials under Anaesthesia

We have applied advanced data analysis techniques based on wavelet analysis to detect instantaneous partial synchronisation in experimental data [5].

7.2.1.3. Statistical Frequency-dependent Analysis by Recurrence Plots

Participants : Axel Hutt, Mariia Fedotenkova, Tamara Tosic
In collaboration with Flavio Frohlich, Peter Beim Graben and Kristin K. Sellers
For decades, research in neuroscience has supported the hypothesis that brain dynamics exhibits recurrent metastable states connected by transients, which together encode fundamental neural information processing. To understand the system's dynamics it is important to detect such recurrence domains, but it is challenging to extract them from experimental neuroscience datasets due to the large trial-to-trial variability. We proposed a methodology to extract recurrent metastable states in univariate time series by transforming datasets into their time-frequency representations and computing recurrence plots based on instantaneous spectral power values in various frequency bands [6]. Additionally, a new statistical inference analysis compares different trial recurrence plots with corresponding surrogates to obtain statistically significant recurrent structures. This combination of methods is validated by applying it to two artificial datasets. In a final study of visually-evoked Local Field Potentials in partially anesthetized ferrets, the methodology is able to reveal recurrence structures of neural responses with trial-to-trial variability. Focusing on different frequency bands, the delta-band activity is much less recurrent than alpha-band activity. Moreover, alpha-activity is susceptible to pre-stimuli, while delta-activity is much less sensitive to pre-stimuli. This difference in recurrence structures in different frequency bands indicates diverse underlying information processing steps in the brain.

7.2.2. Motor System

Participants: Laurent Bougrain, Axel Hutt, Benjamin le Golvan, Cecilia Lindig-Leon, Sébastian Rimbert, Guillaume Serrière

7.2.2.1. Motor Patterns during General Anesthesia

Participants: Laurent Bougrain, Axel Hutt, Cecilia Lindig-Leon, Sébastian Rimbert, Guillaume Serrière
The dosage of the anesthetic agent is tricky: too low, it does not achieve a sufficient loss of consciousness and may lead to a partial memorization during surgery and a post-operative trauma; too strong, it is dangerous for people with respiratory or heart problems. To better monitor the effect of the current dosage, we propose to study the dynamics of the motor brain activity during anesthesia. The relationship between motor brain activity and anesthesia is not intensively studied. Yet even if no physical movement by the patient is visually detectable, an electroencephalographic analysis of brain activity in motor areas may reveal an intention movement. This information is important because it demonstrates that the patient is conscious. We started to define a clinical protocol in collaboration with anesthesiologists of the hospital in Nancy to investigate is possibility. To reduce the duration of the protocol, we studied the minimum duration of a motor imagery to allow its detection from EEG recordings [23]. A large number of Brain-Computer Interfaces (BCIs) are based on the detection of motor imagery related features in the electroencephalographic signal. In most BCI experimental paradigms, subjects realize continuous motor imagery, i.e. a prolonged intention of movement, during a time window of a few seconds. Then, the system detects the movement based on the event-related desynchronization (ERD) and the event-related synchronization (ERS) principles. We studied if a discrete motor imagery, corresponding to a single short motor imagery, would allow a better detection of ERD and ERS patterns than a continuous motor

imagery. Indeed, the results of experiments involving 11 healthy subjects suggest that a continuous motor imagery generates a later ERS as well as a more variable and less detectable ERD than discrete motor imagery [11]. This finding suggests an improved experimental paradigm. We deeper investigated the amplitude and latency of EEG Beta activity during real movements, discrete and continuous motor imageries [22].

7.2.2.2. Motor Patterns during Combined Movements

Participants: Laurent Bougrain, Cecilia Lindig-Leon
Imaginary motor tasks cause brain oscillations that can be detected through the analysis of electroencephalographic (EEG) recordings. We studied whether or not the characteristics of the brain activity induced by the combined motor imagery (MI) of both hands can be assumed as the superposition of the activity generated during simple hand MIs. After analyzing the sensorimotor rhythms in EEG signals of five healthy subjects, results show that the imagination of both hands movement generates in each brain hemisphere similar activity as the one produced by each simple hand MI in the contralateral side [8]. Furthermore, during simple hand MIs, brain activity over the ipsilateral hemisphere presents similar characteristics as those observed during the rest condition. Thus, it is shown that the proposed scheme is valid and promising for brain-computer interfaces (BCI) control, allowing to easily detect patterns induced by combined MIs. Based on these results, we proposed a new method to extend the classic Common Spatial Pattern (CSP) algorithm to a multi-class approach which analyses both brain hemispheres separately to solve, together with a stepwise classification strategy, a multi-label BCI problem. After testing the proposed approach over the EEG signals of six healthy subjects performing a four-class multi-label task involving simple and combined hand MIs together with the rest condition, results show that this technique is plausible for BCI control [7]. In terms of accuracy, it outperforms the classical one-vs-one approach by 20% and has the same performance as the one-vs-all method. Nevertheless, to solve a multi-label classification problem involving k classes, the proposed method requires only log2(k) classifiers, whereas the one-vs-one method uses k(k-1)/2 classifiers and the one-vs-all k classifiers, thereby the new approach simplifies the classification task and seems promising for solving multi-label problems involving numerous classes.

7.2.2.3. On-line Detection of the End of Motor Imageries

Participants : Cécilia Lindig-León, Laurent Bougrain and Sébastien Rimbert
Limb movement execution or imagination induce sensorimotor rhythms that can be detected in electroencephalographic (EEG) recordings. We presented the interest of considering not only the beta frequency band but also the alpha band to detect the elicited EEG rebound, i.e. the increasing of oscillatory power synchronization, at the end of motor imageries [9], [19]. This phenomenon can be stronger over the alpha than the beta band and it is experimentally demonstrated [9] that the analysis on the alpha band improves the detection of the end of motor imageries. Moreover a variant method to compute the oscillatory power without referring to a baseline period is proposed; such capacity is useful for self-paced BCI control.

7.2.3. Pain under General Anaesthesia

7.2.3.1. Detection of EEG-signal Features for Pain under General Anaesthesia

Participants : Axel Hutt, Mariia Fedotenkova
In collaboration with Peter Beim Graben and James W. Sleigh
Nowadays, surgical operations are impossible to imagine without general anaesthesia, which involves loss of consciousness, immobility, amnesia and analgesia. Understanding mechanisms underlying each of these effects guarantees well-controlled medical treatment. Our work focuses on analgesia effect of general anaesthesia, more specifically, on patients reaction on nociception stimuli. The study was conducted on dataset consisting of 230 EEG signals: pre- and post-incisional recordings for 115 patients, who received desflurane and propofol. Initial analysis was performed by power spectral analysis, which is a widespread approach in signal processing. Power spectral information was described by fitting the background activity and measuring power contained in delta and alpha bands according to power of background activity. The fact that power spectrum of background activity decays as frequency increasing is well known and thoroughly studied. Here, traditional $1/f^{\alpha}$ behaviour of the decay was replaced by a Lorentzian model to describe the power spectrum of background activity. Due to observed non-stationary nature of EEG signals spectral analysis does not suffice

to reveal significant changes between two states. A further improvement was done by expanding spectra with time information. To obtain time-frequency representations of the signals conventional spectrograms were used as well as a spectrogram reassignment technique. The latter allows to ameliorate readability of a spectrogram by reassigning energy contained in spectrogram to more precise positions. Subsequently, obtained spectrograms were used in recurrence analysis and its quantification by complexity measure. Recurrence analysis allows to describe and visualise dynamics of a system and discover structural patterns contained in the data. Structure of each recurrence plot is characterised by Lempel–Ziv complexity measure [5], which shows a difference between pre- and post-incision [13].

ORPAILLEUR Project-Team

7. New Results

7.1. The Mining of Complex Data

Participants: Mehwish Alam, Aleksey Buzmakov, Victor Codocedo, Miguel Couceiro, Adrien Coulet, Esther Galbrun, Nicolas Jay, Florence Le Ber, Luis-Felipe Melo, Amedeo Napoli, Chedy Raïssi, Mohsen Sayed, My Thao Tang, Yannick Toussaint.

Keywords: formal concept analysis, relational concept analysis, pattern structures, pattern mining, association rule, graph mining, sequence mining, biclustering

Pattern mining and Formal Concept Analysis are suitable symbolic methods for KDDK, that may be used for real-sized applications. Global improvements are carried out on the scope of applicability, the ease of use, the efficiency of the methods, and on the ability to fit evolving situations. Accordingly, the team is extending these symbolic data mining methods for working on complex data (e.g. textual documents, biological, chemical or medical data), involving objects with multi-valued attributes (e.g. domains or intervals), n-ary relations, sequences, trees and graphs.

7.1.1. FCA and Variations: RCA, Pattern Structures and Biclustering

Advances in data and knowledge engineering have emphasized the needs for pattern mining tools working on complex data. In particular, FCA, which usually applies to binary data-tables, can be adapted to work on more complex data. In this way, we have contributed to two main extensions of FCA, namely Pattern Structures and Relational Concept Analysis. Pattern Structures (PS [92]) allow to build a concept lattice from complex data, e.g. numbers, sequences, trees and graphs. Relational Concept Analysis (RCA) is able to analyze objects described both by binary and relational attributes [101] and can play an important role in text classification and text mining. Following this way, and regarding itemset and association rule discovery, we improved standard algorithms for building lattices from large data and for completing the algorithm collection of the Coron platform [103].

Many developments were carried out in pattern mining and FCA for improving data mining algorithms and their applicability, and for solving some specific problems such as information retrieval, discovery of functional dependencies and biclustering. We designed new information retrieval methods based on FCA where the concept lattice is considered as an index space for answering disjunctive queries [54]. We developed also a whole line of work on pattern structures for the discovery of functional dependencies [80], text classification and heterogeneous pattern structures [83], and pattern structures for structured attribute sets [46]. FCA can also be considered as a clustering method and we adapted pattern structures to clustering for analyzing numerical datatables supporting recommendation problems [13]. Projections can be associated with pattern structures for leveraging the volume and the complexity of the computation [53]. We designed also a quasi-polynomial algorithm for mining top patterns w.r.t. measures satisfying special properties in a FCA framework [52]. We also proposed new visualization techniques and tools able to display important and useful information (e.g. stable concepts) from large concept lattices [49].

Still considering complex data, we worked on the analysis of molecular structures (or molecular graphs) [34]. The mining of molecular graphs is an important task for many reasons, among which the challenges it represents regarding knowledge discovery, life sciences and healthcare, and, as well, the industrial needs that can be met whenever substantial results are obtained (especially in pharmacology).

7.1.2. Text Mining

Ontologies help software and human agents to communicate by providing shared and common domain knowledge, and by supporting various tasks, e.g. problem-solving and information retrieval. In practice, building an ontology or at least "ontological concept definitions" depends on a number of ontological resources having different types: thesaurus, dictionaries, texts, databases, and ontologies themselves. We are currently working on the design of a methodology based on FCA and RCA for ontology engineering from heterogeneous ontological resources. This methodology is based on both FCA and RCA, and was previously successfully applied in domains such as astronomy and biology.

In the framework of the ANR Hybride project (see 8.2.1.2), an engineer is implementing a robust system based on these previous research results, for preparing the way to new research directions involving trees and graphs. Moreover, we led a first successful experiment on extracting drug-drug interactions applying "lazy pattern structure classification" to syntactic trees [66]. In addition, in his thesis work, Mohsen Sayed focused on extracting relations between named entities using graph mining methods applied to dependency graphs. We are currently investigating how this approach can be generalized, i.e. how to detect a relation between complex expressions which are not previously recognized as named entities [64].

The notion of "Jumping Emerging Patterns" (JEP) previously used in chemistry [12], was updated and adapted to the context of text mining within the ANR Termith project. The objective is to design a learning method for filtering candidate terms within a full text and to decide whether an occurrence should be tagged as a term, i.e. as a positive example, or as a simple word, i.e. as a negative example. The method extracts from a training set all JEPs which are considered as hypotheses [7]. To reduce the number of JEPs and to only retain the most significant from a linguistic point of view, JEPs are weighted and a constraint solver is used to check the maximal coverage of the positive examples. Results are currently under evaluation.

7.1.3. Mining Sequences and Trajectories

Sequence data is widely used in many applications. Computing the similarity between sequences is a very important challenge for many different data mining tasks. There is a plethora of similarity measures for sequences in the literature, most of them being designed for sequences of items. In a recent work with Elias Egho, we study the problem of measuring the similarity between sequences of itemsets [32]. We focus on the notion of common subsequences as a way to measure similarity between a pair of sequences composed of a list of itemsets. In this work, we present new combinatorial results for efficiently counting distinct and common subsequences. These theoretical results are the cornerstone of an effective dynamic programming approach to deal with this problem. In addition, we develop an approximate method to speed up the computation process for long sequences. We have applied the method to various data sets: healthcare trajectories, on-line handwritten characters and synthetic data. The results confirm that the current similarity measure produces competitive scores and indicate that the method is relevant for large scale sequential data analysis.

Nowadays data sets are available in very complex and heterogeneous ways. Mining of such data collections is essential to support many real-world applications ranging from healthcare to marketing. In a recent work, we focused on the analysis of "complex sequential data" by means of interesting sequential patterns [19]. We approach the problem using FCA and pattern structures, where the subsumption relation ordering patterns is defined w.r.t. the partial order on sequences. We show how pattern structures along with projections, i.e. a data reduction of sequential structures, are able to enumerate more meaningful patterns and increase the computing efficiency of the approach. Finally, we demonstrate the applicability of the method for discovering and analyzing patient patterns from a French healthcare data set on cancer. The quantitative and qualitative results –with annotations and analysis from a physician– are reported in this use case which is one main motivation for this work.

7.1.4. Mining with Preferences

In the last decade, the pattern mining community has witnessed a sharp shift from efficiency-based approaches to methods which can extract more meaningful patterns. Recently, new methods adapting results from studies of economic efficiency and multi-criteria decision analysis such as Pareto efficiency, or skylines, have been studied. Within pattern mining, this novel line of research allows the easy expression of preferences according

to a dominance relation. We have developed approaches that are useful from a user-preference point of view, tending to promote the use of pattern mining algorithms for non-experts. These approaches are based on the discovery of skyline patterns, or "skypatterns", in relation with condensed representations of patterns. This last relationship facilitates the computation of skypatterns, providing a flexible and efficient approach to mine skypatterns reusing a dynamic constraint satisfaction problems (CSP) framework [8].

7.1.5. Aggregation

Aggregation or consensus theory studies any process dealing the merging of several objects (numerical values, qualitative data, preferences, etc.) into a single (or several) object of similar type and that, in some way, is the best representation. The need to aggregate objects in a meaningful way has become more and more present in an increasing number of areas not only of mathematics, statistics or physics, but especially in applied fields such as engineering, computer science, social sciences and biology. In social choice and multicriteria decision aid, objects are preferences that are expressed by users, voters or criteria, and are modeled by order relations or utility functions. In cluster analysis, the objects to merge are classifications (such as partitions, hierarchies or trees) or related functions (such as similarity/dissimilarity measures).

With the proliferation of massive databases and new fields such as computational advertising, search engines and recommender systems, the need for information retrieval and knowledge discovery processes became emergent as well as the construction of user preference models for classification and prediction purposes. Also in biology and phylogenetics, aggregation is used to find consensus patterns among DNA sequences or finding consensus trees within taxonomies. As algorithms are often heuristic in such large datasets, they rarely produce the same output, highlighting the importance of finding means of aggregation to produce consensus structures. The difficulty in extracting such consensus structures comes down to define appropriate aggregation rules (e.g., counting and median procedures), and their impossibility is many times revealed by Arrowian results. A way to avoid such impossibility results is the consideration of alternative aggregation rules or the weakening of underlying structures, for instance weak hierarchies that allow overlapping clusters while keeping desirable tree-like properties.

We are working on a theoretical basis of a unified theory of consensus and to set up a general machinery for the choice and use of aggregation functions. This choice depends on properties specified by users or decision makers, the nature of the objects to aggregate as well as computational limitations due to prohibitive algorithmic complexity. This problem demands an exhaustive study of aggregation functions that requires an axiomatic treatment and classification of aggregation procedures as well as a deep understanding of their structural behavior. Moreover, Arrowian results are also envisioned since they constitute an important tool in the identification of reasonable algebraic/relational structures for representing data as well as in the identification of meaningful aggregation processes.

Direct applications of this theory are preference learning and cluster analysis. In the first case, preferences are represented by global utility functions and alternatives with higher utilities are preferred. Moreover, simplified versions of this model will be explored in the context of feature selection for both dimension reduction of data as well as classifier design. In the second case, we consider median structures that include several ordered/relational structures (trees, graphs, orders) and that allow several consensus procedures. This is particularly useful in a context of classification that takes into account evolutionary relations between classes, for instance, in taxonomical biology and phylogenetics.

7.1.6. Video Game Analytics

The video game industry has enormously grown over the last twenty years, bringing new challenges to the artificial intelligence and data analysis communities. We are studying the automatic discovery of strategies in real-time strategy games through pattern mining. Such patterns are the basic units for many tasks such as automated agent design, but also to build tools for the professionally played video games in the electronic sports scene. Continuing our joint collaboration with researchers from the MIT GameLab we successfully extended our previous work to a journal paper that will be published in 2016.

7.2. Knowledge Discovery in Healthcare and Life Sciences

Participants: Miguel Couceiro, Adrien Coulet, Amedeo Napoli, Chedy Raïssi, Mohsen Sayed, Malika Smaïl-Tabbone, Yannick Toussaint.

Life Sciences constitute a challenging domain for KDDK. Biological data are complex from many points of views, e.g. voluminous, high-dimensional and deeply inter-connected. Analyzing such data is a crucial issue in healthcare, environment and agronomy. Besides, many bio-ontologies are available and can be used to enhance the knowledge discovery process. Accordingly, the research work of the Orpailleur team in KDDK applied to Life Sciences is in concern with the use of bio-ontologies to improve KDDK, and as well information retrieval, access to "Linked Open Data" (LOD) and data integration.

7.2.1. Ontology-based Clustering of Biological Linked Open Data

Increasing amounts of biomedical data provided as Linked Open Data (LOD) offer novel opportunities for knowledge discovery in bio-medicine. We proposed an approach for selecting, integrating, and mining LOD with the goal of discovering genes responsible for a disease [99]. We are currently working on the integration of LOD about known phenotypes and genes responsible for diseases along with relevant bio-ontologies. We are also defining a corpus-based semantic distance. One possible application of this work is to build and compare possible diseaseomes, i.e. global graphs representing all diseases connected according to their pairwise similarity values.

7.2.2. Suggesting Valid Pharmacogenes by Mining Linked Open Data and Electronic Health Records

A standard task in pharmacogenomics research is identifying genes that may be involved in drug response variability and called "pharmacogenes". As genomic experiments in this domain tend to generate many false positives, computational approaches based on background knowledge may generate more valuable results. Until now, the later have used only molecular networks databases or biomedical literature. We are studying and working on a novel method that take advantage of an eclectic set of linked data sources to validate uncertain drug–gene relationships, i.e. pharmacogenes [3]. One advantage relies on the standard implementation of linked data that facilitates the joint use of various sources and makes easier the consideration of features of various origins. Accordingly, we proposed an initial selection of linked data sources relevant to pharmacogenomics. We formatted these data to train a random forest algorithm, producing a model that classify drug–gene pairs as related or not, thus validating candidate pharmacogenes.

With this same motivation of validating state-of-the-art knowledge in pharmacogenomics, a new ANR project called "PractiKPharma" will be initiated in 2016 and will rely on similar ideas. The originality of "PractiKPharma" is to use "Electronic Health Records" to constitute cohorts of patients that are then mined for validating extracted pharmacogenomics knowledge units (http://practikpharma.loria.fr/).

7.2.3. Biological Data Aggregation for Knowledge Discovery

During this year, in collaboration with the Capsid Team, we contributed to write up two multi-disciplinary projects with a group of clinicians from the Regional University Hospital (CHU Nancy) and bio-statisticians from the Maths Lab (IECL). The first project, entitled ITM2P [0] lying in the so-called CPER 2015–2020 framework, was accepted and granted. The funding is mainly intended for medical and computing equipments and will be used to set up four scientific platforms. We are involved in the SMEC platform as a support for "Simulation, Modeling and Knowledge Extraction from Bio-Medical Data".

The second project is a RHU [0] project entitled *Fight Heart Failure* (FHF) and was accepted as a so-called "investissement d'avenir" and granted. We are in charge of a workpackage which will give us the opportunity of exploring important research questions. Among these questions, one is to define "data aggregation" mechanisms with a twofold objective: (i) the definition of pairwise patient similarity given that patients are described by complex dimensions involving relations and time and (ii) the efficient clustering of patients based

[0] "Innovations Technologiques, Modélisation et Médecine Personnalisée"
[0] "Recherche Hospitalo-Universitaire"

on this similarity measure. Each cluster should correspond to a bioprofile, i.e. a subgroup of patients sharing the same form of the disease and thus the same diagnosis and care strategy. For doing that, we are currently investigating consensus theories [95] and their applicability to a bio-medical context, and as well aggregation operators as defined in various contexts, e.g. databases, data-warehouses, web of data, and graph theory. The idea is to consider relational and temporal data aggregation as a first class citizen in the data preparation phase of the knowledge discovery. This allows to assess the contribution of aggregation for such a task and in this context.

Another question is related to the construction of a prediction model for each bioprofile/subgroup –once validated by the clinicians– to be used in a decision support system. This will likely require the combination of symbolic and numerical methods for the classification task.

7.2.4. Analysis of biomedical data annotated with ontologies

Annotating data with concepts of an ontology is a common practice in the biomedical domain. Resulting annotations define links between data and ontologies that are key for data exchange, data integration and data analysis. Since 2011, we collaborate with the National Center for Biomedical Ontologies (NCBO) to develop a large repository of annotations named the NCBO Resource Index. This repository contains annotations of 36 biomedical databases annotated with concepts of more than 200 ontologies of the BioPortal (http://bioportal.bioontology.org/). In the preceding years, we compared the annotations of a database of biomedical publications (Medline) with two databases of scientific funding (Crisp and ResearchCrossroads) to profile disease research. One main challenge is to mine these annotations.

As a first attempt, we adapted pattern structures to analyze the annotations of biomedical databases [85]. We considered annotated biomedical documents as objects and the corresponding annotations were classified according to various dimensions, i.e. a particular aspect of domain knowledge. The resulting classification of annotations allowed not only to discover correlations between annotations but also incomplete annotations that could be fixed afterward. This adaptation of pattern structures opens many perspectives in term of ontology reengineering and knowledge discovery.

7.3. Knowledge Engineering and Web of Data

Participants: Mehwish Alam, Aleksey Buzmakov, Victor Codocedo, Emmanuelle Gaillard, Florence Le Ber, Jean Lieber, Amedeo Napoli, Emmanuel Nauer.

Keywords: knowledge engineering, web of data, classification-based reasoning, case-based reasoning, belief revision, semantic web

7.3.1. Around the Taaable Research Project

The Taaable project was originally created as a challenger of the Computer Cooking Contest (ICCBR Conference) [84] (http://intoweb.loria.fr/taaable3ccc/). Beyond its participation to the CCC challenges, the Taaable project aims at federating various research themes: case-based reasoning (CBR), information retrieval, knowledge acquisition and extraction, knowledge representation, minimal change theory, ontology engineering, semantic wikis, text-mining, etc. CBR performs adaptation of recipes w.r.t. user constraints. The reasoning process is based on a cooking domain ontology (especially hierarchies of classes) and adaptation rules. The knowledge base is encoded within a semantic wiki containing the recipes, the domain ontology and adaptation rules.

As acquiring knowledge from experts is costly, a new approach was proposed to allow a CBR system to use partially reliable, non expert, knowledge from the Web for reasoning. This approach is based on notions such as belief, trust, reputation and quality, as well as their relationships and rules to manage the knowledge reliability. The reliability estimation is used to filter knowledge with high reliability as well as to rank the results produced by the CBR system. Performing CBR with knowledge resulting from an e-community is improved by taking into account the knowledge reliability [61].

Another study shows how the case retrieval of a CBR system can be improved using typicality. Typicality discriminates subclasses of a class in the domain ontology depending of how a subclass is a good example for its class. An approach has been proposed to partition the subclasses of some classes into atypical, normal and typical subclasses in order to refine the domain ontology. The refined ontology allows a finer-grained generalization of the query during the retrieval process, improving at the same time the final results of the CBR system [62].

The Taaable system also includes a module for adapting textual preparations (from a source recipe text to an adapted recipe text, through a formal representation in the qualitative algebra INDU). The evaluation of this module as a whole thanks to users has been carried out and has shown its efficiency (w.r.t. text quality and recipe quality), when compared with another approach to textual adaptation [4].

FCA allows to organize objects according to the properties they share into a concept lattice. A lattice has been built on a large set a cooking recipes according to the ingredients they use, producing a hierarchy of ingredient combinations. When a recipe R has to be adapted, this lattice can be used to search the best ingredient combinations in the concepts that are the closest to the concept representing R [63].

Minimal change theory and belief revision can be used as tools to support adaptation in CBR, i.e. the source case is modified to be consistent with the target problem using a revision operator. Belief revision was applied to Taaable to adjust the ingredient quantities using engines included in the Revisor library (see § 6.4.5). This year, a mixed linear optimization has implemented to produce human easy understandable quantities. For example, when the ingredient is a lemon, its quantity will take the form of a quarter, a half, etc., instead of *54 g* (which corresponds to a half lemon) [63].

7.3.2. Exploring and Classifying the Web of Data

A part of the research work in Knowledge Engineering is oriented towards knowledge discovery in the web of data, as, with the increased interest in machine processable data, more and more data is now published in RDF (Resource Description Framework) format. The popularization and quick growth of Linked Open Data (LOD) has led to challenging aspects regarding quality assessment and data exploration of the RDF triples that shape the LOD cloud. Particularly, we are interested in the completeness of the data and the their potential to provide concept definitions in terms of necessary and sufficient conditions [1]. We have proposed a novel technique based on Formal Concept Analysis which organizes subsets of RDF data into a concept lattice. This allows data exploration as well as the discovery of implication rules which are used to automatically detect missing information and then to complete RDF data and to provide definitions. Moreover, this is also a way of reconciling syntax and semantics in the LOD cloud. Experiments on the DBpedia knowledge base shows that this kind of approach is well-founded and effective.

Other important aspects are concerned with data access, data visualization w.r.t. the SPARQL query language [46], [49]. SPARQL queries over the web of data usually produce lists of tuples as answers that may be voluminous and hard to interpret. We introduced Lattice-Based View Access (LBVA), a framework based on FCA, which provides a classification of the answers of SPARQL queries based on a concept lattice. This concept lattice can be considered as a materialized view of the data resulting from a SPARQL query and can be navigated for retrieving or mining specific patterns. We associate a VIEW-BY clause to SPARQL for facilitating the interaction between analysts and LOD. The organization of answers is based on an original proposition on pattern structures for structured sets of attributes, which appears to be quite efficient and very well-adapted to the classification and analysis of RDF data. The visualization and the navigation of the concept lattice are guided by RV-Xplorer (i.e. RDF View eXplorer), an adapted interactive visualization system. Experiments show that the approach is well-founded and that it opens many new perspectives in the domain.

7.4. Advances in Graph Theory

Participants: Miguel Couceiro, Amedeo Napoli, Chedy Raïssi, Jean-Sébastien Sereni, Mario Valencia.

Keywords: graph theory, extremal graph theory, chromatic number, triangle-free graph, planar graph, graph coloring

We announced in the last report that we started to work on a conjecture by Heckman and Thomas from 1999. We managed to confirm the conjecture and the demonstration was published in January 2014. A classical result by Staton, from 1979, states that every triangle-free graph G with maximum degree at most 3 contains an independent set of order at least $5n/14$, where n is the number of vertices of G. Heckman and Thomas conjectured a stronger fact: the fractional chromatic number of such a graph is at most $14/5$. We confirmed their conjecture by establishing the following stronger assertion: for any assignment of weights (i.e., real numbers) to the vertices of such a graph G, there exists an independent set I such that the weights of the vertices in I is at least $5/14$ times the total weight of the G.

Exploring further the methods we introduced to solve this conjecture, we obtained new results concerning the fractional chromatic number of planar triangle-free graphs. While the fractional chromatic number of such graphs is at most 3 (because their chromatic number is), a construction of Jones proved the existence of triangle-free planar graphs with fractional chromatic number arbitrarily close to 3. Thus one wonders whether there could be such graphs with fractional chromatic number exactly 3. We demonstrated this not to be the case, by proving a general upper bound of $\frac{9n}{3n+1} = 3(1 - \frac{1}{3n+1})$ for every triangle-free planar graph G with n vertices. This bound is qualitatively the best possible: Jones's construction yields graphs with fractional chromatic number $3 - \frac{c}{n}$ for some constant c. In addition, a tight bound was obtained if the graphs considered are furthermore required to have maximum degree at most 4. In this case, the bound becomes $\frac{3n}{3n+1}$.

Motivated by frequency assignment in office blocks, we study the chromatic number of the adjacency graph of a 3-dimensional parallelepiped arrangement. In the case each parallelepiped is within one floor, a direct application of the Four-Colour Theorem yields that the adjacency graph has chromatic number at most 8. We provide an example of such an arrangement needing exactly 8 colors. We also discuss bounds on the chromatic number of the adjacency graph of general arrangements of 3-dimensional parallelepipeds according to geometrical measures of the parallelepipeds (side length, total surface area or volume).

SEMAGRAMME Project-Team

6. New Results

6.1. Syntax-semantics interface

6.1.1. Lexical Semantics

The interpretation of natural language utterances relies on two complementary elements of natural language modeling. On the one hand, the description of the combinatorics of natural language expresses how elementary units, or *lexical units* (typically the word), combine in order to build more complex elements, such as sentences or discourses. On the other hand, the description of these elementary units specifies how they contribute to the meaning of the whole by their *lexical meaning*. This specification should also take into account how the different parts of the lexical meanings combine during the *composition* process and how they relate to their underlying meaning concepts. For instance, the verbs *buy* and *sell* should refer to a common conceptual representation. However, their syntactic arguments (e.g., the subject) play a different (semantic) role with respect to the *transaction* concept that they share.

The modeling of these concepts and how they relate to each other gave rise to Frames Semantics as a representation format of conceptual and lexical knowledge [40], [31], [26], [59]. Frames consists of directed graphs where nodes correspond to entities (individuals, events, ...) and edges correspond to (functional or non-functional) relations between these entities. Providing a fine-grained representation of the internal concept structure allows both for a *decomposition* of the lexical meaning and for a precise description of the sub-structural interactions in the semantic composition process [58].

Frames can be formalized as extended typed feature structures [71], [50] and specified as models of a suitable logical language. Such a language allows for the composition of lexical frames on the sentential level by means of an explicit syntax-semantics interface [50]. Yet, this logical framework does not provide a direct link between Frames and truth-conditional semantics, where natural language utterances are considered with respect to the conditions under which they are true or false. In particular, it does not provide means for the lexical items to introduce explicit quantification over entities or events.

To overcome these limitations, we proposed use Hybrid Logic (HL) [27], [25]. HL is an extension of modal logic. As such, it is well-suited to the description of graph structures. Moreover, HL introduce *nominals*, that allow the logical formulas to refer to specific nodes of the graph. It is then possible, for example, to specify when two edges should meet. Moreover, it introduces *variables* for nodes, and the associated *quantifiers*, that can appear in the logical formulas. We used this framework to model quantification in Frame Semantics [23], [18]

6.1.2. Compositionality and Modularity

One says that a semantics is compositional when it allows the meaning of a complex expression to be computed from the meaning of its constituents. One also says that a system is modular if it is made of relatively independent components. In the case of a semantic system (e.g, a Montague grammar), we say that it is modular if the ontology on which it is based (including notions such as *truth, entities, events, possible worlds, time intervals, state of knowledge, state of believe, ...*) is obtained by combining relatively independent simple ontologies.

The intensionalization procedure introduced in [4] provides a first step towards modularity. It allows the extensional interpretation of a language to be transformed into an intensionalized interpretation that offers room for accommodating truly intensional phenomena. Moreover, this procedure is conservative in the sense that it preserves the truth conditions of sentences. Another instance of such a procedure is provided by the dynamization procedure described in [57], which allows a static interpretation to be turned into a dynamic one capable of accommodating phenomena related to discourse dynamics.

In [15], we showed that both the intensionalization and dynamization procedures are instances of an abstract general scheme for which conservativity results may be established using the notion of logical relation.

6.1.3. Abstract Categorial Parsing

Kanazawa [53], [54] has shown how parsing and generation may be reduced to datalog queries for a class of grammars that encompasses mildly context-sensitive formalisms. These grammars, which he calls *context-free λ-term grammars*, correspond to second-order abstract categorial grammars.

In [14], we showed how Kanazawa's reduction may be carried out in the case of abstract categorial grammars of a degree higher than two. To this end, we reduced the parsing problem for general Abstract Categorial Grammars to a provability problem in Multiplicative Exponential Linear Logic.

6.2. Discourse dynamics

6.2.1. Discourse Structure Modeling

It is usually assumed that the internal structure of a text, typically characterized by discourse or rhetorical relations, plays an important role in its overall interpretation. In order to build such a structure, some approaches rely on discourse grammars. The key idea is to consider the structural regularities in discourse structure similarly as syntactic regularities. A particular trend relies on tree grammars. This trend has been further developed by integrating the modeling of both clausal syntax and semantics, and discourse syntax and semantics within the frameworks of Tree-Adjoining Grammar (TAG) [48], [49] and TAG for Discourse (D-LTAG) [79], [41], [80], [42].

Two important features characterize these approaches. First, while they use a single grammatical formalism, two different grammars are used for syntactic parsing and then for discourse parsing. In addition to adding an intermediate processing step, this two-tiered treatment both complicates the modeling of connectives that are ambiguous in their syntactic and discourse use, and prevents using standard disambiguation techniques. Second, some discourse structures better represented by directed acyclic graphs (DAG) than by trees are not accounted for.

In order to address the second issue of building DAG structures, [36], [37] have proposed Discourse Synchronous TAG (D-STAG), a TAG based approach together with a higher-order interpretation of sentences using Synchronous Tree-Adjoining Grammar (STAG) [67], [77].

We developed a method to interface a sentential grammar and a discourse grammar without resorting to an intermediate processing step. The method is general enough to build discourse structures that are DAG and not only trees. Our analysis is based on D-STAG. We also use an encoding of TAG into ACG. This encoding allows us to express a higher-order semantic interpretation that enables building DAG discourse structures on the one hand, and to smoothly integrate the sentential and the discourse grammar thanks to the modular capability of ACG. The results has been published [13] and all the examples of the article have been implemented and may be run and tested with the ACGtk software (see 5.1).

6.2.2. Effects and Handlers

We made the argument that pragmatics are to semantics what side effects are to calculations in a programming language. We demonstrated this parallel on two aspects.

First off, both pragmatics and side effects serve the same function. Side effects in programming languages account for the effects of expressions that reach beyond their scope and for the way a language interacts with the world of its users. Pragmatics is concerned with phenomena that also involve the non-immediate effects of expressions (e.g., discourse anaphora, presupposition accommodation) and with the way language interacts with the world of its users. Secondly, we pointed out that very similar formal theories are being used to treat the both of them (i.e. monads and continuations).

Having established this parallel, we then put forward a preliminary proposal of integrating semantics and pragmatics while keeping them separate by assigning effectful computations of truth values as meanings of linguistic expressions. In this way, we can implement the pragmatics at the level of the side effects and then focus on pure semantics at the level of values.

6.3. Common basic resources

6.3.1. Graph Rewriting

Bruno Guillaume and Guy Perrier have proposed to use Graph Rewriting for parsing syntactic dependencies [17]. It is an application of a Graph Rewriting formalism that they have established with Guillaume Bonfante and Mathieu Morey [32] and implemented in the Grew software [47]. They have developed a system of rewriting rules dedicated to French, which they have evaluated by parsing the Sequoia corpus [33].

6.3.2. Categorial Logic

Elaborating on the work of Grishin [45], Moortgat has introduced the non-associative Lambek-Grishin calculus (**LG**) as the foundations of a new kind of symmetric categorial grammar [63], [64], which allows for the treatment of linguistic phenomena such as displacement or discontinuous dependencies.

In [16], we compared **LG** with the non-associative classical Lambek calculus (**CNL**) introduced by de Groote and Lamarche [81]. We provided a translation of **LG** into **CNL**, which allows **CNL** to be seen as a non-conservative extension of **LG**. We then introduced a bimodal version of **CNL** that we called 2-**CNL**. This allowed us to define a faithful translation of **LG** into 2-**CNL**. Finally, we showed how to accommodate Grishin's interaction principles by using an appropriate notion of polarity. From this, we derived a new one-sided sequent calculus for **LG**.

6.3.3. Deep Syntax Annotation of the Sequoia French Treebank

Deep-sequoia introduces a deep syntactic representation scheme for French, built from the surface annotation scheme of the Sequoia corpus and abstracting away from it [69]. This scheme expresses the grammatical relations between content words. When these grammatical relations take part into verbal diatheses, the diatheses are considered as resulting from redistributions from the canonical diathesis, which is retained in the annotation scheme. The first version of the deep-sequoia corpus was released in 2014.

In November 2015, a new version (7.0) of the corpus was release (see http://deep-sequoia.inria.fr). Most of the modifications were corrections of annotations that improve the overall consistency of the corpus. Marie Candito and Guy Perrier have published the annotation guidelines associated with the corpus in [22].

6.3.4. Large Scale Grammatical Resources

Guy Perrier and Bruno Guillaume have achieved the development of a French grammar FRIGRAM with a large coverage [12] in the formalism of Interaction Grammars [5]. The originality of the formalism lies in its system of polarities which expresses the resource sensitivity of natural languages and which is used to guide syntactic composition. We present the principles underlying grammar design, highlight its modular architecture and show that the lexicon used is independent of the grammar formalism. We also introduce the "companion property", and show that it helps to enforce grammar consistency.

6.3.5. Universal Dependency Treebank

Bruno Guillaume participates with Marie-Catherine de Marneffe to the production of the French sub-corpus of the Universal Dependency Treebank [68]. In November 2015, the version 1.2 was released. On the French sub-corpus, Grew was used to detect inconsistency and to correct automatically systematic errors.

SPHINX Team

7. New Results

7.1. Analysis, control and stabilization of heterogeneous systems

Motivated by the collision problem for rigid bodies in a perfect fluid, Munnier and Ramdani investigated in [9] the asymptotics of a 2D Laplace Neumann problem in a domain with cusp. The small parameter involved in the problem is the distance between the solid and the cavity's bottom. Denoting by $\alpha > 0$ the tangency exponent at the contact point, the authors prove that the solid always reaches the cavity in finite time, but with a non zero velocity for $\alpha < 2$ (real shock case), and with null velocity for $\alpha \geqslant 2$ (smooth landing case). The proof is based on a suitable change of variables transforming the Laplace Neumann problem into a generalized Neumann problem set on a domain containing a horizontal rectangle whose length tends to infinity as the solid approached the cavity.

The paper [14] presents the first positive result on approximate controllability for bilinear Schrödinger equations in presence of mixed spectrum when the interaction term is unbounded.

In [15], Tucsnak, Valein and Wu study the numerical approximation of the solutions of a class of abstract parabolic time optimal control problems. The main results assert that, provided that the target is a closed ball centered at the origin and of positive radius, the optimal time and the optimal controls of the approximate time optimal problems converge to the optimal time and to the optimal controls of the original problem. This is based on a nonsmooth data error estimate for abstract parabolic systems.

A vesicle is an elastic membrane containing a liquid and surrounded by another liquid. Such a vesicle can be found in nature or it can be created in laboratory. They can store and/or transport substances. Modeling vesicles is also a first step in order to study and understand the behavior of more complex cells such as red cells. Their studies are important for many applications, in particular in biological and physiological subjects. Recent papers have been devoted to both experimental studies to the modeling and finally to the mathematical analysis of the obtained models. There are many different models to describe the motion of the membrane and one can for instance optimize the shape in order to minimize the elastic energy of the membrane. Such a problem is tackled in [4] in the 2D case and in [6] in the 3D case. In [4], the optimization is done among convex domains whereas in [6], the authors consider the problem of minimizing the total mean curvature in order to understand the differences between the Helfrich energy and the Willmore energy. Up to now, these models are considered without any fluid.

In [13], San Martin, Takahashi and Tucsnak consider a class of low Reynolds number swimmers, of prolate spheroidal shape, which can be seen as simplified models of ciliated microorganisms. Within this model, the form of the swimmer does not change, the propelling mechanism consisting in tangential displacements of the material points of swimmer's boundary. They obtain the exact controllability of the prolate spheroidal swimmer and the existence of an optimal control problem (in the sense of the efficiency during a stroke). They also provide a method to compute an approximation of the efficiency by using explicit formulas for the Stokes system at the exterior of a prolate spheroid, with some particular tangential velocities at the fluid-solid interface. They analyze the sensitivity of this efficiency with respect to the eccentricity of the considered spheroid and show that for small positive eccentricity, the efficiency of a prolate spheroid is better than the efficiency of a sphere. Finally, they use numerical optimization tools to investigate the dependence of the efficiency on the number of inputs and on the eccentricity of the spheroid.

7.2. Inverse problems for heterogeneous systems

In [7], David Dos Santos Ferreira *et al.* obtain global stability estimates for a potential in a Schrödinger equation on an open bounded set in dimension $n = 3$ from the Dirichlet-to-Neumann map with partial data. This improves previous results where local stability was proved : the region under control was the penumbra

delimited by a source of light outside of the convex hull of the open set. These local estimates provided stability of log-log type corresponding to the uniqueness results in Calderón's inverse problem with partial data proved by Kenig, Sjöstrand and Uhlmann. The corresponding global estimates are proved in all dimensions higher than three. The estimates are based on the construction of solutions of the Schrödinger equation by complex geometrical optics developed in the anisotropic setting by Dos Santos Ferreira, Kenig, Salo and Uhlmann to solve the Calderón problem in certain admissible geometries.

In [20], David Dos Santos Ferreira *et al.* proved uniform L^p resolvent estimates for the stationary damped wave operator. Uniform L^p resolvent estimates for the Laplace operator on a compact smooth Riemannian manifold without boundary were first established by Shen on the Torus, then by Dos Santos Ferreira-Kenig-Salo for general compact manifolds and advanced further by Bourgain-Shao-Sogge-Yao. An alternative proof relying on the techniques of semiclassical Strichartz estimates allows to handle non-self-adjoint perturbations of the Laplacian and embeds very naturally in the semiclassical spectral analysis framework, and applies in the damped wave context.

In [10], Munnier and Ramdani considered the 2D inverse problem of recovering the positions and the velocities of slowly moving small rigid disks in a bounded cavity filled with a perfect fluid. Using an integral formulation, they first derive an asymptotic expansion of the DtN map of the problem as the diameters of the disks tend to zero. Then, combining a suitable choice of exponential type data and the DORT method (french acronym for Diagonalization of the Time Reversal Operator), a reconstruction method for the unknown positions and velocities is proposed. Let us emphasize here that this reconstruction method uses in the context of fluid-structure interaction problems a method which is usually used for waves inverse scattering (the DORT method).

In [24], Munnier and Ramdani proposed a new method to tackle a geometric inverse problem related to Calderón's inverse problem. More precisely, they proposed an explicit reconstruction formula for the cavity inverse problem using conformal mapping. This formula is derived by combining two ingredients: a new factorization result of the DtN map and the so-called generalized Polia-Szegö tensors of the cavity.

In [11], Ramdani, Tucsnak and Valein tackled a state estimation problem for an infinite dimensional system arising in population dynamics (a linear model for age-structured populations with spatial diffusion). Assume the initial state to be unknown, the considered inverse problem is to estimate asymptotically on time the state of the system from a locally distributed observation in both age and space. This is done by designing a Luenberger observer for the system, taking advantage of the particular spectral structure of the problem (the system has a finite number of unstable eigenvalues).

In [12], San Martin, Schwindt and Takahashi consider the geometrical inverse problem consisting in recovering an unknown obstacle in a viscous incompressible fluid by measurements of the Cauchy force on the exterior boundary. They deal with the case where the fluid equations are the non stationary Stokes system and using the enclosure method, they can recover the convex hull of the obstacle and the distance from a point to the obstacle. With the same method, they can obtain the same result in the case of a linear fluid–structure system composed by a rigid body and a viscous incompressible fluid. They also tackle the corresponding nonlinear systems: the Navier–Stokes system and a fluid–structure system with free boundary. Using complex spherical waves, they obtain some partial information on the distance from a point to the obstacle.

7.3. Numerical analysis and simulation of heterogeneous systems

In optics, metamaterials (also known as negative or left-handed materials), have known a growing interest in the last two decades. These artificial composite materials exhibit the property of having negative dielectric permittivity and magnetic permeability in a certain range of frequency, leading hence to materials with negative refractive index and super lens effects. In [5], Bunoiu and Ramdani studied a complex wave system involving such materials. More precisely, they consider a periodic homogenization problem involving two isotropic materials with conductivities of different signs: a classical material and a metamaterial (or negative material). Combining the $\mathbf{T}-$coercivity approach and the unfolding method for homogenization, they prove well-posedness results for the initial and the homogenized problems and obtain a convergence result, provided that the contrast between the two conductivities is large enough (in modulus).

Several results on domain decomposition were obtained in the frame of the collaboration of Xavier Antoine with the team of Christophe Geuzaine (Belgium). The paper [3] deals with a Schwarz-type solver for domain decomposition, the paper [8] proposes a Schwarz-type domain decomposition for high frequency electromagnetism equations, the paper [1] exposes how to use of GPELab to solve Gross-Pitaevskii equations.

The paper [2] deals with domain decomposition for nonlinear Schrödinger equations and the book chapter [16] is focused on the modeling of Bose-Einstein condensates.

TONUS Team

7. New Results

7.1. Particle-in-cell simulations for highly oscillatory Vlasov-Poisson systems

Participants: Edwin Chacon Golcher, Sever Adrian Hirstoaga [correspondent], Mathieu Lutz.

The aim of the following works is to study the dynamics of charged particles under the influence of a strong magnetic field by numerically solving in an efficient way the Vlasov-Poisson and guiding center models.

First, we work on the development of the time-stepping method introduced in [7], [8] in two directions: improve the accuracy of the algorithm and adapt the algorithm for general configuration of magnetic field.

Second, by using appropriate data structures, we implement an efficient (from the memory access point of view) Particle-In-Cell method which enables simulations with a large number of particles. Thus, we present in [13] numerical results for classical one-dimensional Landau damping and two-dimensional Kelvin-Helmholtz test cases. The implementation also relies on a standard hybrid MPI/OpenMP parallelization. Code performance is assessed by the observed speedup and attained memory bandwidth. A convergence result is also illustrated by comparing the numerical solution of a four-dimensional Vlasov-Poisson system against the one for the guiding center model.

7.2. Eulerian simulations of parallel transport in the SOL

Participants: David Coulette, Sever Adrian Hirstoaga [correspondent], Giovanni Manfredi.

We continue to investigate kinetic models for simulating the heat load on the divertor plates during transient events as edge-localised modes (ELMs). Our previous work [36] deals with Vlasov-Poisson equations for two particle species for the dynamics of their transport parallel to the magnetic field. We started to improve this model by adding an equation for the evolution in time of the perpendicular temperatures. These equations take also into account the collisions between species which may play a role over long times. The first numerical results are encouraging, showing different features with respect to the older (simpler) model when computing total particles and energy fluxes on the divertor plates.

7.3. Quasi-neutrality equation in a polar mesh

Participants: Christophe Steiner [correspondent], Michel Mehrenberger, Nicolas Crouseilles, Philippe Helluy.

In this work [21], we are concerned with the numerical resolution of the quasi-neutrality equation arising in plasma physics. A classic method is based on a Padé approximation. Two other methods are proposed in this paper: a high order Padé approximation and a direct method in the space configuration which consists in integrating on the gyrocircles using an interpolation operator. Numerical comparisons are performed with analytical solutions and considering the 4D drift-kinetic model with one Larmor radius. This is a preliminary study; further study in GYSELA is envisioned.

7.4. The Semi-Lagrangian method on curvilinear grids

Participants: Aurore Back, Adnane Hamiaz, Michel Mehrenberger [correspondent], Pierre Navaro, Hocine Sellama, Eric Sonnendrücker.

We study the semi-Lagrangian method on curvilinear grids [18], [9]. The classical backward semi-Lagrangian method preserves constant states but is not mass conservative. Natural reconstruction of the field permits nevertheless to have at least first order in time conservation of mass, even if the spatial error is large. Interpolation is performed with classical cubic splines and also cubic Hermite interpolation with arbitrary reconstruction order of the derivatives. High odd order reconstruction of the derivatives is shown to be a good ersatz of cubic splines which do not behave very well as time step tends to zero. A conservative semi-Lagrangian scheme is then described; here conservation of mass is automatically satisfied and constant states are shown to be preserved up to first order in time.

Semi-Lagrangian guiding center simulations are performed on sinusoidal perturbations of cartesian grids, and on deformed polar grids with different boundary conditions. Key ingredients are: the use of a B-spline finite element solver for the Poisson equation and the classical backward semi-Lagrangian method (BSL) for the advection. We are able to reproduce standard Kelvin-Helmholtz and diocotron instability tests on such grids. When the perturbation leads to a strong distorted mesh, we observe that the solution differs if one takes standard numerical parameters that are used in the cartesian reference case. We can recover good results together with correct mass conservation, by diminishing the time step.

7.5. Solving the Guiding-Center model on a regular hexagonal mesh

Participants: Michel Mehrenberger [correspondent], Laura Mendoza, Charles Prouveur, Eric Sonnendrücker.

This work [11] introduces a Semi-Lagrangian solver for the Vlasov-Poisson equations on a uniform hexagonal mesh. The latter is composed of equilateral triangles, thus it doesn't contain any singularities, unlike polar meshes. We focus on the guiding-center model, for which we need to develop a Poisson solver for the hexagonal mesh in addition to the Vlasov solver. For the interpolation step of the Semi-Lagrangian scheme, a comparison is made between the use of box-splines and of Hermite finite elements. The code will be adapted to more complex models and geometries in the future.

7.6. High-order Hamiltonian splitting for Vlasov-Poisson equations

Participants: Fernando Casas, Nicolas Crouseilles, Erwan Faou, Michel Mehrenberger [correspondent].

In this work [12], we consider the Vlasov-Poisson equation in a Hamiltonian framework and derive new time splitting methods based on the decomposition of the Hamiltonian functional between the kinetic and electric energy. Assuming smoothness of the solutions, we study the order conditions of such methods. It appears that these conditions are of Runge-Kutta-Nyström type. In the one dimensional case, the order conditions can be further simplified, and efficient methods of order 6 with a reduced number of stages can be constructed. In the general case, high-order methods can also be constructed using explicit computations of commutators. Numerical results are performed and show the benefit of using high-order splitting schemes in that context. Complete and self-contained proofs of convergence results and rigorous error estimates are also given.

7.7. Velocity space transformations: collisional case

Participants: Emmanuel Franck, Philippe Helluy [correspondent], Laurent Navoret.

The method of "velocity space transformations" allows to obtain an interesting discretization of the Kinetic equations like Vlasov-Poisson or Vlasov Maxwell equations as has been proved in the works of P. Helluy, L. Navoret and N. Pham. During this year, we have begun to extend this method to the collisional case using the entropy variable to write a general collisional operator. To treat all the regimes (small or large collisional regime), asymptotic preserving schemes (stability and convergence independent of the collisional frequency) have been designed. However, this method admits some numerical difficulties if we use the physical entropy to construct the collisional operator. Now we propose to use modified entropy, which has good numerical properties and gives limit regime close to the real one in the low Mach context. If this new approach gives interesting results, we will study the adaptivity of the velocity discrete basis which would allow to treat the collisional and non-collisional regimes with the same method.

7.8. Preconditioning and implicit solvers

Participants: Emmanuel Franck [correspondent], Philippe Helluy, Matthias Hoelzl, Ahmed Ratnani, Malcolm Roberts, Eric Sonnendrücker, Stefano Serra-Capizzano.

The Viscous-resistive MHD model used to simulate the instabilities is a multi-scale models with fast waves. In this context, it is not possible to use full explicit time schemes. However the classical implicit schemes are not usable directly since the matrices are ill-conditioned. For this reason it is necessary to use a preconditioning method. During this year we have studied a method called "physic based preconditioning" for the wave equations which consists to approximate the solution by suitable smaller and simpler systems. The results are very good. After this, we have extended this method to the Linearized Euler equation. During this new study, we have found additional difficulties which appear in some regimes. Two methods to treat this problem will be tested in 2016. We have also implemented a version of this preconditioning for the reduced MHD models of JOREK. The first results are positive. To finish, we have begun a collaboration with S. Serra-Capizzano to study at the theoretical level the physic based preconditioning and propose new preconditioning for each sub-systems of the Physic-Based PC efficient in all the physics regimes and for an arbitrary order.
We have also developed an implicit solver for the transport equation based on the upwind nature of the DG numerical flux. This solver will be used for solving Vlasov models or fluid models thanks to the Lattice-Boltzmann methodology. We have obtained recently a SPPEXA support (http://www.sppexa.de) in a joint french-german-japanese project.

7.9. Finite element for full-MHD problems

Participants: Emmanuel Franck [correspondent], Eric Sonnendrücker.

This work have begun at the end of 2015. It is organized around a PhD: Mustafa Gaja supervised by E. Sonnendrücker, A. Ratnani and E. Franck at the Max-Planck Institute of Plasma Physic. The aim of this work is to design and study compatible finite element method (finite element method which preserve the DeRham sequence and the inclusion between the functional space) for B-Splines. This method will allow to discretize efficiently the Maxwell equations, the MHD model and some operators as curl-curl or grad-div vectorial operators which appear in the physic-based PC. For now, we have begun to study the finite element discretization of vectorial operators which appears in the linearized Euler equations and in the physic-based PC associated.

7.10. Lagrangian averaged gyrokinetic-waterbag continuum

Participant: Nicolas Besse [correspondent].

In this paper [26], we first present the derivation of the anisotropic Lagrangian averaged gyrowaterbag continuum (LAGWBC-α) equations. The gyrowaterbag (nickname for gyrokinetic-waterbag) continuum can be viewed as a special class of exact weak solution of the gyrokinetic-Vlasov equation, allowing to reduce this latter into an infinite dimensional set of hydrodynamic equations while keeping its kinetic features such as Landau damping. In order to obtain the LAGWBC-α equations from the gyrowaterbag continuum we use an Eulerian variational principle and Lagrangian averaging techniques introduced by Holm, Marsden, Ratiu and Shkoller for the mean motion of ideal incompressible flows, extended to barotropic compressible flows by Bhat and co-workers and some supplementary approximations for the electrical potential fluctuations. Regarding to the original gyrowaterbag continuum, the LAGWBC-α equations show some additional properties and several advantages from the mathematical and physical viewpoints, which make this model a good candidate for describing accurately gyrokinetic turbulence in magnetically confined plasma. In the second part of this paper we prove local-in-time well-posedness of an approximate version of the anisotropic LAGWBC-α equations, that we call the “isotropic” LAGWBC-α equations, by using quasilinear PDE type methods and elliptic regularity estimates for several operators.

7.11. Hamiltonian structure, fluid representation, stability for the Vlasov-Dirac-Benney equation

Participants: Claude Bardos, Nicolas Besse [correspondent].

This contribution [23] is an element of a research program devoted to the analysis of a variant of the Vlasov−Poisson equation that we dubbed the Vlasov−Dirac−Benney equation or in short V−D−B equation. As such it contains both new results and efforts to synthesize previous observations. One of main links between the different issues is the use of the energy of the system. In some cases, such energy becomes a convex functional and allows to extend to the present problem the methods used in the study of conservation laws. Such use of the energy is closely related to the Hamiltonian structure of the problem.

7.12. Semi-classical limit of an infinite dimensional system of nonlinear Schrödinger equations

Participants: Claude Bardos, Nicolas Besse [correspondent].

In this paper [24], we study the semi-classical limit of an infinite dimensional system of coupled nonlinear Schrödinger equations towards exact weak solutions of the Vlasov-Dirac-Benney equation, for initial data with analytical regularity in space. After specifying the right analytic extension of the problem and solutions, the proof relies on a suitable version of the Cauchy-Kowalewski Theorem and energy estimates in Hardy type spaces with convenient analytic norms.

7.13. Aligned interpolation for gyrokinetic Tokamak simulations

Participants: Guillaume Latu, Michel Mehrenberger [correspondent], Maurizio Ottaviani, Eric Sonnendrücker.

This work is devoted to study the aligned interpolation method in semi-Lagrangian codes. The scheme is presented and algorithms used implementing the scheme are given. A theoretical justification of the method is given with convergence estimates in the simplified context of 2D constant advection, assuming stability of the scheme. The stability is here studied numerically, letting the formal proof as an open problem. The solution is successfully applied in the gyrokinetic context: first in a simplified case in cylindrical geometry and then in toroidal geometry. In the first case, the solutions provided by simulations based on the scheme are in accordance with linear dispersion analysis; in the second case, numerical simulations produced by the Gysela code are presented, simulation based on the standard scheme are compared to those based on the new aligned scheme. This work will lead to a project of paper, which will be submitted in 2016.

TOSCA Project-Team

7. New Results

7.1. Probabilistic numerical methods, stochastic modelling and applications

Participants: Mireille Bossy, Nicolas Champagnat, Madalina Deaconu, Coralie Fritsch, Benoît Henry, James Inglis, Antoine Lejay, Oana-Valeria Lupascu, Sylvain Maire, Paolo Pigato, Alexandre Richard, Denis Talay, Etienne Tanré, Denis Villemonais.

7.1.1. Published works and preprints

- M. Bossy with H. Quinteros (UChile) submitted a paper [36] on the strong convergence of the symmetrized Milstein scheme for some CEV-like SDEs.
- M. Bossy and J.-F. Jabir (University of Valparaíso) submitted a paper [35] on the particle approximation for Lagrangian stochastic models with specular boundary condition.
- M. Bossy with N. Maizi (Mines ParisTech) and O. Pourtallier (Inria) published a book chapter [31] on game theory analysis for carbon auction market through electricity market coupling hypothesis.
- M. Bossy, O. Faugeras (Inria Sophia, EPI NEUROMATHCOMP), and D. Talay published a clarification on the well-posedness of the limit equations to the mean-field N-neuron models proposed in [58] and proven the associated propagation of chaos property. They also have completed the modeling issue in [58] 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 [15].
- M. Bossy, N. Champagnat, S. Maire and L. Violeau worked with H. Leman (CMAP, Ecole Polytechnique) and M. Yvinec (Inria Sophia, GEOMETRICA team) on Monte Carlo methods for the linear and non-linear Poisson-Boltzmann equations [14]. These methods are based on walk on spheres algorithm, simulation of diffusion processes driven by their local time, and branching Brownian motion to deal with the nonlinear case.
- Together with M. Baar and A. Bovier (Univ. Bonn), N. Champagnat studied the adaptive dynamics of populations under the assumptions of large population, rare and small mutations [34]. In this work, the three limits are taken simultaneously, contrary to the classical approach, where the limits of large population and rare mutations are taken first, and next the limit of small mutations [59]. We therefore obtain the precise range of assumptions under which these limits can be taken, and provide explicit biological conditions for which our approximation is valid.
- N. Champagnat and C. Fritsch worked with F. Campillo (Inria Sophia-Antipolis, LEMON team) on the links between a branching process and an integro-differential equation of a growth-fragmentation-death model [37]. They proved that the two representations of the model lead to the same criteria of invasion of a population in a given environment.
- Using a new method to compute the expectation of an integral with respect to a random measure, N. Champagnat and B. Henry obtained explicit formulas for the moments of the frequency spectrum in the general branching processes known as Splitting Trees, with neutral mutations and under the infinitely-many alleles model [40]. This allows them to obtain a law of large numbers for the frequency spectrum in the limit of large time.
- N. Champagnat and P.-E. Jabin (Univ. Maryland) improved significantly the description of the functional spaces in the preprint [41], devoted to the study of strong existence and pathwise uniqueness for stochastic differential equations (SDE) with rough coefficients, typically in Sobolev spaces.

- N. Champagnat and D. Villemonais obtained criteria for existence and uniqueness of quasi-stationary distributions (QSD) and Q-processes for general absorbed Markov processes [17]. A QSD 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 criteria ensure exponential convergence of the t-marginal of the process conditioned not to be absorbed at time t, to the QSD and also the exponential ergodicity of the Q-process.
- N. Champagnat and D. Villemonais obtained criteria for existence, uniqueness and exponential convergence in total variation to QSD for general absorbed and killed diffusion processes [43], [42]. For diffusions without killing [43], 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. The criteria obtained for diffusion processes with killing on $[0,\infty)$ [42] combine the last criteria and conditions on the killing time only close to 0, provided ∞ is an entrance boundary.
- N. Champagnat and D. Villemonais obtained criteria for existence, uniqueness and exponential convergence in total variation to QSD for general multi-dimensional birth and death processes in $\mathbb{Z}_+^d$ absorbed at the boundary $\mathbb{Z}_+^d \setminus \mathbb{N}^d$ [44]. These birth and death models are motivated by population dynamics and the criteria obtained assume stronger intra-spectific competition than inter-specific competition. These resuls are the first one for such processes, except for the particular case of branching processes, which can be studied using very specific methods.
- 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. This method is based on 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 brownian hitting time [19].
- M. Deaconu and O. Lupaşcu worked with L. Beznea (Bucharest, Romania) on the probabilistic interpretation of fragmentation phenomena. They constructed a continuous time branching process and characterized its behavior by using new potential theoretical tools [12].
- M. Deaconu, O. Lupaşcu and L. Beznea (Bucharest, Romania) started a new challenging work on the description of rupture phenomena like avalanches, by using fragmentation models. The physical properties of the model are deeply involved in this study. The first results concern a stochastic equation of fragmentation and branching processes related to avalanches [13].
- M. Deaconu, B. Dumortier and E. Vincent are working with the Venathec SAS on the acoustic control of wind farms. They constructed a new approach to control wind farms with a control model based on real-time source separation. They first designed a deterministic algorithm in order to maximize the electric production of the wind farms under the legal acoustic constraints. They showed that it is a non linear knapsack optimization problem and they proposed an efficient solution in that context using a branch and bound algorithm based on continuous relaxation. This work was published at the EWEA 2015 [30].
- In [49], B. Henry showed a central limit theorem for the population counting process of a supercritical Splitting Tree in the limit of large time. Thanks to the results of [40], he also obtained a central limit theorem for the frequency spectrum of Splitting Trees with neutral mutations and under the infinitely-many alleles model.
- S. Herrmann and E. Tanré have proposed a new very efficient algorithm to simulate the first-passage-time of a one-dimensional Brownian motion over a continuous curved boundary [23].
- 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 [21].

- In collaboration with J. Maclaurin (Inria Sophia, EPI NEUROMATHCOMP) J. Inglis has presented a general framework to rigorously study the effect of spatio-temporal noise on traveling waves and stationary patterns. In particular the framework can incorporate versions of the stochastic neural field equation that may exhibit traveling fronts, pulses or stationary patterns. They have formulated a local SDE that describes the position of the stochastic wave up until a discontinuity time, at which point the position of the wave may jump and studied the local stability of this stochastic front and the long-time behavior of the stochastic wave [50].
- A. Lejay has continued his work on the Snapping Out Brownian motion, especially with regard to the simulation issues, with potential application to brain imaging techniques [33], [53].
- A. Lejay has continued his work on the simulation of processes with either discontinuous drift (with Arturo Kohatsu-Higa, Ritsumeikan Universitey and Kazuhiro Yasuda, Hosei University, Japan) [52] or with discontinuous coefficients (with Lionel Lenêtre and Géraldine Pichot, EPI SAGE, Irisa) [54].
- A. Lejay has continued his work on the theory of rough paths, notably with the sensitivity aspects with Laure Coutin (Univ. Toulouse III) [47].
- In collaboration with Ivan Dimov and Jean-Michel Sellier (BAS), S. Maire developed a new Monte Carlo method, called the walk on equations, to solve linear systems of equations [22].
- In collaboration with Xuan Vu, Caroline Chaux-Moulin and Nadege Thirion-Moreau, S. Maire developed a stochastic algorithm to decompose large non-negative tensors with applications in spectroscopy [28].
- In collaboration with Martin Simon, Sylvain Maire developed a variant of the walk on spheres method to deal with diffusion equations appearing in electrical impedance tomography.
- With Giang Nguyen, Sylvain Maire worked on finite differences techniques to deal with many kinds of boundary conditions that are met during the Monte Carlo simulation of diffusions [25].
- A. Richard submitted a paper [56] on the spectral representation of L^2-indexed increment-stationary processes. The main result states that any random field (i.e. process indexed by a multidimensional parameter of a function in L^2) with stationary increments can be written as an integral against a random measure satisfying certain properties. Applications to sample path properties of a multiparameter fractional Brownian motion are exhibited.
- D. Villemonais worked with P. Del Moral (Univ. Sydney) on the conditional ergodicity of time inhomogeneous diffusion processes [48]. 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 α-positive recurrence of birth and death processes. This criterion also provides a non-trivial subset of the domain of attraction for quasi-stationary distributions. Finally, this study leads to 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 [29].
- J. Inglis and D. Talay ended their work on mean-field limits of a stochastic particle system smoothly interacting through threshold hitting-times and applications to neural networks with dendritic component [51].

7.1.2. Other works in progress

- Together with M. Andrade (Univ. Paris 7) and R. Ferrière (ENS Paris and Univ. Arizona), N. Champagnat is working on the phenomenon of clustering in populations structured by space and traits for which local adaptation favors different trait values at different spatial locations. Two methods are used and numerically validated: a Turing instablity method and a Hamilton-Jacobi approximation of the population density. This work is currently being written.

- N. Champagnat and J. Claisse (Ecole Polytechnique) are currently working on the ergodic and infinite horizon controls of discrete population dynamics with almost sure extinction in finite time. This can either correspond to control problems in favor of survival or of extinction, depending on the cost function. They have proved that these two problems are related to the QSD of the processes controled by Markov controls. This work is currently being written.
- N. Champagnat and C. Fritsch worked with F. Campillo (Inria Sophia-Antipolis, LEMON team) on the variations of the principal eigenvalue (resp. the survival probability) of an integro-differential equation (resp. branching process) of growth-fragmentation-death models with respect to an environmental parameter. This work is currently being written.
- N. Champagnat, K. Coulibaly-Pasquier (Univ. Lorraine) and D. Villemonais are currently working on general criteria for existence, uniqueness and exponential convergence in total variation to QSD for multi-dimensional diffusions in a domain absorbed at its boundary. These results both improve and simplify the existing results and methods. This work is currently being written.
- N. Champagnat and D. Villemonais are currently working on extensions of their work [17] to general penalized processes, including time-inhomogeneous Markov processes with absorption. Their method allows to improve significantly the former results of [60], [61]. This work is currently being written.
- N. Champagnat and D. Villemonais are also working on extensions of the criteria of [17] in the form of Foster-Lyapunov criteria allowing to deal with cases where the convergence of conditional distribution to the QSD is not uniform with respect to the initial distribution. This work is currently being written.
- M. Deaconu and S. Herrmann are working on the numerical approach of the time-space Dirichlet problem.
- M. Deaconu, O. Lupaşcu and L. Beznea (Bucharest, Romania) worked on the numerical scheme for the simulation of an avalanche by using the fragmentation model. This work is currently being written.
- M. Deaconu, B. Dumortier and E. Vincent are working with the Venathec SAS on the acoustic control of wind farms. They plan to submit another article to IEEE transaction on sustainable energy soon. Currently they work on handling uncertainties in the model in order to design a stochastic algorithm.
- C. Fritsch worked with F. Campillo (Inria Sophia-Antipolis, LEMON team) and O. Ovaskainen (Univ. Helsinki) about the numerical analysis of the invasion of mutant populations in a chemostat, using branching processes and integro-differential models.
- C. Fritsch started a collaboration with B. Cloez (INRA, Montpellier) on a central limit theorem of mass-structured individual-based chemostat model.
- With P. Pigato, A. Lejay has continued his work on the estimation of parameters of skew diffusions.
- Within the ANESTOC Associate Team, R. Rebolledo (Pontificia Universidad Católica de Chile) and A. Richard initiated a work on the long-term behavior of a class of non-Markovian stochastic differential equations. These equations of Volterra type can be used to model the motion of a particle subject to friction forces in a heat bath, which could also be interesting in neuroscience for ion channels.
- A. Richard and E. Tanré are working with P. Orio (CINV, Chile) on the measurement of long-range dependence in series of neuronal spikes, and are providing a leaky integrate-and-fire model with fractional noise to include this effect. So far, we produced numerical experiments that confirm the existence of memory in our model, and A. Richard and E. Tanré now work on the convergence of the statistical estimator that measures this phenomenon.

- A. Richard, E. Tanré and S. Torres (Universidad de Valparaíso, Chile) are working on the definition of a skew fractional Brownian motion. The skew Brownian motion (sBm) is a process which is partly reflected when it reaches the horizontal line, making it a natural model for the motion of a particle crossing media with different diffusion properties. The fractional sBm is a modification of this process to incorporate long-range dependences. So far, we constructed a reflected fractional Brownian motion, and we are now investigating its approximation by a discrete-time process.
- During her internship supervised by E. Tanré and Romain Veltz (NEUROMATHCOMP team), Roberta Evangelista worked on "A stochastic model of gamma phase modulated orientation selectivity". Neurons in primary visual cortex (V1) are known to be highly selective for stimulus orientation. Recent experimental evidence has shown that, in awake monkeys, the orientation selectivity of V1 neurons is modulated by gamma oscillations. In particular, neurons' firing rate in response to the preferred orientation changes as a function of the gamma phase of spiking. The effect is drastically reduced for non-preferred orientations. We have introduced a stochastic model of a network of orientation-dependent excitatory and inhibitory spiking neurons. We have found conditions on the parameters such that the solutions of the mathematical model reproduce the experimental behavior.
- During his internship supervised by E. Tanré and Romain Veltz (NEUROMATHCOMP team), Quentin Cormier studies numerically and theoretically a model of spiking neuron in interaction with plasticity. The synaptic weights evolve according to biological law of plasticity. We study the existence of separable time scales. During his internship, Quentin Cormier also develop a numerical code to simulate large networks of neurons evolving according to this dynamics.
- C. Graham (Ecole Polytechnique) and D. Talay have written a large part of the second volume of their series on Mathematical Foundation of Stochastic Simulation.

7.2. Financial Mathematics

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

7.2.1. Published works and preprints

- In collaboration with Jerome Lelong and Christophe Deluigi, Sylvain Maire built a new algorithm for the automatic integration and approximation of irregular functions [18]. This algorithm is tested numerically on the pricing of multidimensional exotic options.
- 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 is in revision [55].

7.2.2. Other works in progress

- K. Salhi works on partial hedging of options in an incomplete market, under constraints on the initial capital of the investor and assuming that the stock price is described by a Lévy process. In this case, perfect hedging is no more possible and we talk about partial hedging and minimization of risk. K. Salhi focuses on the Conditional Value-at-Risk minimization. He tries to give a numerical approximation to the solution in this context.
- 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.

VEGAS Project-Team

6. New Results

6.1. Robustness issues in computational geometry

Participants: Olivier Devillers, Monique Teillaud.

6.1.1. Qualitative Symbolic Perturbation: a new geometry-based perturbation framework

In a classical Symbolic Perturbation scheme, degeneracies are handled by substituting some polynomials in ϵ to the input of a predicate. Instead of a single perturbation, we propose to use a sequence of (simpler) perturbations. Moreover, we look at their effects geometrically instead of algebraically; this allows us to tackle cases that were not tractable with the classical algebraic approach [25].

This work was done in collaboration with Menelaos Karavelas (Univ. of Crete).

6.2. Probabilistic analysis of geometric data structures and algorithms

Participant: Olivier Devillers.

6.2.1. The worst visibility walk in a random Delaunay triangulation is $O(\sqrt{n})$

We show that the memoryless routing algorithms Greedy Walk, Compass Walk, and all variants of visibility walk based on orientation predicates are asymptotically optimal in the average case on the Delaunay triangulation. More specifically, we consider the Delaunay triangulation of an unbounded Poisson point process of unit rate and demonstrate that the worst-case path between any two vertices inside a domain of area n has a number of steps that is not asymptotically more than the shortest path which exists between those two vertices with probability converging to one (as long as the vertices are sufficiently far apart.) As a corollary, it follows that the worst-case path has $O(\sqrt{n})$ steps in the limiting case, under the same conditions. Our results have applications in routing in mobile networks and also settle a long-standing conjecture in point location using walking algorithms. Our proofs use techniques from percolation theory and stochastic geometry [24].

This work was done in collaboration with Ross Hemsley (formerly in Inria Geometrica).

6.2.2. Smooth analysis of convex hulls

We establish an upper bound on the smoothed complexity of convex hulls in $\mathbb{R}^d$ under uniform Euclidean (ℓ^2) noise. Specifically, let $\{p_1^*, p_2^*, ..., p_n^*\}$ be an arbitrary set of n points in the unit ball in $\mathbb{R}^d$ and let $p_i = p_i^* + x_i$, where $x_1, x_2, ..., x_n$ are chosen independently from the unit ball of radius δ. We show that the expected complexity, measured as the number of faces of all dimensions, of the convex hull of $\{p_1, p_2, ..., p_n\}$ is $O\left(n^{2-\frac{4}{d+1}}\,(1+1/\delta)^{d-1}\right)$; the magnitude δ of the noise may vary with n. For $d = 2$ this bound improves to $O\left(n^{\frac{2}{3}}(1+\delta^{-\frac{2}{3}})\right)$.

We also analyze the expected complexity of the convex hull of ℓ^2 and Gaussian perturbations of a nice sample of a sphere, giving a lower-bound for the smoothed complexity. We identify the different regimes in terms of the scale, as a function of n, and show that as the magnitude of the noise increases, that complexity varies monotonically for Gaussian noise but non-monotonically for ℓ^2 noise [13].

This work was done in collaboration with Xavier Goaoc (Univ. Marne la Vallée), Marc Glisse and Remy Thomasse (Inria Geometrica).

6.3. Non-linear computational geometry

Participants: Guillaume Moroz, Sylvain Lazard, Marc Pouget, Laurent Dupont, Rémi Imbach.

6.3.1. Solving bivariate systems and topology of plane algebraic curves

In the context of our algorithm Isotop for computing the topology of plane algebraic curves (see Section 5.1), we work on the problem of solving a system of two bivariate polynomials. We are interested in certified numerical approximations or, more precisely, isolating boxes of the solutions. But we are also interested in computing, as intermediate symbolic objects, a Rational Univariate Representation (RUR) that is, roughly speaking, a univariate polynomial and two rational functions that map the roots of the univariate polynomial to the two coordinates of the solutions of the system. RURs are relevant symbolic objects because they allow to turn many queries on the system into queries on univariate polynomials. However, such representations require the computation of a separating form for the system, that is a linear combination of the variables that takes different values when evaluated at the distinct solutions of the system.

We published this year [11] results showing that, given two polynomials of degree at most d with integer coefficients of bitsize at most τ, (i) a separating form, (ii) the associated RUR, and (iii) isolating boxes of the solutions can be computed in, respectively, $\widetilde{O}_B(d^8 + d^7\tau)$, $\widetilde{O}_B(d^7 + d^6\tau)$ and $\widetilde{O}_B(d^8 + d^7\tau)$ bit operations in the worst case, where $\widetilde{O}$ refers to the complexity where polylogarithmic factors are omitted and O_B refers to the bit complexity.

However, during the publishing process, we have substentially improved these results. We have presented for these three sub-problems new algorithms that have worst-case bit complexity $\widetilde{O}_B(d^6 + d^5\tau)$. We have also presented probabilistic Las Vegas variants of our two first algorithms, which have expected bit complexity $\widetilde{O}_B(d^5 + d^4\tau)$. We also show that it is likely difficult to improve these complexities as it would essentially require to improve bounds on other fundamental problems (e.g., computing resultants, checking squarefreeness and root isolation of univariate polynomials) that have hold for decades.

This work was done in collaboration with Yacine Bouzidi (Inria Saclay), Michael Sagraloff (MPII Sarrebruken, Germany) and Fabrice Rouillier (Inria Rocquencourt). It is published in the research report [22] and submitted to a journal.

A key ingredient of the above work is the classical triangular decomposition algorithm via subresultants [31] on which we obtain two results of independent interest. First, we improved by a factor d the state-of-the-art worst-case bit complexity of this algorithm [22]. One constraint on this algorithm is that it requires that the curves defined by the input polynomials have no common vertical asymptotes. Our second result is a generalization of this algorithm, which removes that restriction while preserving the same worst-case bit complexity of $\widetilde{O}_B(d^6 + d^5\tau)$. Furthermore, we actually present a refined bit complexity in $\widetilde{O}_B(d_x^3 d_y^3 + (d_x^2 d_y^3 + d_x d_y^4)\tau)$ where d_x and d_y bound the degrees of the input polynomials in x and y, respectively. We also prove that the total bitsize of the decomposition is in $\widetilde{O}((d_x^2 d_y^3 + d_x d_y^4)\tau)$.

This work was done in collaboration with Fabrice Rouillier (Inria Rocquencourt). It is published in the research report [27] and submitted to a journal.

6.3.2. Numeric and Certified Isolation of the Singularities of the Projection of a Smooth Space Curve

Let a smooth real analytic curve embedded in $\mathbb{R}^3$ be defined as the solution of real analytic equations of the form $P(x, y, z) = Q(x, y, z) = 0$ or $P(x, y, z) = \frac{\partial P}{\partial z} = 0$. Our main objective is to describe its projection $\mathcal{C}$ onto the (x, y)-plane. In general, the curve $\mathcal{C}$ is not a regular submanifold of $\mathbb{R}^2$ and describing it requires to isolate the points of its singularity locus Σ. After describing the types of singularities that can arise under some assumptions on P and Q, we present a new method to isolate the points of Σ. We experimented our method on pairs of independent random polynomials (P, Q) and on pairs of random polynomials of the form $(P, \frac{\partial P}{\partial z})$ and got promising results [14].

On the same topic but with a different approach, we improved our research report [26] by including experimental data using SubdivisionSolver (see Section 5.2) and submitted this work to a journal.

6.3.3. Mechanical design of parallel robots

In collaboration with F. Rouillier, D. Chablat and our PhD student Ranjan Jha, we analyzed the singularities and the workspace of different families of robots.

The first result is a certified description of the workspace and the singularities of a Delta like family robot [16]. Workspace and joint space analysis are essential steps in describing the task and designing the control loop of the robot, respectively. This paper presents the descriptive analysis of a family of delta-like parallel robots by using algebraic tools to induce an estimation about the complexity in representing the singularities in the workspace and the joint space. A Gröbner based elimination is used to compute the singularities of the manipulator and a Cylindrical Algebraic Decomposition algorithm is used to study the workspace and the joint space. From these algebraic objects, we propose some certified three dimensional plotting describing the shape of workspace and of the joint space which will help the engineers or researchers to decide the most suited configuration of the manipulator they should use for a given task. Also, the different parameters associated with the complexity of the serial and parallel singularities are tabulated, which further enhance the selection of the different configurations of the manipulator by comparing the complexity of the singularity equations.

The second result is an algebraic method to check the singularity-free paths for parallel robots [15]. Trajectory planning is a critical step while programming the parallel manipulators in a robotic cell. The main problem arises when there exists a singular configuration between the two poses of the end-effectors while discretizing the path with a classical approach. This paper presents an algebraic method to check the feasibility of any given trajectories in the workspace. The solutions of the polynomial equations associated with the trajectories are projected in the joint space using Gröbner based elimination methods and the remaining equations are expressed in a parametric form where the articular variables are functions of time t unlike any numerical or discretization method. These formal computations allow to write the Jacobian of the manipulator as a function of time and to check if its determinant can vanish between two poses. Another benefit of this approach is to use a largest workspace with a more complex shape than a cube, cylinder or sphere. For the Orthoglide, a three degrees of freedom parallel robot, three different trajectories are used to illustrate this method.

6.3.4. Reflection through quadric mirror surfaces

We addressed the problem of finding the reflection point on quadric mirror surfaces, especially ellipsoid, paraboloid or hyperboloid of two sheets, of a light ray emanating from a 3D point source P_1 and going through another 3D point P_2, the camera center of projection. We previously proposed a new algorithm for this problem, using a characterization of the reflection point as the tangential intersection point between the mirror and an ellipsoid with foci P_1 and P_2. The computation of this tangential intersection point is based on our algorithm for the computation of the intersection of quadrics [5], [28]. Unfortunately, our new algorithm is not yet efficient in practice. This year, we made several improvements on this algorithm. First, we decreased from 11 to 4 the degree of a critical polynomial that we need to solve and whose solutions induce the coefficients in some other polynomials appearing later in the computations. Second, we implemented Descarte's algorithm for isolating the real roots of univariate polynomials in the case where the coefficients belong to extensions of $\mathbb{Q}$ generated by at most two square roots. Furthermore, we are currently implementing the generalization of that algorithm when the coefficients belong to extensions of $\mathbb{Q}$ generated by one root of an arbitrary polynomial. These undergoing improvements should allow us to compute more directly the wanted reflexion point, thus avoiding problematic approximations and making the overall algorithm tractable.

VERIDIS Project-Team

7. New Results

7.1. Automated and Interactive Theorem Proving

Participants: Gabor Alági, Haniel Barbosa, Jasmin Christian Blanchette, Martin Bromberger, Simon Cruanes, Pablo Dobal, Mathias Fleury, Pascal Fontaine, Maximilian Jaroschek, Marek Košta, Stephan Merz, Martin Riener, Thomas Sturm, Hernán Pablo Vanzetto, Uwe Waldmann, Daniel Wand, Christoph Weidenbach.

7.1.1. Combination of Satisfiability Procedures

Joint work with Christophe Ringeissen from the CASSIS project-team at Inria Nancy – Grand Est, and Paula Chocron, a student at the University of Buenos Aires.

A satisfiability problem is often expressed in a combination of theories, and a natural approach consists in solving the problem by combining the satisfiability procedures available for the component theories. This is the purpose of the combination method introduced by Nelson and Oppen. However, in its initial presentation, the Nelson-Oppen combination method requires the theories to be signature-disjoint and stably infinite (to ensure the existence of an infinite model). The design of a generic combination method for non-disjoint unions of theories is clearly a hard task, but it is worth exploring simple non-disjoint combinations that appear frequently in verification. An example is the case of shared sets, where sets are represented by unary predicates. Another example is the case of bridging functions between data structures and a target theory (e.g., a fragment of arithmetic).

We defined [24] a sound and complete combination procedure à la Nelson-Oppen for the theory of absolutely free data structures (including lists and trees) connected to another theory via bridging functions. This combination procedure has also been refined for standard interpretations. The resulting theory has a nice politeness property, enabling combinations with arbitrary decidable theories of elements. We also investigated [25] other theories amenable to similar combinations: this class includes the theory of equality, the theory of absolutely free data structures, and all the theories in between.

7.1.2. Adapting Real Quantifier Elimination Methods for Conflict Set Computation

The satisfiability problem in real closed fields is decidable. In the context of satisfiability modulo theories, the problem restricted to conjunctive sets of literals, that is, sets of polynomial constraints, is of particular importance. One of the central problems is the computation of good explanations of the unsatisfiability of such sets, i.e. obtaining a small subset of the input constraints whose conjunction is already unsatisfiable. We have adapted two commonly used real quantifier elimination methods, cylindrical algebraic decomposition and virtual substitution, to provide such conflict sets and demonstrate the performance of our method in practice [27].

7.1.3. Codatatypes and Corecursion

Joint work with Andrei Popescu and Dmitriy Traytel (Technische Universität München) and Andrew Reynolds (EPFL).

Datatypes and codatatypes are useful for specifying and reasoning about (possibly infinite) computational processes. The Isabelle/HOL proof assistant is being extended with flexible and convenient support for (co)datatypes and (co)recursive functions on them. We extended the emergent framework for (co)codatatypes with automatic generation of nonemptiness witnesses [22], nonemptiness being a proviso for introducing types in many logics, including Isabelle's higher-order logic. As a theoretical step towards a definitional mechanism in Isabelle, we formalized a framework for defining corecursive functions safely, based on corecursion up-to and relational parametricity [21]. The end product is a general corecursor that allows corecursive (and even recursive) calls under "friendly" operations—an improvement over the inflexible syntactic criteria of systems such as Agda and Coq.

In a related line of work, we improved the automation of the SMT solver CVC4 by designing, implementing, and evaluating a combined decision procedure for datatypes and codatatypes [31]. The procedure decides universal problems and is composable via the Nelson–Oppen method, as implemented in SMT solvers. The decision procedure for (co)datatypes is useful both for proving and for model finding. We have commenced work on a higher-order model finder based on CVC4, called Nunchaku, that relies heavily on the decision procedure.

7.1.4. Analysis and Generation of Structured Proofs

Joint work with Sascha Böhme (QAware GmbH), Maximilian Haslbeck and Tobias Nipkow (Technische Universität München), Daniel Matichuk (NICTA), and Steffen J. Smolka (Cornell University).

Isabelle/HOL is probably the most widely used proof assistant besides Coq. The Archive of Formal Proofs is a vast collection of computer-checked proofs developed using Isabelle, containing nearly 65 000 lemmas. We performed an in-depth analysis of the archive, looking at various properties of the proof developments, including size, dependencies, and proof style [18]. This give some insights into the nature of formal proofs.

In the context of the Sledgehammer bridge between automatic theorem provers and proof assistants, we designed a translation of machine-generated proofs into (semi-)intelligible Isabelle proofs that users can simply insert into their proof texts to discharge proof obligations [16]. While the output is designed for certifying the machine-generated proofs, it also has a pedagogical value: Unlike Isabelle's automatic tactics, which are black boxes, the proofs delivered by Sledgehammer can be inspected and understood. The direct proofs also form a good basis for manual tuning.

7.1.5. Encoding Set-Theoretic Formulas in Many-Sorted First-Order Logic

TLA$^+$ is a language for the formal specification of systems and algorithms whose first-order kernel is a variant of untyped Zermelo-Fraenkel set theory. Typical proof obligations that arise during the verification of TLA$^+$ specifications mix reasoning about sets, functions, arithmetic, tuples, and records. Encoding such formulas in the input languages of standard first-order provers (SMT solvers or superposition-based provers for first-order logic) is paramount for obtaining satisfactory levels of automation. For set theory, the basic idea is to represent membership as an uninterpreted predicate for the backend provers, and to reduce set-theoretic expressions to this basic predicate. This is not straightforward for formulas involving set comprehension or for proofs that rely on extensionality for inferring equality of sets. Moreover, a full development of set-theoretic expressions may lead to large formulas that can overwhelm backend provers. We describe a technique that transforms set-theoretic formulas by successively applying rewriting and abstraction until a fixed point is reached. The technique is extended to handling functions, records, and tuples, and it is the kernel of the SMT backend of the TLA$^+$ proof system (section 6.3). A paper describing our technique has been presented at the SETS workshop 2015 [46].

Although the approach was mainly intended to support proofs, we have also started work on adapting it for constructing models of formulas in set theory. Being able to construct (counter-)models can help users understand why proof attempts fail. During his internship, Glen Mével from ENS Rennes designed translation rules for a core fragment of TLA$^+$ set theory. He validated them by using the finite model finding functionality of the SMT solver CVC4 for constructing models, with encouraging preliminary results.

7.1.6. Linear Constraints in Integer Arithmetic

We have investigated linear integer constraint solving. Many existing algorithms rely on solving the rational relaxation and transferring the results to an integer branch and bound approach. This algorithm eventually terminates due to the well-known a priori exponential bounds of an integer solution. De Moura and Jovanović proposed the first model-driven terminating algorithm where the termination relies on the structure of the problem itself but not on a priori bounds [62]. However, the algorithm contained some bugs, in particular it did not terminate. We fixed the bugs by introducing the notion of Weak Cooper elimination. Termination requires adding more rules to the algorithm and refining some existing ones [23].

7.1.7. Decidability of First-Order Clause Sets

Recursion is a necessary source for first-order undecidability of clause sets. If there are no cyclic, i.e., recursive definitions of predicates in such a clause set, (ordered) resolution terminates, showing decidability. In this work we present the first characterization of recursive clause sets enabling non-constant function symbols and depth increasing clauses but still preserving decidability. For this class called BDI (Bounded Depth Increase) we present a specialized superposition calculus. This work was published in the Journal of Logic and Computation [63]. Recursive clause sets also become decidable in the context of finite domain axioms. For this case we developped a new calculus that incorporates explicit partial model assumptions guiding the search [19].

7.1.8. Building Blocks for Automated Reasoning

There are automated reasoning building blocks shared between today's prime calculi for propositional logic (CDCL), propositional logic modulo theories (CDCL(T)), and first-order logic with equality (superposition). Underlying all calculi is a partial model assumption guiding inferences that are not redundant. Deciding the abstract redundancy notion is basically as difficult as the overall satisfiability problem for the respective logic, but for well-chosen partial model assumptions inferences can be guaranteed to be non-redundant at much lower cost. For example, for SAT it is possible to computed inferences in linear time [40] that are guaranteed to be non-redundant.

7.1.9. Beagle – A Hierarchic Superposition Prover

Joint work with Peter Baumgartner and Joshua Bax from NICTA, Canberra, Australia.

Hierarchic superposition is a calculus for automated reasoning in first-order logic extended by some background theory. In [20] we describe an implementation of hierarchic superposition within the Beagle theorem prover, and report on Beagle's performance on the TPTP problem library. Currently implemented background theories are linear integer and linear rational arithmetic. Beagle features new simplification rules for theory reasoning and implements calculus improvements like weak abstraction and determining (un)satisfiability w.r.t. quantification over finite integer domains.

7.1.10. Modal Tableau Systems with Blocking and Congruence Closure

Joint work with Renate A. Schmidt from the University of Manchester, UK.

For many common modal and description logics there are ways to avoid the explicit use of equality in a tableau calculus. For more expressive logics, e.g., with nominals as in hybrid modal logics and description logics, avoiding equality becomes harder, though, and for modal logics where the binary relations satisfy frame conditions expressible as first-order formulae with equality, explicit handling of equations is the easiest and sometimes the only known way to perform equality reasoning. In [32] we describe an approach for efficient handling of equality in tableau systems. We combine Smullyan-style tableaux with a congruence closure algorithm, and demonstrate that this method also permits the use of common blocking restrictions such as ancestor blocking.

7.1.11. Subtropical Real Root Finding

This research is motivated by a series of studies of Hopf bifurcations [60], [59] for reaction systems in chemistry and gene regulatory networks in systems biology. The relevant systems are originally given in terms of ordinary differential equations, for which Hopf bifurcations can be described algebraically [54], [74], [58], [57], typically resulting in one very large multivariate polynomial equation $f = 0$ subject to a few much simpler polynomial side conditions $g_1 > 0$, ..., $g_n > 0$. For these algebraic systems one is interested in feasibility over the reals and, in the positive case, in at least one feasible point. It turns out that, generally, scientifically meaningful information can be obtained already by checking only the feasibility of $f = 0$, which is the focus of this project. For further details on the motivating problems, we refer to our earlier publications [72], [71], [56], [55].

With one of our models, viz. *Mitogen-activated protein kinase (MAPK)*, we obtain and solve polynomials of considerable size. Our currently largest instance `mapke5e6` contains 863,438 monomials in 10 variables. One of the variables occurs with degree 12, all other variables occur with degree 5. Such problem sizes are clearly beyond the scope of classical methods in symbolic computation. To give an impression, the size of an input file with `mapke5e6` in infix notation is 30 MB large. LaTeX-formatted printing of `mapke5e6` would fill more than 5000 pages in this report.

We have developed an incomplete but terminating algorithm for finding real roots of large multivariate polynomials [33]. The principal idea is to take an abstract view of the polynomial as the set of its exponent vectors supplemented with sign information on the corresponding coefficients. To that extent, out approach is quite similar to tropical algebraic geometry [73]. However, after our abstraction we do not consider tropical varieties but employ linear programming to determine certain suitable points in the Newton polytope, which somewhat resembles successful approaches to sum-of-square decompositions [67].

We have implemented our approach in Reduce [61] using direct function calls to the dynamic library of the LP solver Gurobi [48]. In practical computations on several hundred examples originating from the work within an interdisciplinary research group our method has failed due to its incompleteness in only 10 percent of the cases. The longest computation time observed was around 16 s for the above-mentioned `mapke5e6`. With a publication of our computational results in a physics journal, our research had considerable impact beyond computer science [17].

7.1.12. Standard Answers for Virtual Substitution

Joint work with A. Dolzmann from Leibniz-Zentrum für Informatik in Saarbrücken, Germany.

We consider existential problems over the reals. Extended quantifier elimination generalizes the concept of regular quantifier elimination by additionally providing answers which are descriptions of possible assignments for the quantified variables. Implementations of extended quantifier elimination via virtual substitution have been successfully applied to various problems in science and engineering.

So far, the answers produced by these implementations included infinitesimal and infinite numbers, which are hard to interpret in practice. This has been explicitly criticized in the scientific literature. In our article [44], we introduce a complete post-processing procedure to convert, for fixed values of parameters, all answers into standard real numbers. We furthermore demonstrate the successful application of an implementation of our method within Redlog to a number of extended quantifier elimination problems from the scientific literature including computational geometry, motion planning, bifurcation analysis for models of genetic circuits and for mass action, and sizing of electrical networks.

7.1.13. A Generalized Framework for Virtual Substitution

We generalize the framework of virtual substitution for real quantifier elimination to arbitrary but bounded degrees [45]. We make explicit the representation of test points in elimination sets using roots of parametric univariate polynomials described by Thom codes. Our approach follows an early suggestion by Weispfenning, which has never been carried out explicitly.

We give necessary and sufficient conditions for the existence of a root with a given test point representation. These conditions are used to rule out redundant test points. Our encoding allows us to distinguish between test points that represent lower bounds and test points representing upper bounds of a satisfying interval for a given input formula. Furthermore, we show how to reduce the size of elimination sets by generalizing a well-known idea from linear virtual substitution, namely to consider only test points representing lower bounds of a satisfying interval.

Our framework relies on some external algorithm $\mathcal{A}$, which is used to eliminate a single existential quantifier from a finite set of generic formulas. The existence of $\mathcal{A}$ is guaranteed by the fact that $\mathbb{R}$ admits quantifier elimination. We briefly refer to experiments which compared the performance of our framework—when Cylindrical Algebraic Decomposition is used as the external algorithm—to other quantifier elimination algorithms. Unfortunately, our approach is not yet able to compete with other state-of-the-art quantifier

elimination algorithms. However, currently ongoing research suggests the possibility for drastic improvements in practice. Investigating this is left for future work.

7.2. Formal Methods for Developing Algorithms and Systems

Participants: Manamiary Andriamiarina, Noran Azmy, Gabriel Corona, Marie Duflot-Kremer, Marion Guthmuller, Souad Kherroubi, Dominique Méry, Stephan Merz, Martin Quinson, Christoph Weidenbach.

7.2.1. Incremental Development of Distributed Algorithms

Joint work with Mike Poppleton, University of Southampton, UK, and with Neeraj Kumar Singh from the Department of Computing and Software, McMaster University, Hamilton, Canada.

The development of distributed algorithms and, more generally, of distributed systems, is a complex, delicate, and challenging process. The approach based on refinement helps to gain formality by using a proof assistant, and proposes to apply a design methodology that starts from the most abstract model and leads, in an incremental way, to the most concrete model, for producing a distributed solution. Our work helps formalizing pre-existing algorithms, developing new algorithms, as well as developing models for distributed systems.

More concretely, we aim at an integration of the correct-by-construction refinement-based approach for distributed algorithms. Our main results during 2015 are:

- An integrated formal method for verification of liveness properties in distributed systems is introduced [43], and the verification of a self-stabilizing leader election protocol for population protocols illustrates the proposed methodology.
- Manamiary Andriamiarina completed his PhD, illustrating a method for developing distributed algorithms based on a combination of Event-B and fragment of temporal logic TLA.
- The methodology has been applied to take into account resilience in distributed systems. We describe a fully mechanized proof of correctness of self-☆ systems [42] along with an interesting case study related to P2P-based self-healing protocols.

7.2.2. Modeling Medical Devices

Joint work with Neeraj Kumar Singh from the Department of Computing and Software, McMaster University, Hamilton, Canada.

Formal modeling techniques and tools have attained sufficient maturity for formalizing highly critical systems in view of improving their quality and reliability, and the development of such methods has attracted the interest of industrial partners and academic research institutions. Building high quality and zero-defect medical software-based devices is a particular domain where formal modelling techniques can be applied effectively. Medical devices are very prone to showing unexpected system behaviour in operation when traditional methods are used for system testing. Device-related problems have been responsible for a large number of serious injuries. Officials of the US Food and Drug Administration (FDA) found that many deaths and injuries related to these devices are caused by flaws in product design and engineering. Cardiac pacemakers and implantable cardioverter-defibrillators (ICDs) are among the most critical medical devices and require closed-loop modelling (integrated system and environment modelling) for verification purposes before obtaining a certificate from the certification bodies.

Clinical guidelines systematically assist practitioners in providing appropriate health care in specific clinical circumstances. Today, a significant number of guidelines and protocols are lacking in quality. Indeed, ambiguity and incompleteness are likely anomalies in medical practice. The analysis of guidelines using formal methods is a promising approach for improving them.

Analyzing requirements is a major challenge in the area of safety- critical software, where the quality of requirements is an important issue for building a dependable critical system. Many projects fail due to lack of understanding of user needs, missing functional and non-functional system requirements, inadequate methods and tools, and inconsistent system specifications. This often results from the poor quality of system requirements. Based on our experience and knowledge, an environment model has been recognized to be a promising approach to support the requirements engineering to validate a system specification. It is crucial to get an approval and feedback at an early stage of the system development to guarantee the completeness and correctness of the requirements. In [29], we propose a method for analyzing the system requirements using closed-loop modelling technique. The closed-loop model in an integration of system model and environment model, where both the system and environment models are formalized using formal techniques. Formal verification of this closed-loop model helps to identify hidden or missing system requirements and peculiar behaviours, which are not covered earlier during requirements elicitation process. Moreover, the environment model assists in the construction, clarification, and validation of a given system requirements.

7.2.3. Verification of the Pastry routing protocol

In his PhD thesis at Saarbrücken University in 2013, Tianxiang Lu had studied the routing protocol of the Pastry algorithm [69] for maintaining a distributed hash table in a peer-to-peer network. He had discovered several problems in the published algorithm and proposed a modification of the protocol, together with a correctness proof under the hypothesis that no node ever disconnects. The proof had been checked using TLAPS, but it made many assumptions on the underlying data structures that were left unchecked. In particular, support for (modulus) arithmetic in TLAPS was too weak at the time when the proof was written.

As part of her PhD thesis, Noran Azmy studied the assumptions that had been left unproved, and found that several of them were not valid. As a consequence, she was able to find a counter-example to one of the invariants underlying the correctness proof. She corrected the assumptions, proved all of the ones that were needed for the proof using the current version of TLAPS, and also introduced higher-level abstractions that allowed her to rewrite the specification and the correctness proof of the routing protocol in a way that avoids low-level arithmetic reasoning throughout the proof. As a result, she obtained a complete machine-checked proof of Lu's variant of Pastry, still under the assumption that no node leaves the network. A paper describing the result is being submitted.

7.2.4. Proof of Determinacy of PharOS

Joint work with Selma Azaiez and Matthieu Lemerre (CEA Saclay), and Damien Doligez (Inria Paris).

The main contribution of our team to the ADN4SE project (section 8.1), in cooperation with colleagues from CEA, was to write a high-level specification of the real-time operating system PharOS in the TLA$^+$ language, and to prove a determinacy property of the model using TLAPS. Roughly speaking, determinacy means that the sequence of local states of each process during a computation does not depend on the order in which processes are scheduled, as long as there are no missed deadlines. This property simplifies the analysis and verification of programs that run on PharOS. It relies on the fact that every instruction is associated with a time window of execution, and a message can only be received by an instruction if the earliest possible execution time of that instruction is later than the latest possible execution time of the instruction sending the message. The model and proof are based on Lemerre et al. [65]. However, the underlying assumptions are made fully explicit in the formal model, and the proof is carried out in assertional rather than behavioral style. The proof was completed in 2015, and a paper describing the result is being submitted.

7.2.5. Formal Development of Component Semantics in B

Joint work with David Déharbe of Universidade Federal do Rio Grande de Norte (UFRN), Brazil.

We develop a formal model in Isabelle/HOL of the behavioral semantics of software components designed with the B method. We formalize semantic objects, based on labeled transition systems, notions of internal and externally visible behavior, and simulation. In particular, we study a variant of simulation that corresponds to refinement in the B method. We also formally represent the composition of components in the B method.

This work was presented at an invited talk at FACS 2015 in Rio de Janeiro, and an article will be published in LNCS.

7.2.6. Analysis of Distributed Legacy Applications

SimGrid is a toolkit for the study of Large-Scale Distributed Systems. It contains both a simulator with sound and validated performance models for the network, CPUs, and disks, but also an explicit model checker exploring all possible message interleavings in the application, and searching for states violating some properties specified by the user.

We recently added the ability to assess liveness properties over arbitrary and legacy codes, thanks to a system-level introspection tool that provides a detailed view of the running application to the model checker. This can for example be leveraged to verify both safety and liveness properties, on arbitrary MPI code written in C, C++ or Fortran. This work has been published in the Workshop on Formal Approaches to Parallel and Distributed Systems (4PAD) [26], while the full details appear in Guthmuller's PhD thesis [12].

In his master project, Gabriel Rodrigues Santos investigated the feasibility of implementing algorithms for statistical model checking within SimGrid. The basic idea is to sample sufficiently many executions of a program, based on probabilistic parameters associated with the execution platform, for quantifying correctness and reliability properties. By construction, the answers obtained in this way are not exact, but their imprecision can be bounded by an interval of confidence. The results are very encouraging, and we intend to pursue this approach in further work.

7.2.7. Evaluating and Verifying Probabilistic Systems

Joint work with colleagues at ENS Cachan, University Paris Est Créteil, and Ecole Centrale Paris.

Since its introduction in the 1980s, model checking has become a prominent technique for the verification of complex systems. The aim was to decide whether or not a system fulfills its specification. With the rise of probabilistic systems, new techniques have been designed to verify this new type of systems, and appropriate logics have been proposed to describe more subtle properties to be verified. However, some characteristics of such systems fall outside the scope of model checking. In particular, it is often of interest not to decide wether a property is satisfied but how well the system performs with respect to a certain measure. We have designed a statistical tool for tackling both performance and verification issues. Following several conference talks, two journal papers have been published. The first one [14] presents the approach in details together with illustrative applications to flexible manufacturing systems, and to the study of a biological mechanism knwon as circadian clock. The second one [15] focuses on biological applications, and more precisely the use of statistical model checking to detect and measure several indicators of oscillating biological systems.