RESEARCH CENTER
Nancy - Grand Est

FIELD

Activity Report 2018

Section New Results

Edition: 2019-03-07

ALICE Project-Team

7. New Results

7.1. Hex-dominant meshing: Mind the gap!

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

We proposed a robust pipeline that can generate hex-dominant meshes from any global parameterization of a tetrahedral mesh (Figure 1). We focus on robustness in order to be able to benchmark different parameterizations on a large database. Our main contribution is a new method that integrates the hexahedra (extracted from the parameterization) into the original object. The main difficulty is to produce the boundary of the result, composed of both faces of hexahedra and tetrahedra. Obviously, this surface must be a good approximation of the original object but, more importantly, it must be possible to remesh the volume bounded by this surface minus the extracted hexahedra (called void). We enforce these properties by carefully tracking and eliminating all possibilities of failure at each step of our pipeline.

We tested our method on a large collection of objects (200+) with different settings. In most cases, we obtained results of very good quality as compared to the state-of-the-art solutions.

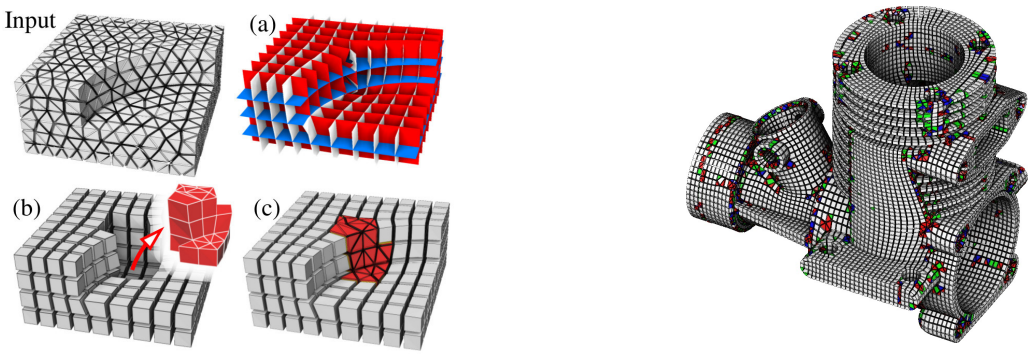

Figure 1. Our hexahedral-dominant meshing procedure: Start from an input tetrahedral mesh. Compute a global parameterization ***(a)****. Extract hexahedra by contouring the isovalues of the parameterization. Isolate the boundary of the void (in red), i.e., the volume with a degenerate / singular parameterization* ***(b)*** *(also called "gap" or "cavity"), shown in red. Remesh the void and stitch it into the hexahedral mesh* ***(c)****.*

7.2. Meshless Voronoi on the GPU

Participants: Nicolas Ray, Dmitry Sokolov, Sylvain Lefebvre, Bruno Lévy.

We proposed a GPU algorithm that computes a 3D Voronoi diagram (Figure 2). Our algorithm is tailored for applications that solely make use of the geometry of the Voronoi cells, such as Lloyd's relaxation used in meshing, or some numerical schemes used in fluid simulations and astrophysics. Since these applications only require the geometry of the Voronoi cells, they do not need the combinatorial mesh data structure computed by the classical algorithms (Bowyer-Watson). Thus, by exploiting the specific spatial distribution of the point-sets used in this type of applications, our algorithm computes each cell independently, in parallel, based on its nearest neighbors. In addition, we show how to compute integrals over the Voronoi cells by decomposing them on the fly into tetrahedra, without needing to compute any combinatorial information. The advantages

of our algorithm is that it is fast, very simple to implement, has constant memory usage per thread and does not need any synchronization primitive. These specificities make it particularly efficient on the GPU: it gains one order of magnitude as compared to the fastest state-of-the-art multicore CPU implementations. To ease the reproducibility of our results, the full documented source code is included in the supplemental material.

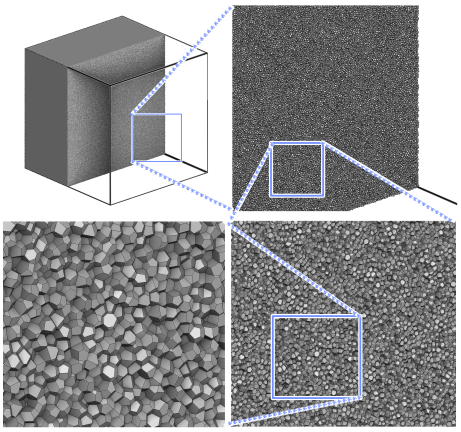

Figure 2. The 3D Voronoi diagram of 10 million points computed on the GPU in 800 ms (NVidia V100). We do not compute the tetrahedra, but in terms of equivalent computation speed, this corresponds to 84 million Delaunay tetrahedra per second.

7.3. Computational Optimal Transport

Participants: Bruno Lévy, Erica Schwindt.

We continued working on Optimal Transportation and its applications in fluid simulation and astrophysics [21], [20]. We developed an efficient and robust algorithm to compute Laguerre diagrams and intersections with tetrahedralized domains, that is, the geometric structure involved in a specific form of optimal transport that we are interested in. In addition, we developed an efficient parallel algorithm to compute Laguerre diagrams, with the possibility of handling periodic boundaries (3-torus), that is to say that the domain is a unit cube with opposite faces that are identified (if one leaves the domain from the left, it enters the domain from the right, etc..., like in the PacMan game). Such a topology is interesting for some simulations in astrophysics, or in material science, that consider a huge domain with homogeneous behavior and replace it with a tiny fraction and periodic boundary conditions (equivalent to a periodic material). We made the algorithms available in the geogram programming library (http://alice.loria.fr/software/geogram/doc/html/index.html). In cooperation with Roya Mohayaee (Institut d'Astrophysique de Paris) and Jean-Michel Alimi (Observatoire de Paris), we started applying the method to some inverse problems in astrophysics (Early Universe Reconstruction), that is reconstructing the past history of the universe from a 3D map of the galaxy clusters. Under some simplifying assumptions, the problem is precisely an instance of semi-discrete optimal transport that our algorithm solves efficiently. Our algorithm does the computation on a desktop PC within hours for several tenths of million points. With Quentin Merigot and Hugo Leclerc (U. Paris Sud), we are designing a new algorithm with the aim of scaling up to billions points (as requested by our astrophysicist colleagues).

BIGS Project-Team

6. New Results

6.1. Stochastic modelling

Participants: A. Gégout-Petit, S. Mézières, Y. Petot, P. Vallois

In the framework of the esca-illness of vines, we 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 purpose (2), we developed an autologistic model (centered in a new way), estimators of the parameters, and showed their good properties, and proposed a way to choose between several neighborhood models. It is the object of preprint [41].

In the framework of chalara of ashand, through a collaboration with INRA researchers, we have proposed a mechanistic model of propagation whose parameters are estimated by bayesian estimation. It is the object of the communication [26].

In a collaboration with physicists from Nancy CHRU, we have worked about the interest to use the whole distribution of telomeres lengths until the mean that is usually used to characterise ageing of a cell. We have shown that the shape of the distribution can be seen as a individuals's signature. It is the object of the accepted paper [10].

We analyse the probabilistic features of the Choquet integral with respect to a capacity over a finite set where the entries are random variables. Despite the amount of studies, the question of uncertainty remains under-considered. Such a question is of first importance in many applications and uses.

In the multifactorial context of modelling for gliomas, we focused our attention on the acquisition of the tumor diameter from clinical-collected data [1]. 3-D reconstruction via an equivalent sphere from multiple contouring of the tumor leads us to characterize its infiltrating phenotype (infiltration rate, direction of infiltration, evolution of morphology over time), current work. Our aim is to incorporate this new factor in the modeling already started (to appear in JBHI, beginning 2019).
A brain cartography obtained by sensorial simulations during awake surgery with the aid of clustering analysis is in revision.

6.2. Estimation and control for Markov Processes

Participants: R. Azaïs, F. Bouguet, A. Gégout-Petit, F. Greciet, B. Scherrer

Piecewise-deterministic Markov processes form a class of stochastic models with a sizeable scope of applications. Such processes are defined by a deterministic motion punctuated by random jumps at random times, and offer simple yet challenging models to study. The issue of statistical estimation of the parameters ruling the jump mechanism is far from trivial. Responding to new developments in the field as well as to current research interests and needs, the book "Statistical Inference for Piecewise-deterministic Markov Processes" edited by Romain Azaïs and Florian Bouguet [33] gather 7 chapters by different authors on the topic. The idea for this book stemmed from a workshop organized in Nancy in the 2016-17 winter. Two chapters [48][31] have been co-authored by one or more BIGS members.

Multiple-step lookahead policies have demonstrated high empirical competence in Reinforcement Learning, via the use of Monte Carlo Tree Search or Model Predictive Control. In [13], multiple-step greedy policies and their use in vanilla Policy Iteration algorithms were proposed and analyzed. In [14], [12], we study multiple-step greedy algorithms in more practical setups: we describe and analyze a stochastic approximation variation and general sensitivity analyses to approximations. In [15], we describe a short study on an Anderson acceleration of the fixed point computation involved in Reinforcement Learning. These contributions resulted in one publication in ICML, in NeurIPS, and two in EWRL (the European Workshop on Reinforcement Learning).

6.3. Algorithms and Estimation for graph data

Participants: A. Gégout-Petit, A. Gueudin, C. Karmann

In the purpose to deal with inference for network of zero-inflated variables, we have developped a new regression model. We consider the problem of variable selection when the response is ordinal, that is an ordered categorical variable. In particular, we are interested in selecting quantitative explanatory variables linked with the ordinal response variable and we want to determine which predictors are relevant. In this framework, we choose to use the polytomous ordinal logistic regression model using cumulative logits which generalizes the logistic regression. We then introduce the Lasso estimation of the regression coefficients using the Frank-Wolfe algorithm. To deal with the choice of the penalty parameter, we use the stability selection method and we develop a new method based on the knockoffs idea. This knockoffs method is general and suitable to any regression and besides, gives an order of importance of the covariates. Finally, we provide some experimental results to corroborate our method and we present an application of this regression method for zero-inflated network inference. This work is the object of a presentation in a conference [28] and a preprint submitted in a journal [40].

6.4. Regression and machine learning

Participants: E. Albuisson, R. Azaïs (Inria, Lyon), T. Bastogne, L. Batista, K. Duarte, S. Ferrigno, A. Gégout-Petit, P. Guyot, J.-M. Monnez, N. Sahki, S. Mézières

In the purpose to detect change of health state for lung-transplanted patient, we have begun to work on breakdowns in multivariate physiological signals. Based on the CUSUM statistics, we have used dynamical thresholds of detection [27]. A more general talk about statistical learning and connected patient was given in a workshop "Evaluation des objets en santé connectée" [35].

We consider the analysis of cardiomyocyte signals (cardiac cells) for the cardiotoxicity assessment of new pharmaceutical compounds in preclinical assays. The experimental data are either impedance signals measuring the contractility of cardiomyocytes [39], [4], field potential signals measuring their functionality or fluorescence signals measuring the activity of some ion channels such as calcium pumps (Ca2+). At this preclinical level, our main contribution is the estimation of important characteristics such the field potential duration [17] or the identification of cardiotoxic events such as the early-afterdepolarization.We have also developed new methods for the analysis of electrocardiograms at patient level and more precisely the estimation of parameters such as the RR and QT intervals in long and noisy signals provided by wearable sensors [24], [30], [23], [25], [29]. We also study the efficacy of a new biomarker in radiotherapy. The objective is to compute a score able to predict risk of radiosensitivity for patients in radiotherapy [19], [20].We are also developing a new method to characterize the potential interactions between nanoparticles and biological compounds of complex media such as blood. This new method aims at predicting risks on the biodistribution and toxicity of the nanoparticles [16], [36].

In [7], we present a methodology for constructing a short-term event risk score from an ensemble predictor using bootstrap samples, two different classification rules, logistic regression and linear discriminant analysis for mixed data, continuous or categorical, and random selections of variables into the construction of predictors. We establish a property of linear discriminant analysis for mixed data and define an event risk measure by an odds-ratio. This methodology is applied to heart failure patients on whom biological, clinical and medical history variables were measured and the results obtained from our data are detailed.

The study [8] addresses the problem of sequential least square multidimensional linear regression, particularly in the case of a data stream, using a stochastic approximation process. To avoid the phenomenon of numerical explosion which can be encountered and to reduce the computing time in order to take into account a maximum of arriving data, we propose using a process with online standardized data instead of raw data and the use of several observations per step or all observations until the current step. Herein, we define and study the almost sure convergence of three processes with online standardized data: a classical process with a variable step-size and use of a varying number of observations per step, an averaged process with a constant step-size and use of a varying number of observations per step, and a process with a variable or constant step-size and use of all observations until the current step. Their convergence is obtained under more general assumptions than classical ones. These processes are compared to classical processes on 11 datasets for a fixed total number of observations used and thereafter for a fixed processing time. Analyses indicate that the third-defined process typically yields the best results.

Many articles were devoted to the problem of estimating recursively the eigenvectors and eigenvalues in decreasing order of the expectation of a random matrix using an i.i.d. sample of it. In [43], we make the following contributions. The convergence of a normed process is proved under more general assumptions: the random matrices are not supposed i.i.d. and a new data mini-batch or all data until the current step are taken into account at each step without storing them; three types of processes are studied; this is applied to online principal component analysis of a data stream, assuming that data are realizations of a random vector Z whose expectation is unknown and must be estimated online, as well as possibly the metrics used when it depends on unknown characteristics of Z.

Let $Y = m(X) + \sigma(X)\varepsilon$ be a regression model, where $m(\cdot)$ is the regression function, $\sigma^2(\cdot)$ the variance function and ε the random error term. Methods to assess how well a model fits a set of observations fall under the banner of goodness-of-fit tests. Many tests have been developed to assess the different assumptions for this kind of model. Most of them are “directional” in that they detect departures from mainly a given assumption of the model. Other tests are “global” in that they assess whether a model fits a data set on all its assumptions. We focus on the task of choosing the structural part $m(\cdot)$. It gets most attention because it contains easily interpretable information about the relationship between X and Y. To valid the form of the regression function, we consider three nonparametric tests based on a generalization of the Cramér-von Mises statistic. The first two are directional tests, while the third is a global test. To perform these goodness-of-fit tests based on a generalization of the Cramér-von Mises statistic, we have used Wild bootstrap methods and we also proposed a method to choose the bandwidth parameter used in nonparametric estimations. Then, we have developed cvmgof R package (being submitted), an easy-to-use tool for many users. The use of the package is illustrated using simulations to compare the three implemented tests [37].

In epidemiology, we are working with clinicians to study fetal development in the last two trimesters of pregnancy. We have data from the "Service de foetopathologie et de placentologie" of the "Maternité Régionale Universitaire" (CHU Nancy) and from the EDEN cohort (INSERM). We propose to use non parametric methods of estimation to obtain reference curves of fetus and child growth. In addition, we want to develop a test, based on Z-scores, to detect any slope breaks in the fetal development curves (work in progress).

CAMUS Team

7. New Results

7.1. AutoParallel: A Python module for automatic parallelization and distributed execution of affine loop nests

Participant: Philippe Clauss.

The last improvements in programming languages, programming models, and frameworks have focused on abstracting the users from many programming issues. Among others, recent programming frameworks include simpler syntax, automatic memory management and garbage collection, which simplifies code re-usage through library packages, and easily configurable tools for deployment. For instance, Python has risen to the top of the list of the programming languages due to the simplicity of its syntax, while still achieving a good performance even being an interpreted language. Moreover, the community has helped to develop a large number of libraries and modules, tuning the most commonly used to obtain great performance.

However, there is still room for improvement when preventing users from dealing directly with distributed and parallel computing issues. This work proposes AutoParallel, a Python module to automatically find an appropriate task-based parallelization of affine loop nests to execute them in parallel in a distributed computing infrastructure. This parallelization can also include the building of data blocks to increase task granularity in order to achieve a good execution performance. Moreover, AutoParallel is based on sequential programming and only contains a small annotation in the form of a Python decorator so that anyone with little programming skills can scale up an application to hundreds of cores.

This work has been published in [18] and is the result of a collaboration between Philippe Clauss, Cristian Ramon-Cortes, PhD student, and Rosa M. Badia, his PhD advisor, both from the Barcelona Supercomputing Center, Spain.

7.2. Optimization of recursive functions by transformation into loops

Participants: Salwa Kobeissi, Philippe Clauss.

Recursion is a fundamental computing concept that offers the opportunity to elegantly solve various kinds of problems, particularly those whose solutions depend on solutions of smaller instances of their own. Nevertheless, today in imperative languages, recursive functions are still not considered sufficiently time-efficient in comparison with the alternative equivalent iterative code. Although many advanced and aggressive optimizers have been developed to enhance the performance of iterative control structures, there are still no such sophisticated and advanced techniques built for the sake of optimizing recursions.

We propose an approach that makes possible applying powerful optimizations on recursive functions through transforming them into loops. We are particularly interested in applying polyhedral optimization techniques which usually tackle affine loops. Therefore, the scope of our study is restricted to recursive functions whose control flow and memory accesses exhibit an affine behavior, which means that there exists a semantically equivalent affine loop nest, candidate for polyhedral optimizations. Accordingly, our approach is based on analyzing early executions of a recursive program using a Nested Loop Recognition algorithm, performing the convenient recursion-to-iteration transformation of the original program and, finally, applying further loop optimizations using the polyhedral compiler Polly. This approach brings recursion optimization techniques into a higher level in addition to widening the scope of the polyhedral model to include originally non-loop programs.

This work is the topic of Salwa Kobeissi's PhD. A first paper has been submitted to an international workshop.

7.3. Impact Study of Data Locality on Task-Based Applications Through the Heteroprio Scheduler

Participant: Bérenger Bramas.

Task-based parallelization is massively used in high-performance computing on heterogeneous hardware because it allows programmers to finely describe the intrinsic parallelism of the algorithms while ignoring the hardware details. However, this approach delegates the main decisions to the scheduler, making it a critical component responsible for the distribution of the tasks on the different types of processing unit. In a former work, Bérenger Bramas has proposed the Heteroprio scheduler, which has demonstrated to be extremely efficient in the computation of the fast multipole method or linear algebra factorizations/decompositions. However, the original version was not taking into account data locality leading to loss of execution efficiency from important data movements between the memory nodes.

The current work aimed at improving the Heteroprio scheduler by making it locality sensitive. The idea is to divide the task-lists to have as many lists as there are memory nodes. Then, the two main issues are to find where to store the new ready tasks and to decide how to iterate over all the task-lists. For the first problem, we have studied different locality scores to find the best memory node for each task, and we have demonstrated that taking into account the type of data access - read or write - allows for significant improvement. Concerning the iteration order, we have proposed to use a priority distance and a memory distance such that the tasks are stolen from memory nodes that are close but also that have opposite priorities.

All these ideas were implemented in a scheduler inside StarPU and have been validated on two applications: QrMumps from Alfredo Buttari (IRIT) and SpLDLT from Florent Lopez (Rutherford Appleton Laboratory, UK.). The performance study demonstrated the benefit of our approach with a significant improvement in terms of execution time and data movement. The executions were accelerated by 30% for QrMumps and 80% for SpLDLT. The results will now be written into a dedicated paper for publication.

7.4. Combining Locking and Data Management Interfaces

Participants: Jens Gustedt, Maxime Mogé, Mariem Saied, Daniel Salas.

7.4.1. Ordered Read-Write Locks

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 [2] shows how to produce deadlock-free iterative programs that implement strong overlapping between communication, IO and computation.

An implementation (ORWL) of our ideas of combining control and data management in C has been undertaken, see Section 6.6 . In previous work it has demonstrated its efficiency for a large variety of platforms.

In the context of the thesis of Mariem Saied, a new domain specific language (DSL) has been completed that largely eases the implementation of applications with ORWL. In its first version it provides an interface for stencil codes. The approach allows to describe stencil codes quickly and efficiently, and leads to substantial speedups.

In the framework of the ASNAP project (see 9.1.2) we have used ordered read-write locks (ORWL) as a model to dynamically schedule a pipeline of parallel tasks that realize a parallel control flow of two nested loops; an outer *iteration* loop and an inner *data traversal* loop. Other than dataflow programming we emphasize on upholding the sequential modification order of each data object. As a consequence the visible side effects on any object can be guaranteed to be identical to a sequential execution. Thus the set of optimizations that are performed are compatible with C's abstract state machine and compilers could perform them, in principle, automatically and unobserved. See [19] for first results.

In the context of the Prim'Eau project (see 9.1.1) we use ORWL to integrate parallelism into an already existing `Fortran` application that computes floods in the region that is subject to the study. A first step of such a parallelization has been started by using ORWL on a process level. Our final goal will be to extend it to the thread level and to use the application structure for automatic placement on compute nodes.

Within the framework of the thesis of Daniel Salas we have successfully applied ORWL to process large histopathology images. We are now able to treat such images distributed on several machines or shared in an accelerator (Xeon Phi) transparently for the user.

7.4.2. Low level locks

Our low level locks algorithm that is based on atomics and Linux' futexes [25] [26] has been integrated into the `musl` C library (see Section 6.7) and is thus deployed in several Linux distributions that use musl as their base.

7.5. High-Performance Particle-in-Cell Simulations

Participants: Arthur Charguéraud, Yann Barsamian, Alain Ketterlin.

Yann Barsamian's PhD thesis focuses on the development of efficient programs for Particle-in-Cell (PIC) simulations, with application to plasma physics. On recent multi-core hardware, performance of this code is often limited by memory bandwidth. We describe a multi-core PIC algorithm that achieves close-to-minimal number of memory transfers with the main memory, while at the same time exploiting SIMD instructions for numerical computations and exhibiting a high degree of OpenMP-level parallelism. Our algorithm keeps particles sorted by cell at every time step, and represents particles from the same cell using a linked list of fixed-capacity arrays, called chunks. Chunks support either sequential or atomic insertions, the latter being used to handle fast-moving particles. To validate our code, called Pic-Vert, we consider a 3d electrostatic Landau-damping simulation as well as a 2d3v transverse instability of magnetized electron holes. Performance results on a 24-core Intel Skylake hardware confirm the effectiveness of our algorithm, in particular its high throughput and its ability to cope with fast moving particles. A paper describing this work was published at Euro-par [13] and is described in more details in Yann Barsamian's PhD thesis [6].

7.6. Granularity Control for Parallel Programs

Participant: Arthur Charguéraud.

Arthur Charguéraud contributes to the ERC DeepSea project, which is hosted at Inria Paris (team Gallium). With his co-authors, he focused recently on the development of techniques for controlling granularity in parallel programs. Granularity control is an essential problem because creating too many tasks may induce overwhelming overheads, while creating too few tasks may harm the ability to process tasks in parallel. Granularity control turns out to be especially challenging for nested parallel programs, i.e., programs in which parallel constructs such as fork-join or parallel-loops can be nested arbitrarily. This year, the DeepSea team investigated two different approaches.

The first one is based on the use of asymptotic complexity functions provided by the programmer, combined with runtime measurements to estimate the constant factors that apply. Combining these two sources of information allows to predict with reasonable accuracy the execution time of tasks. Such predictions may be used to guide the generation of tasks, by sequentializing computations of sufficiently-small size. An analysis is developed, establishing that task creation overheads are indeed bounded to a small fraction of the total runtime. These results extend prior work by the same authors [22], extending them with a carefully-designed algorithm for ensuring convergence of the estimation of the constant factors deduced from the measures, even in the face of noise and cache effects, which are taken into account in the analysis. The approach is demonstrated on a range of benchmarks taken from the state-of-the-art PBBS benchmark suite. These results have been accepted for publication at PPoPP'19.

The second approach is based on an instrumentation of the runtime system. The idea is to process parallel function calls just like normal function calls, by pushing a frame on the stack, and only subsequently promoting these frames as threads that might get scheduled on other cores. The promotion of frames takes place at regular time interval, hence the name *heartbeat scheduling* given to the approach. Unlike in prior approaches such as *lazy scheduling*, in which promotion is guided by the work load of the system, hearbeat scheduling can be proved to induce only small scheduling overheads, and to not reduce asymptotically the amount of parallelism

inherent to the parallel program. The theory behind the approach is formalized in Coq. It is also implemented through instrumented C++ programs, and evaluated on PBBS benchmarks. A paper describing this approach was published at PLDI'18 [12].

7.7. Program verification and formal languages

Participant: Arthur Charguéraud.

- Armaël Guéneau, PhD student advised by A. Charguéraud and F. Pottier, has developed a Coq library formalizing the asymptotic notation (big-O), and has developed an extension of the CFML verification tool to allow specifying the asymptotic complexity of higher-order, imperative programs. This new feature has been tested on several classic examples of complexity analyses, including: nested loops in $O(n^3)$ and $O(nm)$, selection sort in $O(n^2)$, recursive functions in $O(n)$ and $O(2^n)$, binary search in $O(\log n)$, and Union-Find in $O(\alpha(n))$. A paper was describing this work was published at ESOP'18 [15].
- A. Charguéraud, together with Ralf Jung and Jan-Oliver Kaiser and Derek Dreyer (MPI-SWS), Robbert Krebbers (Delft University of Technology), Jacques-Henri Jourdan (Inria), Joseph Tassarotti (Carnegie Mellon University), and Amin Timany (KU Leuven), developed MoSel, a general and extensible Coq framework for carrying out separation-logic proofs mechanically using an interactive proof assistant. This tool extends the Iris Proof Mode (IPM) to make it applicable to both affine and linear separation logics (and combinations thereof), and to provide generic tactics that can be easily extended to account for the bespoke connectives of the logics with which it is instantiated. To demonstrate the effectiveness of MoSeL, the tool has been instantiated to provide effective tactical support for interactive and semi-automated proofs in six very different separation logics. This work was published at ICFP'18 [17].
- A. Charguéraud advised Ramon Fernandez for a 4-month internship. The aim of that internship was to formalize, using the Coq proof assistant, several data layout transformations such as the transformation from an array of structures to a structure of arrays (AoS-to-SoA). Such transformations are routinely employed to develop high-performance code. Ramon investigated the literature on data layout transformations, listed the most useful transformations exploited in practice, and identified several core transformations from which almost all others can be derived. He then successfully carried out proofs of semantic preservation for the three most important transformations: field grouping, tiling, and AoS-to-SoA.
- A. Charguéraud, together with Alan Schmitt (Inria Rennes) and Thomas Wood (Imperial College), developed an interactive debugger for JavaScript. The interface, accessible as a webpage in a browser, allows to execute a given JavaScript program, following step by step the formal specification of JavaScript developed in prior work on *JsCert* [24]. Concretely, the tool acts as a double-debugger: one can visualize both the state of the interpreted program and the state of the interpreter program. This tool is intended for the JavaScript committee, VM developers, and other experts in JavaScript semantics. A paper describing the tool appeared at the international conference Web Programming [14].

7.8. Flexible Runtime System with High Throughput for Many-to-Many Data Stream Problems

Participants: Paul Godard, Vincent Loechner, Cédric Bastoul.

In the context of our collaboration with the Caldera company, we are interested in high throughput data stream problems, that require low latency, maximal bandwidth usage, and that avoid starvations. We suppose that we receive jobs from an external system through a queue, each job including a description of its computation needs, output data requirements and output locations.

The computations are distributed on a cluster organized in a many-to-many logical topology where one or many computing tasks (producers) send data to one or many consumer tasks (consumers). The runtime system is orchestrated by a centralized scheduler, which decomposes jobs into tasks and dynamically assigns them to producers. The producers perform the computations and send their output data to the consumers. The consumers collect and order output data to make them available to the final user.

We implemented our framework, and performed some experiments on a real-world use case: real time professional digital printing, that may require tens of Gbit/s sustained output rates. We show in our measurements that our system scales and reaches data rates that are close to the maximum throughput of our experimental hardware. The architecture as a cluster and using the standard TCP/IP network protocol allow our system to be highly adaptive to the user's requirements. We are in the process of writing a paper describing our framework architecture for many-to-many data stream problems and results.

7.9. Visual Program Manipulation in the Polyhedral Model

Participants: Cédric Bastoul, Oleksandr Zinenko, Stéphane Huot.

While a plethora of libraries and frameworks focus on expressing parallelism, identifying and extracting it remains a challenging task. Automatic parallelization relies on imprecise heuristics resulting in cumbersome manual code analysis and transformation in case of underperformance. Alternatively, directive-based approaches often require transforming the program from scratch when a slightly modified version of an automatically-computed transformation would suffice. We propose an interactive visual approach building on the polyhedral model that (1) visualizes exact dependences and parallelism, (2) decomposes a complex automatically-computed transformation into simple steps for replay and easier modification, and (3) allows for directly manipulating the visual representation as a means of transforming the program with immediate feedback. User studies suggest that our visualization is understood by experts and non-experts alike, and that it may favor an exploratory approach to transformation. Finally, an eye-tracking study suggests that programmers may resort to visualizations instead of code if visualizations are clearly efficient for a given task.

This is a joint work with PARKAS team at Inria Paris (contact: Oleksandr Zinenko) and MJOLNIR team at Inria Lille (contact: Stéphane Huot), published in TACO [10].

7.10. A language extension set to generate adaptive versions automatically

Participants: Maxime Schmitt, Cédric Bastoul.

A large part of the development effort of compute-intensive applications is devoted to optimization, i.e., achieving the computation within a finite budget of time, space or energy. Given the complexity of modern architectures, writing simulation applications is often a two-step workflow. Firstly, developers design a sequential program for algorithmic tuning and debugging purposes. Secondly, experts optimize and exploit possible approximations of the original program to scale to the actual problem size. This second step is a tedious, time-consuming and error-prone task. During this year, we investigated language extensions and compiler tools to achieve that task semi-automatically in the context of approximate computing. We identified the semantic and syntactic information necessary for a compiler to automatically handle approximation and adaptive techniques for a particular class of programs. We proposed a set of language extensions generic enough to provide the compiler with the useful semantic information when approximation is beneficial. We implemented the compiler infrastructure to exploit these extensions and to automatically generate the adaptively approximated version of a program. We conducted an experimental study of the impact and expressiveness of our language extension set on various applications.

These language extensions and the underlying compiler infrastructure are a significant output of 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. A paper presenting these extensions has been accepted to the OGST journal, targeting typical end-users.

CAPSID Project-Team

7. New Results

7.1. Drug Targeting and Adverse Drug Side Effects

Identifying new molecular targets using comparative genomics and knowledge of disease mechanisms is a rational first step in the search for new preventative or therapeutic drug treatments [63]. We are mostly concerned with three global health problems, namely fungal and bacterial infections and hypertension. Through on-going collaborations with several Brazilian laboratories (at University of Mato Grosso State, University of Maringá, Embrapa, and University of Brasilia), we previously identified several novel small-molecule drug leads against *Trypanosoma cruzi*, a parasite responsible for Chagas disease [91]. With the University of Maringá, we subsequently found several active molecules against the flavoenzyme TRR1 in *Candida albicans*, and two manuscripts are in preparation. We also proposed several small-molecule inhibitors against *Fusarium graminearum*, a fungal threat to global wheat production [63], [43]. Two further manuscripts on this topic are currently in preparation. Concerning hypertension, we continued our collaboration with Prof. Catherine Llorens-Cortes at Collège de France to study the interaction between the apelin receptor (a transmembrane protein important for blood pressure regulation) and the aminopetidase A enzyme [47].

It is well known that many therapeutic drug molecules can have adverse side effects. However, when patients take several combinations of drugs it can be difficult to determine which drug is responsible for which side effect. In collaboration with Adrien Coulet (Orpailleur team co-supervisor of Gabin Personeni) and Prof. Michel Dumontier (Biomedical Informatics Research Laboratory, Stanford), we developed an approach which combines multiple ontologies such as the Anatomical Therapeutical Classification of Drugs, the ICD-9 classification of diseases, and the SNOMED-CT medical vocabulary together with the use of Pattern Structures (an extension of Formal Concept Analysis) in order to extract association rules to analyse the co-occurrence of adverse drug effects in patient records [74], [73]. A paper describing this work has been published in the Journal of Biomedical Semantics [70].

7.2. Docking Symmetrical Protein Structures

Many proteins form symmetrical complexes in which each structure contains two or more identical copies of the same sub-unit. We recently developed a novel polar Fourier docking algorithm called "Sam" for automatically assembling symmetrical protein complexes. A journal article describing the Sam algorithm has been published [8]. An article describing the results obtained when using Sam to dock several symmetrical protein complexes from the "CASP/CAPRI" docking experiment has also been published [53]. This study showed that many of the models of protein structures built by members of the "CASP" fold prediction community are "dockable" in the sense that Sam is able to find acceptable docking solutions from amongst the CASP models.

More recently, we are working to extend the polar Fourier correlation algorithm to use very high angular resolution spherical Bessel basis functions. As part of this work, we have developed a very fast recursive algorithm for calculating high order Clebsch-Gordan coupling coefficients [30]. A manuscript describing this work has been submitted to a quantum mechanics journal.

7.3. Multiple Flexible Protein Structure Alignments

Comparing two or more proteins by optimally aligning and superposing their backbone structures provides a way to detect evolutionary relationships between proteins that cannot be detected by comparing only their primary amino-acid sequences. The latest version of our "Kpax" protein structure alignment algorithm can flexibly align pairs of structures that cannot be completely superposed by a single rigid-body transformation, and can calculate multiple alignments of several similar structures flexibly [9]. In collaboration with Alain

Hein of the INRA lab "Agronomie et Environnement", we used Kpax to help study the structures of various "Cyp450" enzymes in plants [81]. In collaboration with Emmanuel Levy of the Weizmann Institute, we used Kpax to superpose and compare all of the symmetrical protein complexes in the Protein Databank in order to verify or remediate their quaternary structure annotations. A manuscript describing this work has been published in Nature Methods [15].

7.4. Large-Scale Annotation of Protein Domains and Sequences

Many protein chains in the Protein Data Bank (PDB) are cross-referenced with Pfam domains and Gene Ontology (GO) terms. However, these annotations do not explicitly indicate any relation between EC numbers and Pfam domains, and many others lack GO annotations. In order to address this limitation, as part of the PhD thesis project of Seyed Alborzi, we developed the CODAC approach for mining multiple protein data sources (i.e. SwissProt, TremBL, and SIFTS) in order to associate GO molecular function terms with Pfam domains, for example. We named the software implementation "GO-DomainMiner". This work was first presented at IWBBIO 2017 [36]. A full paper has recently been accepted for a special issue of *BMC Bioinformatics* [13].

In collaboration with Maria Martin's team at the European Bioinformatics Institute (EBI), we combined the CODAC approach with a novel combinatorial association rule based approach called "CARDM" for annotating protein sequences. When applied to the large UniProt/TrEMBL sequence database of 63 million protein entries, CARDM predicted over 24 million Enzyme Commission (EC) numbers and 188 million GO terms for those entries. A journal paper in collaboration with the EBI on comparing the quality of these predicted annotations with other state of the art annotation methods is in preparation, and a poster was presented at ISMB-ECCB-2017 [35]. As part of the PhD thesis of Bishnu Sarker, we also developed GrAPFI, a graph-based protein function annotation approach. GrAPFI applies a label propagation algorithm to a complex network representation of protein sequence data. A full paper on this work has recently been accepted by the International Conference on Complex Networks and their Applications [24].

7.5. Distributed Protein Graph Processing

The huge number of protein sequences in protein databases such as UniProtKB calls for rapid procedures to annotate them automatically. We are using existing protein annotations to predict the annotations of new or non-reviewed proteins. In this context, we developed the "DistNBLP" method for annotating protein sequences using a graph representation and a distributed label propagation algorithm. DistNBLP uses the BLADYG framework [38] to process protein graphs on multiple compute nodes by applying a neighbourhood-based label propagation algorithm in a distributed way. We applied DistNBLP in the recent "CAFA 3" (critical Assessment of Protein Function Annotation) community experiment to annotate new protein sequences automatically. This work was presented as a poster at ISMB/ECCB-2017 [34]. We are also interested in feature selection for subgraph patterns. In collaboration with the LIMOS laboratory at Université Clermont Auvergne we also developed a scalable approach using MapReduce for identifying sub-graphs having similar labels in very large graphs [51].

7.6. Flexible Docking of Protein-GAG Complexes

Modeling how flexible polymers bind to proteins presents enormous computational challenges due to the large conformational search space that arises from the many internal rotational degrees of freedom in polymer structures. In collaboration with Sergey Samsonov (Gdansk University, Poland), we extended our fragment-based flexible docking approach [83], [42] to model how flexible Glycosaminoglycans (GAGs) might bind to the surface of a known protein structure. A paper has been submitted to the Journal of Computational Chemistry.

In collaboration with Sjoerd de Vries (Univ Paris Diderot), we have created a new protein-glycan interaction force-field and integrated it in the ATTRACT docking engine [83]. We also participated in a comparative study of the main current protein-GAG docking methods.

7.7. Stochastic Decision Trees for Similarity Computation

We have designed a method to compute similarities on unlabeled data using stochastic decision trees [20]. The main idea of Unsupervised Extremely Randomized Trees (UET) is to randomly and iteratively split the data until a stopping criterion is met. Pairwise similarity values are computed based on the co-occurrence of samples in the leaves of each generated tree. We evaluate our method on synthetic and real-world datasets by comparing the mean similarities between samples with the same label and the mean similarities between samples with distinct labels. Empirical studies show that the method effectively gives distinct similarity values between samples belonging to distinct clusters, and gives indiscernible values when there is no cluster structure. We also assessed some interesting properties such as invariance under monotone transformations of variables and robustness to correlated variables and noise. Our experiments show that the algorithm outperforms existing methods in some cases, and can reduce the amount of preprocessing needed with many real-world datasets. We plan to study the application of this "global" pairwise similarity computation to quantify protein structural similarities. Two interesting problems will concern the representation of the protein structure and how to tackle extra constraints such as invariance under rotational and translational transformations.

CARAMBA Project-Team

7. New Results

7.1. A new family of pairing-friendly elliptic curves

Participant: Aurore Guillevic.

In [11], together with M. Scott from Miracl, we presented an algorithm to generate new families of pairing-friendly curves. It generalizes the very popular Barreto-Naehrig curves. This paper jointly received the best paper award of the conference.

7.2. Faster individual discrete logarithms in finite fields of composite extension degree

Participant: Aurore Guillevic.

We improved in [7] the previous work [25] on speeding-up the first phase of the individual discrete logarithm computation, the initial splitting, a.k.a. the smoothing phase. We extended the algorithm to any non-prime finite field $\mathbb{F}_{p^n}$ where n is composite. We also applied it to the new variant Tower-NFS. The paper is now published.

7.3. Polynomial Time Bounded Distance Decoding near Minkowski's Bound in Discrete Logarithm Lattices

Participant: Cécile Pierrot [contact].

In [6], together with Léo Ducas, we proposed a concrete family of dense lattices of arbitrary dimension n in which the lattice Bounded Distance Decoding (BDD) problem can be solved in deterministic polynomial time. The lattice construction needs discrete logarithm computations that can be made in deterministic polynomial time for well-chosen parameters. Each lattice comes with a deterministic polynomial time decoding algorithm able to decode up to a large radius. Namely, we reached decoding radius within $O(\log n)$ Minkowski's bound, for both ℓ_1 and ℓ_2-norms.

7.4. Improved complexity bounds for counting points on hyperelliptic curves

Participants: Simon Abelard, Pierrick Gaudry [contact], Pierre-Jean Spaenlehauer [contact].

In [3], we presented a probabilistic Las Vegas algorithm for computing the local zeta function of a hyperelliptic curve of genus g defined over $\mathbb{F}_q$. It is based on the approaches by Schoof and Pila combined with a modeling of the ℓ-torsion by structured polynomial systems. Our main result improves on previously known complexity bounds by showing that there exists a constant $c > 0$ such that, for any fixed g, this algorithm has expected time and space complexity $O((\log q)^{cg})$ as q grows and the characteristic is large enough.

7.5. Counting points on genus-3 hyperelliptic curves with explicit real multiplication

Participants: Simon Abelard, Pierrick Gaudry [contact], Pierre-Jean Spaenlehauer [contact].

In [9], we proposed a Las Vegas probabilistic algorithm to compute the zeta function of a genus-3 hyperelliptic curve defined over a finite field $\mathbb{F}_q$, with explicit real multiplication by an order $\mathbb{Z}[\eta]$ in a totally real cubic field. Our main result states that this algorithm requires an expected number of $O((\log q)^6)$ bit-operations, where the constant in the $O()$ depends on the ring $\mathbb{Z}[\eta]$ and on the degrees of polynomials representing the endomorphism η. As a proof-of-concept, we computed the zeta function of a curve defined over a 64-bit prime field, with explicit real multiplication by $\mathbb{Z}[2\cos(2\pi/7)]$.

7.6. Counting points on hyperelliptic curves with explicit real multiplication in arbitrary genus

Participant: Simon Abelard.

In [14], we presented a probabilistic Las Vegas algorithm for computing the local zeta function of a genus-g hyperelliptic curve defined over $\mathbb{F}_q$ with explicit real multiplication (RM) by an order $\mathbb{Z}[\eta]$ in a degree-g totally real number field. It is based on the approaches by Schoof and Pila in a more favorable case where we can split the ℓ-torsion into g kernels of endomorphisms, as introduced by Gaudry, Kohel, and Smith in genus 2. To deal with these kernels in any genus, we adapted a technique that Abelard, Gaudry, and Spaenlehauer introduced to model the ℓ-torsion by structured polynomial systems. Applying this technique to the kernels, the systems we obtained are much smaller and so is the complexity of solving them. Our main result is that there exists a constant $c > 0$ such that, for any fixed g, this algorithm has expected time and space complexity $O((\log q)^c)$ as q grows and the characteristic is large enough. We proved that $c \leq 8$ and we also conjecture that the result still holds for $c = 6$.

7.7. A fast randomized geometric algorithm for computing Riemann-Roch spaces

Participants: Aude Le Gluher, Pierre-Jean Spaenlehauer [contact].

In [16], we proposed a probabilistic Las Vegas variant of Brill-Noether's algorithm for computing a basis of the Riemann-Roch space $L(D)$ associated to a divisor D on a projective plane curve $\mathcal{C}$ over a sufficiently large perfect field k. Our main result shows that this algorithm requires at most $O(\max(\deg(\mathcal{C})^{2\omega}, \deg(D_+)^{\omega}))$ arithmetic operations in k, where ω is a feasible exponent for matrix multiplication and D_+ is the smallest effective divisor such that $D_+ \geq D$. This improves the best known upper bounds on the complexity of computing Riemann-Roch spaces. Our algorithm may fail, but we showed that provided that a few mild assumptions are satisfied, the failure probability is bounded by $O(\max(\deg(\mathcal{C})^4, \deg(D_+)^2)/|E|)$, where E is a finite subset of k in which we pick elements uniformly at random. We provide a freely available C++/NTL implementation of the proposed algorithm, and experimental data. In particular, our implementation enjoys a speed-up larger than 9 on several examples compared to the reference implementation in the Magma computer algebra system. As a by-product, our algorithm also yields a method for computing the group law on the Jacobian of a smooth plane curve of genus g within $O(g^\omega)$ operations in k, which slightly improves in this context the best known complexity $O(g^{\omega+\varepsilon})$ of Khuri-Makdisi's algorithm.

7.8. Formal proof of mpfr_add

Participants: Jianyang Pan, Paul Zimmermann [contact].

With the help of Karthik Bhargavan (Prosecco project-team), we proved formally the correctness of the `mpfr_add` code in case where all inputs and the output have the same precision, and this precision is less than one limb (i.e., less than 64 bits on modern computers). The algorithm was proven formally correct using the F^* language, and the extracted code, which was shown to be as efficient as the original MPFR code, is now available in MPFR. A similar work was done for the multiplication `mpfr_mul`, but the proof of correctness was only partly completed.

7.9. Various ways to split a floating-point number

Participant: Paul Zimmermann.

Together with Claude-Pierre Jeannerod and Jean-Michel Muller (AriC project-team), we revisited in an unified way the classical algorithms to split a floating-point number in two parts, and some applications of these algorithms. Some new algorithms were also designed. This work was presented at the 25th IEEE Symposium on Computer Arithmetic [10].

7.10. A polyhedral method for sparse systems with many positive solutions

Participant: Pierre-Jean Spaenlehauer.

Together with Frédéric Bihan (Université Savoie Mont Blanc) and Francisco Santos (Universidad de Cantabria), we investigated in [4] a version of Viro's method for constructing polynomial systems with many positive solutions, based on regular triangulations of the Newton polytope of the system. The number of positive solutions obtained with our method is governed by the size of the largest positively decorable subcomplex of the triangulation. Here, positive decorability is a property that we introduced and which is dual to being a subcomplex of some regular triangulation. Using this duality, we produced large positively decorable subcomplexes of the boundary complexes of cyclic polytopes. As a byproduct we obtained new lower bounds, some of them being the best currently known, for the maximal number of positive solutions of polynomial systems with prescribed numbers of monomials and variables. We also studied the asymptotics of these numbers and observed a log-concavity property.

7.11. Fast Integer Multiplication Using Generalized Fermat Primes

Participants: Svyatoslav Covanov, Emmanuel Thomé [contact].

In [5] we described 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 support this assumption. This article was submitted to Mathematics of Computation and was completely rewritten in late 2017-early 2018. It is now accepted for final publication.

7.12. Improved Methods for Finding Optimal Formulae for Bilinear Maps in a Finite Field

Participant: Svyatoslav Covanov.

In [15], we described a method improving on the exhaustive search algorithm originally developed in [19]. We are able to compute new optimal formulae for the short product modulo X^5 and the circulant product modulo $(X^5 - 1)$. Moreover, we proved that there is essentially only one optimal decomposition of the product of 3×2 by 2×3 matrices up to the action of some group of automorphisms. This work has been submitted to *Theoretical Computer Science* and is tentatively accepted, pending minor revisions.

7.13. Using Constraint Programming to Solve a Cryptanalytic Problem

Participant: Marine Minier.

In [8], we described Constraint Programming (CP) models to solve a cryptanalytic problem: the related key differential attacks against the standard block cipher AES. We improved our models for those attacks and the time required to solve the related key differential attacks for all instances of this particular problem. In particular, we were able to find the best related key differential trails for all the instances of AES-128, AES-192 and AES-256 in less than 5 core-hours except for one instance (AES-128 with 5 rounds) that took 15 core-hours.

7.14. Preparation of a submission for the NIST call dedicated to standardization of lightweight cryptography

Participants: Marine Minier [contact], Paul Huynh, Virginie Lallemand.

During these last six months, we prepared a submission to the NIST call dedicated on lightweight cryptography. The criteria required by this call are various and concern both small embedded micro-controllers and efficient hardware implementation with side channel and fault attack resistance. The proposal will be submitted by the call deadline, at the latest on Feb 25th, 2019.

Coast Project-Team

6. New Results

6.1. Design and Analysis of Collaborative Editing Approaches

Participants: Matthieu Nicolas, Victorien Elvinger, Hoai Le Nguyen, Quentin Laporte Chabasse, Claudia-Lavinia Ignat [contact], Gérald Oster, François Charoy, Olivier Perrin.

Since the Web 2.0 era, the Internet is a huge content editing place on which users collaborate. Thousand of people can edit this shared document. However, current consistency maintenance algorithms are not adapted to massive collaborative updating involving large number of contributors and a high velocity of changes. This year we studied collaborative editing user behaviour and started to work on an optimised solution for sequence CRDTs. Version control systems such as Git became very widespread in the open-source community. In these collaborative systems, conflict resolution that arise during synchronisation of parallel changes might become a burden for the user. We analysed concurrency and conflicts in Git repository of four projects: Rails, IkiWiki, Samba and Linux Kernel. We analysed the collaboration process of these projects at specific periods revealing how change integration and conflict rates vary during the project development life-cycle. Our study suggests that developers should use more intensively awareness mechanisms close to release dates where changes integration rate is higher. We also discussed the mechanism adopted by Git to consider concurrent changes made on two adjacent lines as conflicting. Based on the high rate of false positives of this mechanism, our study suggests that Git should reconsider signalling adjacent line conflicts inside the source code files [4]. Sequence Conflict-free Replicated Data Types (CRDTs) allow one to replicate and edit, without any kind of coordination, sequences in distributed systems. To ensure convergence, existing works from the literature add metadata to each element but they do not bind its footprint, which impedes their adoption. Several approaches were proposed to address this issue but they do not fit a fully distributed setting. We started to work on the design and validation of a fully distributed renaming mechanism, setting a bound to the metadata's footprint [14]. Addressing this issue opens new perspectives of adoption of these CRDTs in distributed applications.

6.2. Trustworthy Collaboration

Participants: Claudia-Lavinia Ignat, Victorien Elvinger, François Charoy, Olivier Perrin, Gérald Oster, Hoang Long Nguyen.

Trust between users is an important factor for the success of a collaboration. Users might want to collaborate only with those users they trust. We are interested in assessing users trust according to their behaviour during collaboration in a large scale environment. We studied the trust assessment problem and designed a computational trust model for collaborative systems [1]. We also studied how to predict the trust relation between users that did not interact in the past. Given a network in which the links represent the trust/distrust relations between users, we aimed to predict future relations. We proposed a link-sign prediction algorithm [6] that does not require full graph information, is suitable for dynamic networks and takes into account the creation time of the links in the network. Our solution combines state-of-the-art techniques in natural language processing (Doc2Vec [25]) and deep learning (Recurrent Neural Networks [31] with Long-Short Term Memory [24]) with the random walk graph sampling [26]. Our algorithm outperforms state-of-the-art approaches on real world signed directed social network datasets. In distributed collaborative systems, participants maintain a replicated copy of shared documents. They edit their own copy and then share their modifications without any coordination. Copies follow successions of divergence and convergence. Convergence is a liveness property of collaborative systems. Some malicious participants may find an advantage to make the collaboration fail. To that end, they can preclude convergence of the copies. To protect convergence of copies, participants can exploit an authenticated log of modifications. New participants have to retrieve the entire log in order to contribute. Unfortunately, the cost of joining a collaboration increases with the size of this log. Causal Stability allows to prune authenticated logs in a static collaborative group without

any malicious participants. We tailored Causal Stability to dynamic groups in the presence of malicious participants. We also proposed a mechanism to verify the consistency of a pruned log and a mechanism to authenticate a snapshot from a pruned log [7]. Public key server is a simple yet effective way of key management in secure end-to-end communication. To ensure the trustworthiness of a public key server, CONIKS [27] employs a tamper-evident data structure on the server and a gossiping protocol among clients in order to detect compromised servers. However, due to lack of incentive and vulnerability to malicious clients, a gossiping protocol is hard to implement in practice. Meanwhile, alternative solutions such as EthIKS [21] are too costly. We proposed Trusternity [13], [12], an auditing scheme relying on Ethereum blockchain that is easy to implement, inexpensive to operate and resilient to malicious clients. We also conducted an empirical study of system behaviour in face of attacks and proposed a lightweight anomaly detection algorithm to protect clients against such attacks.

6.3. Trust and data sharing in crisis management

Participants: François Charoy, Béatrice Linot, Valerie Shalin.

Sharing information between responders is important during crisis management response. Tools and platforms are eagerly developed for that purpose. They are supposed to support people and help them to build a shared situation awareness. However as the scale of crisis increases and as more and more organisations are involved, people get reluctant to use them to share their data. They prefer to rely on one to one communication tools like phones or text. This is why we are studying how these collaborative platforms impact the work of responders positively or negatively. We want to know why most of the time they don't want to use them for their original purpose. We studied reports on past incidents [17], [10] and conducted extensive analysis of the use of existing systems (e.g. the French platform CRISORSEC) through interviews, observation and data analysis. [11]

6.4. Cloud Provisioning for Elastic BPM

Participants: François Charoy, Samir Youcef, Guillaume Rosinosky.

Cloud computing providers do not help consumers to use optimally the available resources. Several approaches have been proposed [33] that take benefit from the elasticity of the Cloud, starting and stopping virtual machines on demand. They suffer from several shortcomings. Often they consider only one objective, the reduction of the cost, or a level of quality of service. We proposed to optimise two conflicting objectives, the number of migrations of tenants that is helpful to reach the optimal cost and the cost incurred considering a set of resources. Our approach allows to take into account the multi-tenancy property and the Cloud computing elasticity, and is efficient as shown by an extensive experimentation based on real data from Bonita BPM customers. In the continuation of our previous work we proposed and validated a more efficient algorithm for elastic execution of processes in the cloud [16]. To ensure a realistic validation, we collaborated with colleagues from the University of Lugano to set up a benchmarking platform in order to evaluate the impact of migration in a multi-tenant setting. This allowed us to execute reproducible experiments and to validate our hypothesis regarding the effect of migration and the parameters that affect them [15]. This platform is now an asset that can be used for all kinds of live migration experiments of software architectures.

6.5. Risk Management for the Deployment of a Business Process in a Multi-Cloud Context

Participants: Amina Ahmed Nacer, Claude Godart, Samir Youcef.

The lack of trust in cloud organisations 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 categorises 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 formalises 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 optimised configuration. It is implemented in a tool to assess cloud providers and decompose processes. Rooted in past years work, we are contributing this year at the methodological and framework levels in two directions:

- At the methodological level, while our risk computing model rested previously only on data provided by cloud providers (provider-side risk model), we are developing a risk model integrating client-side knowledge (client-side risk model).
- Additionally are developing a simulation tool for supporting designer decision with the ability to balance risk with cost when selecting the best cloud configuration [2].

6.6. Scheduling and Resource Allocation in Business Processes

Participants: Khalid Benali, Abir Ismaili-Alaoui.

Business Process Management (BPM) is concerned with continuously enhancing business processes by adapting a systematic approach that enables companies to increase the performance of their existing business processes and achieve their business goals. Business processes are generally considered as blind, stateless and reactive. This means that in each business process execution we do not take into consideration either the results from last process instances nor the context (for most cases). The rise of new technologies such as big and fast data, cloud computing, Internet of Things (IoT), etc, implies new business process scheduling problems. They are linked to limited resources (human and/or machine) or the need to use resources in an optimal and exible way. In order to avoid either under-provisioning (when there is an underestimation for the needed resources, business processes may not be executed) or over-provisioning (the resources planned in advance to cover peak times demands were not used in non-peak time) and also to take into consideration the priority level of each business process instances.

GAMBLE Project-Team

7. New Results

7.1. Non-Linear Computational Geometry

Participants: Sény Diatta, Laurent Dupont, George Krait, Sylvain Lazard, Guillaume Moroz, Marc Pouget.

7.1.1. Reliable location with respect to the projection of a smooth space curve

Consider a plane curve $\mathcal{B}$ defined as the projection of the intersection of two analytic surfaces in $\mathbb{R}^3$ or as the apparent contour of a surface. In general, $\mathcal{B}$ has node or cusp singular points and thus is a singular curve. Our main contribution [6] is the computation of a data structure for answering point location queries with respect to the subdivision of the plane induced by $\mathcal{B}$. This data structure is composed of an approximation of the space curve together with a topological representation of its projection $\mathcal{B}$. Since $\mathcal{B}$ is a singular curve, it is challenging to design a method only based on reliable numerical algorithms.

In a previous work [49], we have shown how to describe the set of singularities of $\mathcal{B}$ as regular solutions of a so-called ball system suitable for a numerical subdivision solver. Here, the space curve is first enclosed in a set of boxes with a certified path-tracker to restrict the domain where the ball system is solved. Boxes around singular points are then computed such that the correct topology of the curve inside these boxes can be deduced from the intersections of the curve with their boundaries. The tracking of the space curve is then used to connect the smooth branches to the singular points. The subdivision of the plane induced by $\mathcal{B}$ is encoded as an extended planar combinatorial map allowing point location. We experimented our method and showed that our reliable numerical approach can handle classes of examples that are not reachable by symbolic methods.

7.1.2. Workspace, Joint space and Singularities of a family of Delta-Like Robots

Our paper [7] presents the workspace, the joint space and the singularities of a family of delta-like parallel robots by using algebraic tools. The different functions of the SIROPA library are introduced and used to estimate the complexity representing the singularities in the workspace and the joint space. A Groebner 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 tools describing the shape of the workspace and of the joint space which will help 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.

In collaboration with Ranjan Jha, Damien Chablat, Luc Baron and Fabrice Rouillier.

7.2. Non-Euclidean Computational Geometry

Participants: Vincent Despré, Iordan Iordanov, Monique Teillaud.

7.2.1. Delaunay Triangulations of Symmetric Hyperbolic Surfaces

We have worked on extending our previous results on the computation of Delaunay triangulations of the Bolza surface [50] (see also the section New Software above), which is the most symmetric surface of genus 2. Elaborating further on previous work [26], we are now considering symmetric hyperbolic surfaces of higher genus, for which we study mathematical properties [14] that allow us to propose algorithms [13].

In collaboration with Gert Vegter and Matthijs Ebbens (University of Groningen).

7.3. Probabilistic Analysis of Geometric Data Structures and Algorithms

Participants: Olivier Devillers, Charles Duménil, Fernand Kuiebove Pefireko.

7.3.1. Stretch Factor in a Planar Poisson-Delaunay Triangulation with a Large Intensity

Let $X := X_n \cup \{(0,0),(1,0)\}$, where X_n is a planar Poisson point process of intensity n. Our paper [4] provides a first non-trivial lower bound for the expected length of the shortest path between $(0,0)$ and $(1,0)$ in the Delaunay triangulation associated with X when the intensity of X_n goes to infinity. Simulations indicate that the correct value is about 1.04. We also prove that the expected length of the so-called upper path converges to $\frac{35}{3\pi^2}$, giving an upper bound for the expected length of the smallest path.

In collaboration with Nicolas Chenavier (Université du Littoral Côte d'Opale).

7.3.2. Delaunay triangulation of a Poisson Point Process on a Surface

The complexity of the Delaunay triangulation of n points distributed on a surface ranges from linear to quadratic. We proved that when the points are evenly distributed on a smooth compact generic surface the expected size of the Delaunay triangulation is can be controlled. If the point set is a good sample of a smooth compact generic surface [22] the complexity is controlled. Namely, good sample means that a sphere of size ϵ centered on the surface contains between 1 and η points. Under this hypothesis, the complexity of the Delaunay triangulation is $O\left(\frac{\eta^2}{\epsilon^2}\log\frac{1}{\epsilon}\right)$. We proved that when the points are evenly distributed on a smooth compact generic surface they form a good sample with high probability for relevant values of ϵ and η. We can deduce [15] that the expected size of the Delaunay triangulation of n random points of a surface is $O(n\log^2 n)$.

7.3.3. On Order Types of Random Point Sets

Let P be a set of n random points chosen uniformly in the unit square. In our paper [19], we examine the typical resolution of the order type of P. First, we showed that with high probability, P can be rounded to the grid of step $\frac{1}{n^{3+\epsilon}}$ without changing its order type. Second, we studied algorithms for determining the order type of a point set in terms of the the number of coordinate bits they require to know. We gave an algorithm that requires on average $4n\log_2 n + O(n)$ bits to determine the order type of P, and showed that any algorithm requires at least $4n\log_2 n - O(n\log\log n)$ bits. Both results extend to more general models of random point sets.

In collaboration with Philippe Duchon (LABRI) and Marc Glisse (project team DATASHAPE *).*

7.4. Classical Computational Geometry and Graph Drawing

Participants: Vincent Despré, Olivier Devillers, Sylvain Lazard.

7.4.1. Delaunay Triangulations of Points on Circles

Delaunay triangulations of a point set in the Euclidean plane are ubiquitous in a number of computational sciences, including computational geometry. Delaunay triangulations are not well defined as soon as 4 or more points are concyclic but since it is not a generic situation, this difficulty is usually handled by using a (symbolic or explicit) perturbation. As an alternative, we proposed to define a canonical triangulation for a set of concyclic points by using a max-min angle characterization of Delaunay triangulations. This point of view leads to a well defined and unique triangulation as long as there are no symmetric quadruples of points. This unique triangulation can be computed in quasi-linear time by a very simple algorithm [18].

In collaboration with Hugo Parlier and Jean-Marc Schlenker (University of Luxembourg).

7.4.2. Improved Routing on the Delaunay Triangulation

A geometric graph $G = (P, E)$ is a set of points in the plane and edges between pairs of points, where the weight of each edge is equal to the Euclidean distance between the corresponding points. In k-local routing we find a path through G from a source vertex s to a destination vertex t, using only knowledge of the present location, the locations of s and t, and the k-neighbourhood of the current vertex. We presented [11] an algorithm for 1-local routing on the Delaunay triangulation, and show that it finds a path between a source vertex s and a target vertex t that is not longer than $3.56|st|$, improving the previous bound of 5.9.

In collaboration with Nicolas Bonichon (Labri), Prosenjit Bose, Jean-Lou De Carufel, Michiel Smid and Daryl Hill (Carleton University)

7.4.3. Limits of Order Types

We completed an extended version of a work published at SoCG 2015, in which we apply ideas from the theory of limits of dense combinatorial structures to study order types, which are combinatorial encodings of finite point sets. Using flag algebras we obtain new numerical results on the Erdös problem of finding the minimal density of 5-or 6-tuples in convex position in an arbitrary point set, and also an inequality expressing the difficulty of sampling order types uniformly. Next we establish results on the analytic representation of limits of order types by planar measures. Our main result is a rigidity theorem: we show that if sampling two measures induce the same probability distribution on order types, then these measures are projectively equivalent provided the support of at least one of them has non-empty interior. We also show that some condition on the Hausdorff dimension of the support is necessary to obtain projective rigidity and we construct limits of order types that cannot be represented by a planar measure. Returning to combinatorial geometry we relate the regularity of this analytic representation to the aforementioned problem of Erdös on the density of k-tuples in convex position, for large k [20].

In collaboration with Alfredo Hubard (Laboratoire d'Informatique Gaspard-Monge) Rémi De Joannis de Verclos (Radboud university, Nijmegen) Jean-Sébastien Sereni (CNRS) Jan Volec (Department of Mathematics and Computer Science, Emory University)

7.4.4. Snap rounding polyhedral subdivisions

Let $\mathcal{P}$ be a set of n polygons in $\mathbb{R}^3$, each of constant complexity and with pairwise disjoint interiors. We propose a rounding algorithm that maps $\mathcal{P}$ to a simplicial complex $\mathcal{Q}$ whose vertices have integer coordinates. Every face of $\mathcal{P}$ is mapped to a set of faces (or edges or vertices) of $\mathcal{Q}$ and the mapping from $\mathcal{P}$ to $\mathcal{Q}$ can be build through a continuous motion of the faces such that (i) the L_∞ Hausdorff distance between a face and its image during the motion is at most 3/2 and (ii) if two points become equal during the motion they remain equal through the rest of the motion. In the worse case, the size of $\mathcal{Q}$ is $O(n^{15})$, but we conjecture a good complexity of $O(n\sqrt{n})$ in practice on non-pathological data [12].

In collaboration with William J. Lenhart (Williams College, USA).

7.4.5. On the Edge-length Ratio of Outerplanar Graphs

We show that any outerplanar graph admits a planar straight-line drawing such that the length ratio of the longest to the shortest edges is strictly less than 2. This result is tight in the sense that for any $\epsilon > 0$ there are outerplanar graphs that cannot be drawn with an edge-length ratio smaller than $2 - \epsilon$. We also show that this ratio cannot be bounded if the embeddings of the outerplanar graphs are given [9].

In collaboration with William J. Lenhart (Williams College, USA) and Giuseppe Liotta (Università di Perugia, Italy).

LARSEN Project-Team

7. New Results

7.1. Lifelong Autonomy

7.1.1. Foundations of Reinforcement Learning

7.1.1.1. ρ-POMDPs have Lipschitz-Continuous ϵ-Optimal Value Functions

Participant: Vincent Thomas.

Collaboration with Jilles Dibangoye (INSA Lyon).

Many state-of-the-art algorithms for solving Partially Observable Markov Decision Processes (POMDPs) rely on turning the problem into a "fully observable" problem—a belief MDP—and exploiting the piece-wise linearity and convexity (PWLC) of the optimal value function in this new state space (the belief simplex Δ). This approach has been extended to solving ρ-POMDPs—i.e., for information-oriented criteria-when the reward ρ is convex in Δ. General ρ-POMDPs can also be turned into "fully observable" problems, but with no means to exploit the PWLC property. In this paper, we focus on POMDPs and ρ-POMDPs with $\lambda\rho$-Lipschitz reward function, and demonstrate that, for finite horizons, the optimal value function is Lipschitz-continuous. Then, value function approximators are proposed for both upper-and lower-bounding the optimal value function, which are shown to provide uniformly improvable bounds. This allows proposing two algorithms derived from HSVI which are empirically evaluated on various benchmark problems.

Publication: [14]

7.1.1.2. Addressing Active Sensing Problem through MCTS

Participants: Vincent Thomas, Geremy Hutin.

The problem of active sensing is of paramount interest for building self awareness in robotic systems. It consists of a system to make decisions in order to gather information (measured through the entropy of the probability distribution over unknown variables) in an optimal way.

In the past, we have proposed an original formalism *ρ-POMDP* and new algorithms for representing and solving active sensing problems [33] by using point-based algorithms. This year, new approaches based on Monte-Carlo Tree Search algorithms (MCTS) and Partially Observable Monte-Carlo Planning (POMCP) [45] have been proposed to build the policies of an agent whose aim is to gather information.

7.1.2. Robot Learning

Our main objective is to design data-efficient trial-and-error learning algorithms (reinforcement learning) that can work with continuous states and continuous actions. The main use-case is robot damage recovery: a robot has to to discover new behaviors by trial-and-error without a diagnosis of the damage.

7.1.2.1. Adaptive and Resilient Soft Tensegrity Robots

Participant: Jean-Baptiste Mouret.

Collaboration with John Rieffel (Union College, USA).

Living organisms intertwine soft (e.g., muscle) and hard (e.g., bones) materials, giving them an intrinsic flexibility and resiliency often lacking in conventional rigid robots. The emerging field of soft robotics seeks to harness these same properties to create resilient machines. The nature of soft materials, however, presents considerable challenges to aspects of design, construction, and control—and up until now, the vast majority of gaits for soft robots have been hand-designed through empirical trial-and-error. In this contribution, we introduced an easy-to-assemble tensegrity-based soft robot capable of highly dynamic locomotive gaits and demonstrating structural and behavioral resilience in the face of physical damage. Enabling this is the use of a machine learning algorithm able to discover effective gaits with a minimal number of physical trials. These results lend further credence to soft-robotic approaches that seek to harness the interaction of complex material dynamics to generate a wealth of dynamical behaviors.

Publication: [10]

7.1.2.2. Bayesian Optimization with Automatic Prior Selection for Data-Efficient Direct Policy Search
Participants: Konstantinos Chatzilygeroudis, Jean-Baptiste Mouret.

One of the most interesting features of Bayesian optimization for direct policy search is that it can leverage priors (e.g., from simulation or from previous tasks) to accelerate learning on a robot. In this contribution, we are interested in situations for which several priors exist but we do not know in advance which one fits best the current situation. We tackle this problem by introducing a novel acquisition function, called Most Likely Expected Improvement (MLEI), that combines the likelihood of the priors and the expected improvement. We evaluate this new acquisition function on a transfer learning task for a 5-DOF planar arm and on a possibly damaged, 6-legged robot that has to learn to walk on flat ground and on stairs, with priors corresponding to different stairs and different kinds of damages. Our results show that MLEI effectively identifies and exploits the priors, even when there is no obvious match between the current situations and the priors.

Publication: [23]

7.1.2.3. Multi-objective Model-based Policy Search for Data-efficient Learning with Sparse Rewards
Participants: Rituraj Kaushik, Konstantinos Chatzilygeroudis, Jean-Baptiste Mouret.

The most data-efficient algorithms for reinforcement learning in robotics are model-based policy search algorithms, which alternate between learning a dynamical model of the robot and optimizing a policy to maximize the expected return given the model and its uncertainties. However, the current algorithms lack an effective exploration strategy to deal with sparse or misleading reward scenarios: if they do not experience any state with a positive reward during the initial random exploration, they are very unlikely to solve the problem. To address this challenge, we proposed a novel model-based policy search algorithm, Multi-DEX, that leverages a learned dynamical model to efficiently explore the task space and solve tasks with sparse rewards in a few episodes. To achieve this, we frame the policy search problem as a multi-objective, model-based policy optimization problem with three objectives: (1) generate maximally novel state trajectories, (2) maximize the cumulative reward and (3) keep the system in state-space regions for which the model is as accurate as possible. We then optimize these objectives using a Pareto-based multi-objective optimization algorithm. The experiments show that Multi-DEX is able to solve sparse reward scenarios (with a simulated robotic arm) in much lower interaction time than VIME, TRPO, GEP-PG, CMA-ES and Black-DROPS.

Publication: [18]

7.1.2.4. Using Parameterized Black-Box Priors to Scale Up Model-Based Policy Search for Robotics
Participants: Konstantinos Chatzilygeroudis, Jean-Baptiste Mouret.

Among the few model-based policy search algorithms, the recently introduced Black-DROPS algorithm exploits a black-box optimization algorithm to achieve both high data-efficiency and good computation times when several cores are used; nevertheless, like all model-based policy search approaches, Black-DROPS does not scale to high dimensional state/action spaces. In this paper, we introduce a new model learning procedure in Black-DROPS that leverages parameterized black-box priors to (1) scale up to high-dimensional systems, and (2) be robust to large inaccuracies of the prior information. We demonstrate the effectiveness of our approach with the "pendubot" swing-up task in simulation and with a physical hexapod robot (48D state space, 18D action space) that has to walk forward as fast as possible. The results show that our new algorithm is more data-efficient than previous model-based policy search algorithms (with and without priors) and that it can allow a physical 6-legged robot to learn new gaits in only 16 to 30 seconds of interaction time.

Publication: [12]

7.1.2.5. Data-efficient Neuroevolution with Kernel-Based Surrogate Models
Participants: Adam Gaier, Jean-Baptiste Mouret.

Collaboration with Alexander Asteroth (Hochschule Bonn-Rhein-Sieg, Germany)

Surrogate-assistance approaches have long been used in computationally expensive domains to improve the data-efficiency of optimization algorithms. Neuroevolution, however, has so far resisted the application of these techniques because it requires the surrogate model to make fitness predictions based on variable topologies, instead of a vector of parameters. Our main insight is that we can sidestep this problem by using kernel-based surrogate models, which require only the definition of a distance measure between individuals. Our second insight is that the well-established Neuroevolution of Augmenting Topologies (NEAT) algorithm provides a computationally efficient distance measure between dissimilar networks in the form of "compatibility distance", initially designed to maintain topological diversity. Combining these two ideas, we introduce a surrogate-assisted neuroevolution algorithm that combines NEAT and a surrogate model built using a compatibility distance kernel. We demonstrate the data-efficiency of this new algorithm on the low dimensional cart-pole swing-up problem, as well as the higher dimensional half-cheetah running task. In both tasks the surrogate-assisted variant achieves the same or better results with several times fewer function evaluations as the original NEAT.

Publication: [17] (best paper, GECCO 2018, Complex System track)

7.1.2.6. Alternating Optimization and Quadrature for Robust Control

Participants: Konstantinos Chatzilygeroudis, Jean-Baptiste Mouret.

Collaboration with Shimon Whiteson (Oxford, UK).

Bayesian optimization has been successfully applied to a variety of reinforcement learning problems. However, the traditional approach for learning optimal policies in simulators does not utilise the opportunity to improve learning by adjusting certain environment variables — state features that are randomly determined by the environment in a physical setting but are controllable in a simulator. In this wwork, we consider the problem of finding an optimal policy while taking into account the impact of environment variables. We present alternating optimization and quadrature (ALOQ), which uses Bayesian optimization and Bayesian quadrature to address such settings. ALOQ is robust to the presence of significant rare events, which may not be observable under random sampling, but have a considerable impact on determining the optimal policy. The experimental results demonstrate that our approach learns more efficiently than existing methods.

Publication: [22]

7.1.2.7. Learning robust task priorities of QP-based whole-body torque-controllers

Participants: Marie Charbonneau, Serena Ivaldi, Valerio Modugno, Jean-Baptiste Mouret.

Generating complex whole-body movements for humanoid robots is now most often achieved with multi-task whole-body controllers based on quadratic programming. To perform on the real robot, such controllers often require a human expert to tune or optimize the many parameters of the controller related to the tasks and to the specific robot, which is generally reported as a tedious and time consuming procedure. This problem can be tackled by automatically optimizing some parameters such as task priorities or task trajectories, while ensuring constraints satisfaction, through simulation. However, this does not guarantee that parameters optimized in simulation will also be optimal for the real robot. As a solution, the present paper focuses on optimizing task priorities in a robust way, by looking for solutions which achieve desired tasks under a variety of conditions and perturbations. This approach, which can be referred to as domain randomization, can greatly facilitate the transfer of optimized solutions from simulation to a real robot. The proposed method is demonstrated using a simulation of the humanoid robot iCub for a whole-body stepping task.

Publication: [11]

7.1.3. Quality Diversity Algorithms

Quality diversity algorithms are a new kind of evolutionary algorithms that focuses on finding a large set of high-performing solutions (instead of the global optimum). We use them for design and as a step for data-efficient robot learning.

7.1.3.1. Data-Efficient Design Exploration through Surrogate-Assisted Illumination

Participants: Adam Gaier, Jean-Baptiste Mouret.

Collaboration with Alexander Asteroth (Hochschule Bonn-Rhein-Sieg, Germany)

Design optimization techniques are often used at the beginning of the design process to explore the space of possible designs. In these domains illumination algorithms, such as MAP-Elites, are promising alternatives to classic optimization algorithms because they produce diverse, high-quality solutions in a single run, instead of only a single near-optimal solution. Unfortunately, these algorithms currently require a large number of function evaluations, limiting their applicability. In this work, we introduce a new illumination algorithm, Surrogate-Assisted Illumination (SAIL), that leverages surrogate modeling techniques to create a map of the design space according to user-defined features while minimizing the number of fitness evaluations. On a 2-dimensional airfoil optimization problem SAIL produces hundreds of diverse but high-performing designs with several orders of magnitude fewer evaluations than MAP-Elites or CMA-ES. We demonstrate that SAIL is also capable of producing maps of high-performing designs in realistic 3-dimensional aerodynamic tasks with an accurate flow simulation. Data-efficient design exploration with SAIL can help designers understand what is possible, beyond what is optimal, by considering more than pure objective-based optimization.

Publication: [7]

7.1.3.2. Discovering the Elite Hypervolume by Leveraging Interspecies Correlation

Participants: Vassilis Vassiliades, Jean-Baptiste Mouret.

Evolution has produced an astonishing diversity of species, each filling a different niche. Algorithms like MAP-Elites mimic this divergent evolutionary process to find a set of behaviorally diverse but high-performing solutions, called the elites. Our key insight is that species in nature often share a surprisingly large part of their genome, in spite of occupying very different niches; similarly, the elites are likely to be concentrated in a specific "elite hypervolume" whose shape is defined by their common features. In this paper, we first introduce the elite hypervolume concept and propose two metrics to characterize it: the genotypic spread and the genotypic similarity. We then introduce a new variation operator, called “directional variation”, that exploits interspecies (or inter-elites) correlations to accelerate the MAP-Elites algorithm. We demonstrate the effectiveness of this operator in three problems (a toy function, a redundant robotic arm, and a hexapod robot).

Publication: [25]

7.1.3.3. Maintaining Diversity in Robot Swarms with Distributed Embodied Evolution

Participants: Amine Boumaza, François Charpillet.

We investigated how behavioral diversity can be maintained in evolving robot swarms by using distributed Embodied Evolution. In these approaches, each robot in the swarm runs a separate evolutionary algorithm, and populations on each robot are built through local communication when robots meet; therefore, genome survival results not only from fitness-based selection but also from spatial spread. To better understand how diversity is maintained in distributed embodied evolution, we propose a postanalysis diversity measure — global diversity (over the swarm), and local diversity (on each robot) —, on two swarm robotic tasks — navigation and item collection —, with different intensities of selection pressure, and compare the results of distributed embodied evolution to a centralized case. We conclude that distributed evolution intrinsically maintains a larger behavioral diversity when compared to centralized evolution, which allows for the search algorithm to reach higher performances, especially in the more challenging collection task.

Publication: [16]

7.2. Natural Interaction with Robotics Systems

7.2.1. Control of Interaction

Because of the AnDy project, we are currently focused on interaction in industrial contexts, in particular to encourage ergonomic motions.

7.2.1.1. Robust Real-time Whole-Body Motion Retargeting from Human to Humanoid

Participants: Serena Ivaldi, Luigi Penco, Brice Clement, Jean-Baptiste Mouret.

Transferring the motion from a human operator to a humanoid robot is a crucial step to enable robots to learn from and replicate human movements. The ability to retarget in real-time whole-body motions that are challenging for the humanoid balance is critical to enable human to humanoid teleoperation. In this work, we design a retargeting framework that allows the robot to replicate the motion of the human operator, acquired by a wearable motion capture suit, while maintaining the whole-body balance. We introduce some dynamic filter in the retargeting to forbid dangerous motions that can make the robot fall. We validate our approach through several experiments on the iCub robot, which has a significantly different body structure and size from the one of the human operator.

Publication: [24]

7.2.1.2. Prediction of Human Whole-Body Movements with AE-ProMPs

Participants: Serena Ivaldi, Oriane Dermy, Francis Colas, François Charpillet.

The ability to predict intended movements is crucial for collaborative robots to anticipate the human actions and for assistive technologies to alert if a particular movement is non-ergonomic and potentially dangerous for humans. In this paper, we address the problem of predicting the future human whole-body movements given early observations. We propose to predict the continuation of the high-dimensional trajectories mapped into a reduced latent space, using autoencoders (AE). The prediction is based on a probabilistic description of the movement primitives (ProMPs) in the latent space, which notably reduces the computational time for the prediction to occur, and hence enables to use the method in real-time applications. We evaluate our method, named AE-ProMPs, for predicting future movements belonging to a dataset of 7 different actions performed by a human, recorded by a wearable motion tracking suit.

Publication: [13]

Publications: [28],

7.2.2. Generating Assistive Humanoid Motions for Co-Manipulation Tasks with a Multi-Robot Quadratic Program Controller

Participants: Karim Bouyarmane, Serena Ivaldi.

Human-humanoid collaborative tasks require that the robot takes into account the goals of the task, interaction forces with the human, and its own balance. We present a formulation for a real-time humanoid controller which allows the robot to keep itself stable, while also assisting the human in achieving their shared objectives. This is achieved with a multi-robot quadratic program controller, which solves for human motion reconstruction and optimal robot controls in a single optimization problem. Our experiments on a simulated robot platform demonstrate the ability to generate interactions motions and forces that are similar to what a human collaborator would produce.

Publication: [21]

7.2.2.1. Activity Recognition With Multiple Wearable Sensors for Industrial Applications

Participants: Francis Colas, Serena Ivaldi, Adrien Malaisé, Pauline Maurice, François Charpillet.

We address the problem of recognizing the current activity performed by a human operator, providing an information useful for automatic ergonomic evaluation for industrial applications. While the majority of research in activity recognition relies on cameras observing the human, here we explore the use of wearable sensors, which are more suitable in industrial environments. We use a wearable motion tracking suit and a sensorized glove. We describe our approach for activity recognition with a probabilistic model based on Hidden Markov Models, applied to the problem of recognizing elementary activities during a pick-and-place task inspired by a manufacturing scenario. We show that our model is able to correctly recognize the activities with 96% of precision if both sensors are used.

Publication: [19],[19]

7.2.2.2. Activity Recognition for monitoring eldely people at home

Participants: Yassine El Khadiri, François Charpillet.

Early detection of frailty signs is important for senior people who prefer to keep living in their homes instead of moving to a nursing home. Sleep quality is a good predictor for frailty monitoring. Thus we are interested in tracking sleep parameters like sleep wake patterns to predict and detect potential sleep disturbances of the monitored senior residents. We use an unsupervised inference method based on actigraphy data generated by ambient motion sensors scattered around the senior's apartment. This enables our monitoring solution to be flexible and robust to the different types of housings it can equip while still attaining accuracy of 0.94 for sleep period estimates.

Publication: [15]

7.2.3. Ethics

7.2.3.1. Ethical and Social Considerations for the Introduction of Human-Centered Technologies at Work

Participants: Serena Ivaldi, Adrien Malaisé, Pauline Maurice, Ludivine Allienne.

Human-centered technologies such as collaborative robots, exoskeletons, and wearable sensors are rapidly spreading in industry and manufacturing because of their intrinsic potential at assisting workers and improving their working conditions. The deployment of these technologies, albeit inevitable, poses several ethical and societal issues. Guidelines for ethically aligned design of autonomous and intelligent systems do exist, however we argue that ethical recommendations must necessarily be complemented by an analysis of the social impact of these technologies.

In a recent paper[20], we report on our preliminary studies on the opinion of factory workers and of people outside this environment on human-centered technologies at work. In light of these studies, we discuss ethical and social considerations for deploying these technologies in a way that improves acceptance.

Publication: [20]

MAGRIT Project-Team

7. New Results

7.1. Matching and localization

Participants: Marie-Odile Berger, Vincent Gaudilliere, Antoine Fond, Gilles Simon.

Vanishing point detection

Accurate detection of vanishing points (VPs) is a prerequisite for many computer vision problems such as camera self-calibration, single view structure recovery, video compass, robot navigation and augmented reality, among many others. More specifically, knowing three orthogonal VPs aligned with the buildings of a scene (the Manhattan directions) allows computing the intrinsic parameters of the camera as well as warped images where the buildings' facades are orthorectified, facilitating their detection and registration. VPs are also used in our work on epipolar geometry estimation to help matching line segments, a particularly difficult task in low-textured environments.

We introduced an *a-contrario* method to solve this problem. Our key contribution was to show that, as soon as the horizon line (HL) is inside the image boundaries, this line can usually be detected as an alignment of oriented line segments. This comes from a simple geometric property, that any horizontal line segment at the height of the camera's optical center projects to the HL regardless of its 3-D direction. This property generally yields statistically meaningful events, detectable from *a-contrario* analysis. Additional candidate HLs are sampled around these events using a Gaussian Mixture Model (GMM), and scored according to the strongest of the VPs hypothesized along them. VP hypotheses are also obtained from an *a-contrario* method, using integral geometry to accurately model the background noise. Experiments made on three urban datasets showed that our method, not only achieves state-of-the-art performance w.r.t. computation times and accuracy of the HL, but also yields much less spurious VPs than the previous top-ranked methods. This work was published at ECCV'2018 [23] and an article is in preparation for submission in a peer-reviewed journal. In this article, we show that our method also outperforms state-of-the-art methods on a new industrial dataset that we built and will make publicly available. We also establish a relation between the Number of False Alarms (NFA) obtained for the meaningful events and the spreads of the GMM. In addition, the Matlab code implementing our method has been made publicly available.

Urban AR

Urban localization plays a major role in many applications including navigation aid, labeling of local touristic landmarks, and robot localization. The outdoor accuracy of mobile phone GPS is only 12.5 meters and can be worse in urban areas where the street is flanked by buildings on both sides. By contrast, buildings' facades are meaningful landmarks to rely on for large-scale localization. Last year, we proposed a method to automatically detect facades in an image, based on image cues that measure facade characteristics such as shape, color, contours, semantic structure and symmetry. Matching the detected facade with a facade database using a metric learned through a siamese neural network allowed us to estimate a first initialization of the registration parameters by solving the least-square problem that maps the four transformed corners of the reference to the four corners of the detection.

This year, we attempted to rely on semantic segmentation to improve the accuracy of that initial registration [11]. Simultaneously, we aimed to iteratively improve the quality of the semantic segmentation through registration. Registration and semantic segmentation were jointly solved in a Expectation-Maximization framework. We especially introduced a Bayesian model that uses prior semantic segmentation as well as geometric structure of the facade reference modeled by Generalized Gaussian Mixtures. We showed the advantages of our method in terms of robustness to clutter and change of illumination on urban images from various databases. We currently are assessing the relevance of the method using the large scale dataset SFM Aachen, in order to compare it with state-of-the-art SFM-based localization.

AR in industrial environments
Industrial environments are normally inundated with textureless objects, specular surfaces, repetitive objects and artificial lights, etc. which may fail traditional 2D/3D matching-based approaches. Line segments are numerous in industrial environments, but contrary to what happens in urban scenes, matching is a tough issue since most segments are silhouette contours whose appearance is viewpoint dependant. The combinatory of segment matches is thus very high, making impossible in practice the use of RANSAC algorithms for pose computation.

Within V. Gaudilliere's PhD thesis [21], [25], we took advantage of global properties of the environment, both geometric - such as the presence of numerous vertical planes - and contextual to guide matching. First, sub-image correspondences based on high level ConvNet features are used as prior for vertical planes detection and matching. Then, local homographies are detected between matched regions. To ensure efficient estimations, we have developed a dedicated RANSAC framework in which model hypotheses are first generated based on vanishing point and visual keypoint correspondences, and then validated on key points and line segments. This potential set of matched features are finally filtered with a robust fundamental matrix estimation. That scheme enables us to circumvent problems encountered in poorly-textured images (sparsity of visual keypoints and difficulties to match segments) while taking advantage of the abundance of segments and vanishing points characteristic of industrial environments

7.2. Handling non-rigid deformations

Participants: Marie-Odile Berger, Jaime Garcia Guevara, Daryna Panicheva, Pierre-Frédéric Villard.

Elastic multi-modal registration
Our previous works about 3D tracking for deformable objects [1] are template-based methods and thus need to carefully register the model onto the image in the first image. In practice this task is performed manually and is especially difficult for deformable organs. Within J. Guevara's PhD thesis, we have proposed an automatic method for registering pre and per-operative imagery which exploits the matching of the vascular trees, visible in most pre and intra-operative images. Although methods dedicated to non-rigid graph registration exist, they are not efficient when large intra-operative deformations of tissues occur. Our contribution is an extension of the graph-matching algorithm based on Gaussian process regression (GPR) proposed in [28]. Our idea is to combine GPR with a biomechanical model of the organ. Indeed, GPR allows for rigorous and fast error propagation but is extremely versatile and may thus produce non physically coherent registration, while biomechanical transformations are slower to compute but are capable of handling non linear deformation while preserving their physical nature. They thus allow earlier incoherent matching hypotheses to be removed. Integrating the two approaches allows us to significantly improve the quality of the registration while reducing computation times. These contributions have been published in the IPCAI conference [20] and in the International Journal of Computer Assisted Radiology and Surgery [13].

Individual-specific heart valve modeling
We first finished up a feasibility study aiming at providing fast image-based mitral valve mechanical simulation from individualized geometry [19]. The method was demonstrated on one dataset which was interactively segmented. In order to extend the pipeline to any data, robust methods to segment the valve components are required. Within the context of D. Panicheva's PhD thesis, we are currently working on means allowing to automatically segment the chordae. Valve chordae are generalized cylinders: Instead of being limited to a line, the central axis is a continuous curve. Instead of a constant radius, the radius varies along the axis. Most of the time, chordae sections are flattened ellipses and classical model-based methods commonly used for vessel enhancement or vessel segmentation fail. In this contribution, we exploit the fact that there are no other generalized cylinders than the chordae in the CT scan and we propose a topology-based method for chordae extraction. This approach is flexible and only requires the knowledge of an upper bound of the maximum radius of the chordae. The method has been tested on three CT scans. Overall, non-chordae structures are correctly identified and detected chordae ending points match up with actual chordae attachment points.

INVIVE: The Individual Virtual Ventilator: Image-based biomechanical simulation of the diaphragm during mechanical ventilation

When intensive care patients are subjected to mechanical ventilation, the ventilator causes damage to the muscles that govern the normal breathing, leading to Ventilator Induced Diaphragmatic Dysfunction (VIDD). The INVIVE project aims to study the mechanics of respiration through numerical simulation in order to learn more about the onset of VIDD. We propose to use a meshfree RBF method. During this year, we have worked on building an implicit representation of the surface of the diaphragm based on the topology coming from last year researches. It has been associated with a linear elasticity model and the boundary conditions have been measured on landmark points extracted by medical experts.

3D catheter navigation from monocular images
In interventional radiology, the 3D shape of the micro-tool (guidewire, micro-catheter or micro-coil) can be very difficult, if not impossible to infer from fluoroscopy images, which may have an impact on the clinical outcome of the procedure. This question is considered as a single view 3D curve reconstruction problem. We follow a constrained non-rigid shape from motion approach, using a physics-based model as a prior for the object shape. The navigation model is implemented through interactive simulation that provides a prediction of the device, taking into account non-linear effects such as friction during contacts.

Raffaella Trivisonne started her PhD thesis in November 2015 (co-supervised by Stéphane Cotin, from MIMESIS team in Strasbourg) to address this research topic. An unscented kalman filter is used as a fusion mechanism of the model with image data (opaque markers placed along the device). Progress has been made this year towards an effective formulation of the filter. In particular, a good estimate is recovered for the device shape in the case of ambiguous views (overlapping anatomy, bifurcations).

The method has been implemented in Sofa simulation software platform. Validation has been performed on porcine in-vivo data, acquired in accordance with UE norms, in collaboration with Pr. Mario Gimenez and Dr. Alain Garcia from IHU-Strasbourg.

In-vivo procedures will be performed under ethical approval of MSER (reference to ethic protocol *APAFIS #15433-2018060815283960*).

Markerless similarity metrics were investigated during Juan Rocha's Master's thesis. The update equations of the filter were generalized to tackle curve to image similarity metrics, traditionally used in multi-view reconstruction methods.

7.3. Image processing

Participants: Gilles Simon, Fabien Pierre, Frédéric Sur.

Variational methods for image processing
In the previous decade, variational methods in image processing have been widely used with a huge number of applications. The convex hypothesis generally makes the proof of convergence easier, whereas it is not fulfilled by the most interesting problems in imaging. The non-convexity may appear in some applications such as image colorization with multiple candidates selection [27] or in the case of M-estimators computation, in particular with an assumption of Cauchy noise [16]. These two points of view of the non-convex variational methods bring two different mathematical challenges to ensure the convergence of the numerical scheme. The choice of one candidate among a collection of possible ones implies bi-convex functions (functions with multiple variables, convex with respect to each ones). The computation of M-estimators with Cauchy noise hypothesis implies smooth but non-convex functions.

Our contributions concern both types of non-convexity. For bi-convex functions, we have demonstrated in [27] the convergence of alternate gradient descent numerical scheme with inertial relaxation of the iterates. Moreover, an application to image colorization has been proposed. In [16], a fixed-point algorithm has been studied to solve the problem of the Myriad filters. The particularity of this work is the convergence of the numerical scheme to a local minimum with probability 1, which is, up to our best knowledge, a novelty in the optimization community.

Computational photomechanics

In computational photomechanics, mainly two methods are available for estimating displacement and strain fields on the surface of a material specimen subjected to a mechanical test, namely digital image correlation (DIC) and localized spectrum analysis (LSA). With both methods, a contrasted pattern marks the surface of the specimen: either a random speckle pattern for DIC or a regular pattern for LSA, this latter method being based on Fourier analysis. It is a challenging problem since strains are tiny quantities giving deformations often not visible to the naked eye. This year's outcomes of our collaboration with Institut Pascal (Clermont-Ferrand) focus on three areas.

We have proposed an algorithm to render synthetic speckle images deformed under a predetermined deformation fixed by the user [17]. The goal is to generate ground truth datasets in order to assess the performance of the numerous variants of DIC and also the influence of extrinsic factors such as the noise or the marking pattern. It is required to carefully design the rendering algorithm in order to ensure that any measurement bias is caused by DIC estimation and not by the rendering algorithm itself. We have proposed to render speckle images based on a Boolean model, a standard model of stochastic geometry, a Monte Carlo estimation giving the gray level at any pixel. A software library and datasets are publicly available.

We have also investigated the optimization of the pattern marking the specimen [15], which is the topic of various recent papers. Checkerboard is the optimized pattern in terms of sensor noise propagation when the signal is correctly sampled, but its periodicity causes convergence issues with DIC. The consequence is that checkerboards are not used in DIC applications although they are optimal in terms of sensor noise propagation. We have shown that it is possible to use LSA to estimate displacement and strain fields from checkerboard images, although LSA was originally designed to process 2D grid images. A comparative study of checkerboards and grids shows that, under similar lighting conditions, the noise level in displacement and strain maps obtained with checkerboards is lower than that obtained with classic 2D grids.

Another scientific contribution concerns the restoration of displacement and strain maps. DIC and LSA both provide displacement fields equal to the actual one convolved by a kernel known a priori. The kernel indeed corresponds to the Savitzky-Golay filter in DIC, and to the analysis window of the windowed Fourier transform used in LSA. While convolution reduces noise level, it also gives a systematic measurement error. We have proposed a deconvolution method to retrieve the actual displacement and strain fields from the output of DIC or LSA [14]. The proposed algorithm can be considered as a variant of Van Cittert deconvolution, based on the small strain assumption. It is demonstrated that it allows enhancing fine details in displacement and strain maps, while improving the spatial resolution.

Cartoon-texture image decomposition

Decomposing an image as the sum of geometric and textural components is a popular problem of image analysis. In this problem, known as cartoon and texture decomposition, the cartoon component is piecewise smooth, made of the geometric shapes of the images, and the texture component is made of stationary or quasi-stationary oscillatory patterns filling the shapes. Microtextures being characterized by their power spectrum, we propose to extract cartoon and texture components from the information provided by the power spectrum of image patches. The contribution of texture to the spectrum of a patch is detected as statistically significant spectral components with respect to a null hypothesis modeling the power spectrum of a non-textured patch. The null-hypothesis model is built upon a coarse cartoon representation obtained by a basic yet fast filtering algorithm of the literature. The coarse decomposition is obtained in the spatial domain and is an input of the proposed spectral approach. We thus design a "dual domain" method. The statistical model is also built upon the power spectrum of patches with similar textures across the image. The proposed approach therefore falls within the family of non-local methods. Compared to variational methods or fast filers, the proposed non-local dual-domain approach [18] is shown to achieve a good compromise between computation time and accuracy. Matlab code is publicly available.

MFX Team

6. New Results

6.1. Carving Large Cavities in Shapes for Fast Fused Deposition Modeling

Participants: Samuel Hornus, Sylvain Lefebvre.

FDM Fused Deposition Modeling: fabricating things by depositing fused material into layers.

In 2016, we developed a technique for modeling a tight shield that protects the part being manufactured during 3D-printing with multi-material. In particular, the shield catches oozing material before it reaches the part [19]. The technique was implemented on a voxel representation of the shape. We also demonstrated its use for the modeling of a large *self-supported* cavity inside the shape.

In this more recent work, we have extended the technique to iteratively carve large cavities in the shape in order to hollow a shape as much as possible while maintaining its ability to be fabricated without internal support. (see Figure 2) We developed a polygonal implementation of the technique that provides much higher quality results. The work was published at the 2018 Eurographics conference as a short paper [14]. An implementation is now available to the general public in our software IceSL.

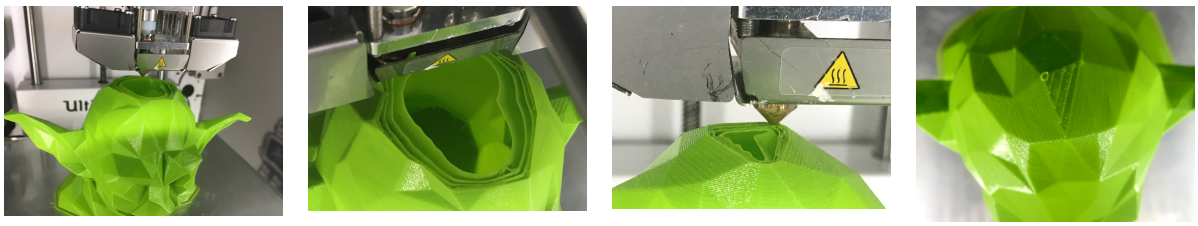

Figure 2. From Section 6.1 . Timelapse of the printing of a Yoda model. (Middle-left.) Note how the print is mostly empty and the nested cavity walls. Middle-right. Approaching the top of the head. Right. Closing the top of the head.

6.2. A Metamaterial for Fused Filament Fabrication

Participants: Jonàs Martínez Bayona, Samuel Hornus, Sylvain Lefebvre.

A critical advantage of additive manufacturing is its ability to fabricate complex small-scale structures. These microstructures can be understood as a *metamaterial*: they exist at a much smaller scale than the volume they fill, and are collectively responsible for an average elastic behavior different from that of the base printing material. For instance, this can make the fabricated object lighter and/or flexible along specific directions. In addition, the average behavior can be graded spatially by progressively modifying the microstructure geometry (see Figure 3).

The definition of a microstructure is a careful trade-off between the geometric requirements of manufacturing and the properties one seeks to obtain within a shape: in our case a wide range of elastic behaviors. Most existing microstructures are designed for stereolithography (SLA) and laser sintering (SLS) processes. The requirements are however different than those of continuous deposition systems such as fused filament fabrication, for which there was a lack of microstructures enabling graded elastic behaviors.

Figure 3. A 3D printed shoe sole. Left: Control fields used on the model, density (top), orthotropy strength (middle) and angle (bottom). Right: Printed shoe, top, side and bending. The shoe is printed without any skin to reveal the foam structure.

We introduced a novel type of metamaterial that *strictly enforces* all the requirements of Fused Filament Fabrication (FFF): continuity, self-support and overhang angles. This metamaterial offers a range of orthotropic elastic responses that can be graded spatially. This allows us to fabricate parts usually reserved to the most advanced technologies on widely available inexpensive printers that also benefit from a continuously expanding range of materials.

This work was presented at the SIGGRAPH conference and published in ACM Transactions on Graphics [12], and is integrated in the publicly available IceSL software. This was a joint work with Haichuan Song, then a post-doctoral researcher in ALICE.

6.3. Topology Optimization of Parametrized Stochastic Microstructures

Participants: Jonàs Martínez Bayona, Sylvain Lefebvre.

Different works have explored the topology optimization of parametrized periodic microstructures by the homogenization method. A promising venue of work lies in Additive Manufacturing technologies, that allow us to physically realize the intricate designs obtained with topology optimization. In order to fabricate the results, the parametrized microstructures must be projected at some finite scale taking into account the minimum printable size. However, for periodic microstructures it remains difficult to project and continuously grade the material properties since the boundary and transition between tiles has to be carefully handled.

We have an ongoing project in collaboration with Perle Geoffroy-Donders and Grégoire Allaire at École Polytechnique, to investigate the applicability of stochastic microstructures for topology optimization. This year we studied two different stochastic microstructures (isotropic and orthotropic) solely parametrized by an anisotropic metric and a Poisson point process. Both stochastic microstructures are amenable to efficient and scalable computation of their geometry. Unlike previous methods dealing with the projection of orthotropic microstructures the presented microstructures are able to easily follow a field of orthotropy orientation (see Figure 4).

6.4. Hash-based CSG Evaluation on GPU

Participants: Cédric Zanni, Sylvain Lefebvre.

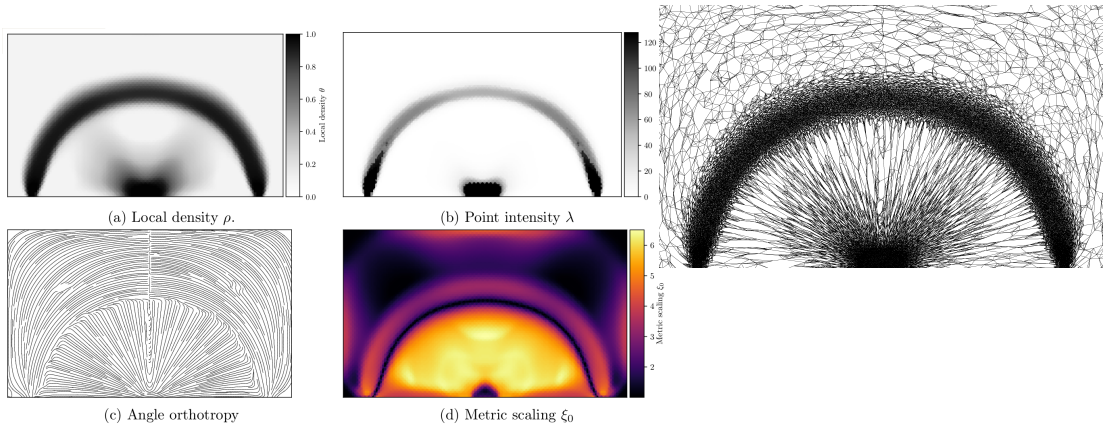

Figure 4. Optimization of a bridge problem with an orthotropic material, and our parametric stochastic microstructure. Left: Optimized parameters of the microstructure (density, angle of orthotropy, and degree of orthotropy). Right: Projection of the stochastic microstructure at a finite scale.

We have developed a new evaluation scheme for Constructive Solid Geometry (CSG) modeling that is well adapted to modern GPUs. The approach falls into the category of screen space techniques and can handle a large range of geometric representations. The proposed method relies on the idea of hashing in order to reduce the memory footprint for the processing of a given ray in the scene (*e.g.*, for discovering which part of the space is within or outside the object) while allowing the evaluation of the CSG in amortized constant time. This memory reduction in turn allows the space to be subdivided in order to apply progressively the rendering algorithm, ensuring that required data fit in the graphic memory. This improvement over previous approaches allows us to handle objects of higher complexity during both modeling and slicing for additive manufacturing.

The work was presented at the 2018 Symposium on Interactive 3D Graphics and Games conference and published in the ACM journal Computer Graphics and Interactive Techniques [15]. It was then integrated in the current version of our software IceSL.

6.5. Tile-based Pattern Design with Topology Control

Participant: Sylvain Lefebvre.

This project is a collaboration with Li-Yi Wei (HKU/Adobe) in the context of the PrePrint3D associated team. We consider the problem of producing tilings with boundary constraints, while enforcing global topology constraints. Tilings are composed by assembling a number of square tiles. Only tiles with compatible boundaries may be placed next to each others. In our context the tiles contain solid shapes connecting some of the borders together (corners, bars, crosses, etc.). Our algorithm is able to produce tilings that enforce border constraints as well as global topology constraints – in particular obtaining a connected network. This has applications in digital manufacturing, for instance to design decorative panels, but also in Computer Graphics, to synthesize large environments guaranteed to be navigable. These results were published in the ACM journal Computer Graphics and Interactive Techniques [10] and presented at the 2018 Symposium on Interactive 3D Graphics and Games conference.

We continue exploring tiling related problems, for instance to encode information within synthesized tilings [17].

6.6. Curved Deposition

Participants: Sylvain Lefebvre, Jimmy Etienne.

This project continues in collaboration with the ALICE team.

We are pursuing a line of research around curved deposition. The objective is to go beyond the flat-layers currently used. Indeed, some processes would allow for deposition along curved paths, however this capability is rarely used: proofs of concept exist, but no general algorithm can generate curved paths given an input geometry.

There are several key potential advantages to curved deposition: reducing the constraints in terms of geometries that can be manufactured, achieving better mechanical properties (*e.g.*, by aligning deposition with respect to a computed stress field), achieving better surface quality.

In this context, we achieved new results to reduce support material, in a joint project with Charlie C.L. Wang (TU Delft) [11]. The 3D printer is a 5-DOF robotic arm equipped with a standard FDM extruder. The algorithm we developed is based on a heuristic growth process within a discretized version of the model (voxels). The growth process attempts to place additional material where it is already supported from below, while avoiding cases where some unfinished parts of the model would become inaccessible due to collisions.

This led us to the first general algorithm for multi-axis 3D printing. It produces tool-paths that allow the robotic arm to fabricate most parts without any support, while avoiding collisions. Many challenges remain, both related to geometry and robotics, and we are pursuing this collaboration, jointly with Nicolas Ray (ALICE-Inria).

6.7. Colored 3D Printing

Participants: Sylvain Lefebvre, Jonàs Martínez Bayona, Noémie Vennin, Pierre Bedell.

In 2018 we kept developing our project regarding colored FDM printing. We have a paper accepted with minor revisions in ACM Transactions on Graphics. This was a joint work with Haichuan Song, then a post-doctoral researcher in ALICE. We worked on revising and refining our initial results throughout the year.

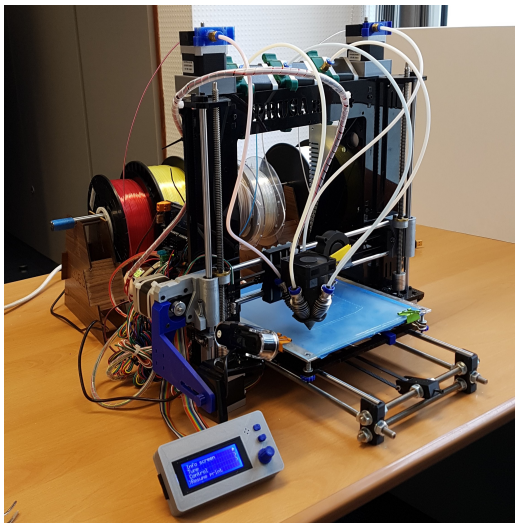

Figure 5. 3D printer Diamonds 5 filaments.

We proposed a novel algorithm for the problem of determining micro-layer mixtures to reproduce a subspace of material mixing ratios. We express the problem as fitting a simplex of minimal volume enclosing a set of points. The vertices of the simplex correspond to micro-layer mixtures, while the point set captures the desired mixtures within the model. This algorithm replaces the previous non-linear, gradient based optimizer. It achieves better results at a fraction of the previous computation time.

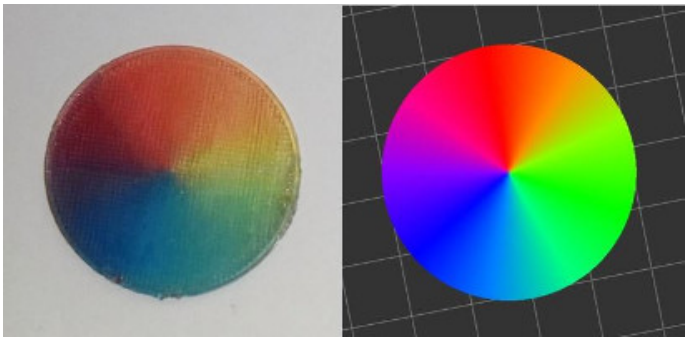

Figure 6. Disc showing all the gradation of color with 3 filaments (red, yellow, blue). Left: 3D printed disc. Right: Numerical view on the software IceSL.

We also developed, through the internship of Pierre Bedell, a 3D printer able to mix up to five filaments. We ran extensive testing and implemented additional improvements regarding flow control during deposition. Pierre Bedell joined the team as a research engineer and will keep participating in this project.

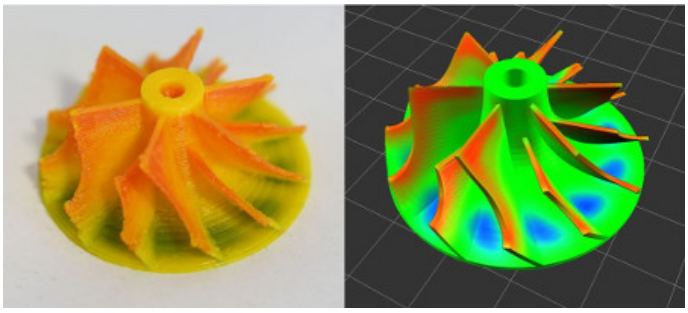

Figure 7. Turbine with a colored simulation of friction. Left: 3D printed turbine. Right: Numerical view on the software IceSL.

Noémie Vennin joined the team on a funding from *Université de Lorraine* to explore material aspects of the project, in a close collaboration with Sandrine Hoppe (LRGP). This part of the project receives support from the CPER Cyber-entreprise, thanks to which we acquired equipment to develop our own filament formulations, mixing pigments and additives to control color and transparency. Pierre Bedell and Noémie Vennin are developing a calibration process to determine the achievable color space given specific filaments, but also to tackle the inverse problem of designing filaments spanning a desired color space.

6.8. IceSL

Participants: Sylvain Lefebvre, Salim Perchy, Cédric Zanni, Samuel Hornus, Jonàs Martínez Bayona, Jimmy Etienne, Noémie Vennin, Pierre Bedell.

IceSL is the software developed within the team that serves as a research platform, a showcase of our research results, a test bed for comparisons and a vector of collaborations with both academic and industry partners. The software is freely available at https://icesl.loria.fr, both as a desktop and an online version.

In 2018, IceSL has been featured in news, exhibitions and fairs as a well-established tool for 3D printing. Additionally, since its inception, IceSL's community has grown significantly together with the number of new features included in it for slicing and modeling.

In February 2018, we organized the first event to introduce basic and advanced features which differentiate IceSL from other 3D printing tools. The event, targeted towards enthusiasts, allowed its participants to follow tutorials, interact with its developers and suggest additions and new directions for the software.

IceSL was also presented in May 2018 at the Strasbourg's Mini MakerFaire to a general audience that included high school students. The audience was introduced to IceSL's new features first hand and their applications to 3D printing. In addition to this, IceSL was shown to designers in November 2018 at Affinité Design (http://www.affinitedesign.com/) with part of the developing team demonstrating and answering questions on the use of IceSL as a modeling tool.

In October 2018, both the desktop and the online versions of IceSL were featured in a list of the 24 best 3D printing software tools. [0]

Regarding new features and additions to the software in 2018, IceSL has added several innovative methods for modeling and slicing. With respect to modeling, these include the ability to interactively paint values in a script (field tweaks), the option to export the shape generated with CSG to a mesh via dual contouring, texture synthesis on 3D objects, better font geometry creation as well as numerous improvements on its user interface and compatibility with hardware.

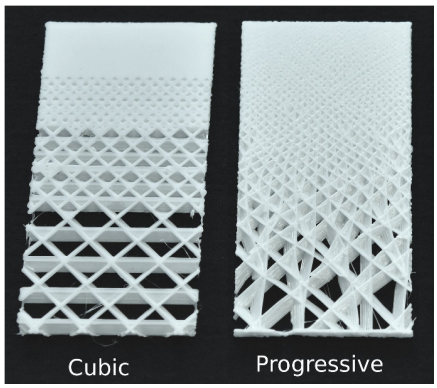

Figure 8. New infill patterns introduced in IceSL, the left one (cubic) has now been adopted by most other slicing software.

On the slicing front, IceSL has introduced new material infilling methods such as *polyfoam* [12], progressive and cubic structures (Figure 8) as well as putting a system in place allowing the user to specify an infill pattern through program image assets or shaders.

[0] https://all3dp.com/fr/1/meilleur-logiciel-imprimante-3d-gratuit-en-ligne/

IceSL also added a new method to compute supports called “wings,” a new framework for mixing colors into a 3D print [20] (presented in several exhibitions), curved printing covers, a faster slicing algorithm (in case of tessellated geometry), and a new geometry renderer [15].

The social community of IceSL is also growing accordingly. Its twitter account has around 200 followers and there are 150 users frequently interacting in its google forum. Downloads have increased around 30% after the first event in February 2018 to make a cumulative of 30k downloads since its initial release. [0] Youtube videos done by third persons on the usage of IceSL are also common (around a dozen in three different languages). And finally, in October 2018 IceSL launched its new website with a more professional look and additional resources (documentation, tutorials, videos, online version and new features).

6.9. Chill

Participants: Jimmy Etienne, Sylvain Lefebvre.

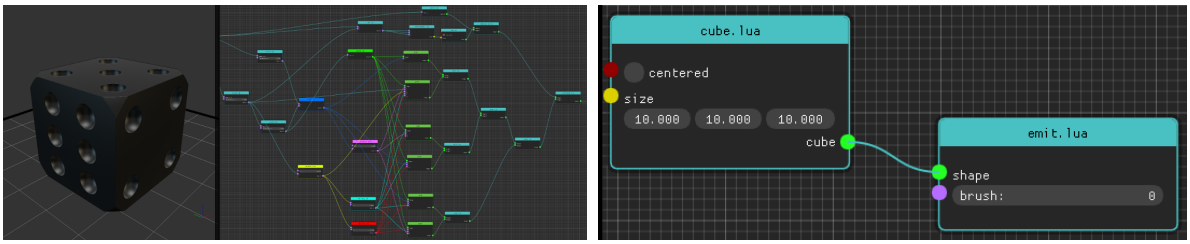

Figure 9. The Chill node-based GUI.

Chill is an open source GUI for IceSL, illustrated in Figure 9 . It features a node-based interface that hides the scripting language used to model shapes in IceSL. This enables a larger group of people to use IceSL in their projects, without having to type any code. The user of Chill creates 3D shapes by connecting various nodes arranged in a directed graph. The shape visualization is updated instantly as the graph is modified.

The source code is publicly available at https://github.com/shapeforge/Chill. We are planning to communicate broadly about the software in the first months of 2019.

[0] See https://gforge.inria.fr/top/toplist.php?type=downloads. Due to the removal of a file, the download counter on gforge.inria.fr is off by 6000 downloads.

MIMESIS Team

7. New Results

7.1. A Unified Bayesian and Physics-Based Approach for Non-rigid 3D Shape Reconstruction from 2D Images

We developed a method to reconstruct the 3D shape of the interventional device, based on a constrained physics-based simulation combined with 2D monocular fluoroscopic images through a Bayesian filter (see Fig. 4). Whereas the physics-based model provides a prediction of the device shape within the blood vessel, taking into account non-linear interactions between the catheter and the surrounding anatomy adds further information on its current position. In addition, an Unscented Kalman Filter is used to combine the navigation model with the 2D external observations. We focused on a medical application as we believe the method could provide an actual solution to some of the current limitations of fluoroscopy-based procedures. The use of a Bayesian formalism allows for retrieving a good estimate in presence of ambiguous views (i.e. in presence of overlapping anatomies) and to take into account uncertainties on the prediction model (errors in constraint definition, as well as inaccuracies in the mechanical characterization of the catheter) and errors in the external measurements.

The method has been implemented through software developed within the team; for more details see sec. 6.4 and 6.3 . Validation has been performed on porcine in-vivo data, acquired in accordance with UE norms, in collaboration with Pr. Mario GIMENEZ and Dr. Alain GARCIA from IHU-Strasbourg.

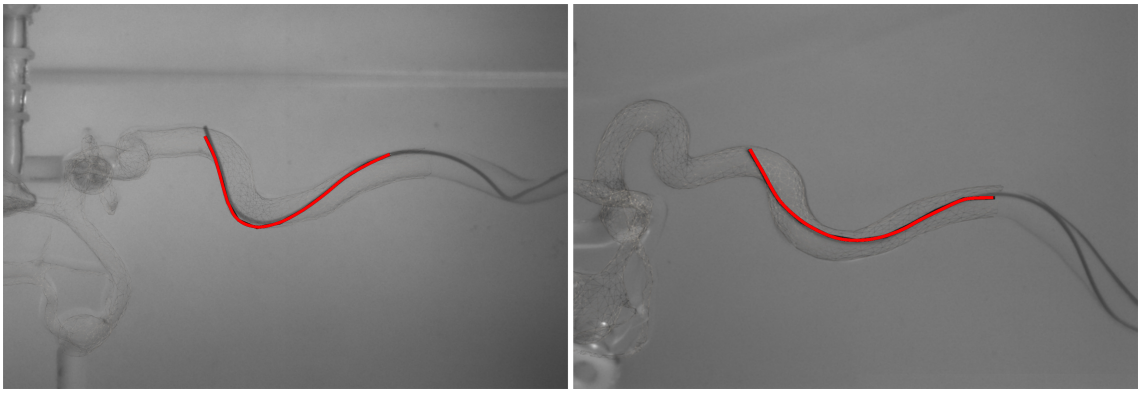

Figure 4. Catheter reconstruction in acquisition view (left) and validation view (right)

- *Authors:* Raffaella Trivisonne and Stéphane Cotin and Erwan Kerrien
- *Type:* PhD Thesis

7.2. Biomechanics-based graph matching for augmented CT-CBCT

Augmenting intraoperative cone beam computed tomography (CBCT) images with preoperative computed tomography (CT) data in the context of image-guided liver therapy is proposed (see Fig. 5). The expected benefit is an improved visualization of tumor(s), vascular system and other internal structures of interest. An automatic elastic registration based on matching of vascular trees extracted from both the preoperative and

intraoperative images is presented. Although methods dedicated to non-rigid graph matching exist, they are not efficient when large intraoperative deformations of tissues occur, as is the case during the liver surgery. First, an improved graph matching algorithm using Gaussian process is introduced by imposing additional constraints during the matching when the number of hypotheses is large; this extended version does not require a manual initialization of matching. Second, a fast biomechanical model is employed to make the method capable of handling large deformations. The proposed automatic intraoperative augmentation is evaluated on both synthetic and real data. It is demonstrated that the algorithm is capable of handling large deformations, thus being more robust and reliable than previous approaches. Moreover, the time required to perform the elastic registration is compatible with the intraoperative navigation scenario. The input data and result of the biomechanics-based graph matching method, which can handle large deformations and augment intraoperative CBCT, is shown in Fig. 5 .

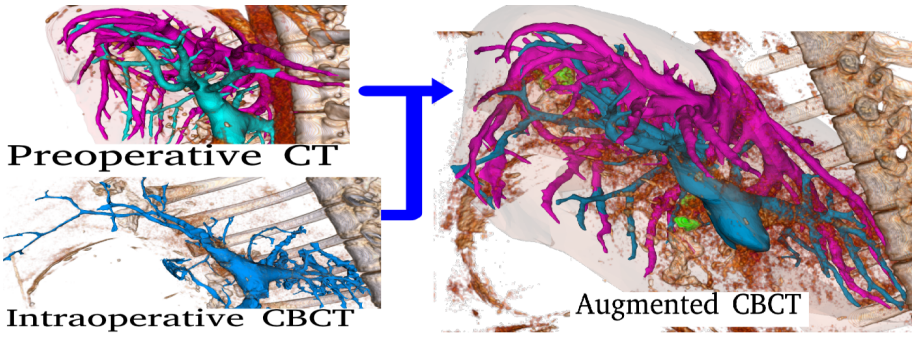

Figure 5. In the top left the preoperative CT image with complete portal and hepatic vessels clearly visible. In the bottom left the intraoperative deform, noisy CBCT image with partial only partial portal vessels visible. In the right the augmented CBCT view with the complete vessels added form preoperative image.

- *Authors:* Jaime Garcia Guevara and Igor Peterlik and Marie-Odile Berger and Stephane Cotin
- *Type:* Journal publication IJCARS 2018

7.3. A Combined Simulation and Machine Learning Approach for Force Classification during Robotized Intravitreal Injections

Intravitreal injection is one of the most common treatment strategies for chronic ophthalmic diseases. The last decade has seen the number of intravitreal injections dramatically increase, and with it, adverse effects and limitations. To overcome these issues, medical assistive devices for robotized injections have been proposed and are projected to improve delivery mechanisms for new generation of pharmacological solutions. In our work, we propose a method aimed at improving the safety features of such envisioned robotic systems. Our vision-based method uses a combination of 2D OCT data, numerical simulation and machine learning to estimate the range of the force applied by an injection needle on the sclera (see Fig. 6). We build a Neural Network (NN) to predict force ranges from Optical Coherence Tomography (OCT) images of the sclera directly. To avoid the need of large training data sets, the NN is trained on images of simulated deformed sclera. We validate our approach on real OCT images collected on five *ex vivo* porcine eyes using a robotically-controlled needle. Results show that the applied force range can be predicted with 94% accuracy. Being real-time, this solution can be integrated in the control loop of the system, allowing for in-time withdrawal of the needle.

- *Authors:* Andrea Mendizabal and Stephane Cotin
- *Type:* Conference publication MICCAI 2018

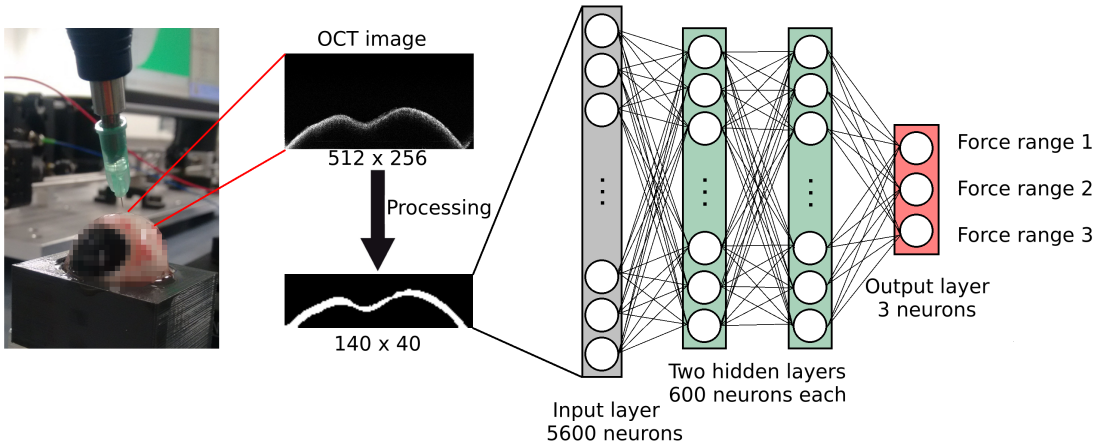

Figure 6. During the robotized intravitreal injection an OCT image of the deformed sclera is collected. The obtained image is processed and given a to a Neural Network that predicts the force range of the force applied by the robotically guided needle.

7.4. Inverse simulation for Robotic control of needle insertion

We recently published in a numerical method allowing for automatic control of a robot during needle insertion procedures (see Fig. 7). Our approach is to develop control models allowing for the correction and prediction of the deformation of structures (needle, tissues or the robot itself) and to adapt the behavior of the robot in order to reach an objective. We showed that inverse steps can be used to control an articulated robot while considering deformations of structures during needle insertion. The method has been used for a needle insertion inside a polyurethane foam using a Mitsubishi RV1A anthropomorphic robot arm. During the insertion vertical and lateral deformations were generated (see Fig. 7) leading to significant modification of the undeformed trajectory, important bending of the needle and even an off-plane shift between the base of the needle and the insertion point. Despite these strong modifications, the method was able to maintain the tip of the needle within the thickness of 1 cm of the foam and followed the desired curved path with accuracy lower than 1 mm without any human intervention.

7.5. Marker-Based Registration for Large Deformations

We proposed an Augmented Reality (AR) system for open liver surgery (see Fig. 8). Although open surgery remains the gold-standard for the treatment of complex tumors and central lesions, technological issues actually prevent using AR with sufficient accuracy for clinical use. We propose a markers-based method allowing for the tracking and the deformation of a preoperative model in real-time during the surgery. Markers are manually placed on the surface of the organ after opening the abdominal cavity, and tracked in real-time by a set of infrared cameras. Our framework is composed of both a non-rigid initial registration method, providing an estimation of the location of the markers in the preoperative model, and a real-time tracking algorithm to deform the model during the surgery (even for large deformation or partial occlusion of the organ). The method is validated on both synthetic and ex-vivo samples; in addition, we demonstrate its applicability in the operating room during a liver resection surgery on a human patient. Preliminary studies provided promising results to improve the location of tumors, and to help surgeons into planning the ideal resection intraoperatively.

- *Authors:* Yinoussa Adagolodjo, Nicolas Golse, Eric Vibert, Michel De Mathelin, Stéphane Cotin, Hadrien Courtecuisse

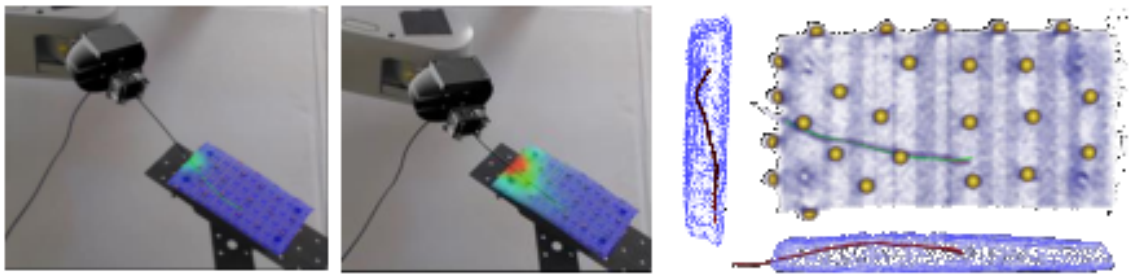

Figure 7. Robotic control based on inverse finite element simulations for needle insertion in deform-able structures. The models are registered in real-time using makers and infrared image-tracking system. We measured an average distance of 1.2 mm between needle's (red) and the desired trajectory (green).

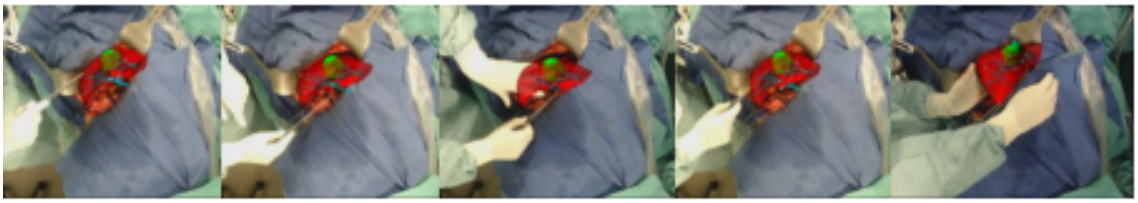

Figure 8. Figure 3: Research prototype for Augmented Reality during open surgery of the liver. We pro-posed a method based on markers for the registration of a preoperative model in real time during surgery. The markers are placed manually on the surface of the body after the opening of the abdominal cavity, and followed in real time by a set of infrared cameras.

- *Type:* PhD Thesis, publication to ICRA

7.6. Automatic and robust 2D/3D registration on fluoroscopic images

We introduce a unified solution to detect, register and track the liver in 2D live fluoroscopic X-ray images, in order to provide augmented reality and guidance during surgery (see Fig. 9). The solution can be decomposed into two phases, with an initial phase to globally estimate the rigid pose through template matching, and a second local rigid refinement step. A main contribution lies in the combination, for the pose refinement step, of intensity and contour based features over the contrasted vessels of the liver and surrounding organs, by integrating corresponding visual cues in a local optimization framework with respect to the pose. The method does not need any 2D segmentation of the contrasted vessels but relies on a synthetic X-ray rendering algorithm, and requires very few assumptions or priors. Our solution has been tested on synthetic and porcine data, showing its efficiency on realistic scenarios.

A non-rigid registration technique to account for local deformations of the target is also investigated. Once the model is rigidly aligned, local estimation of the deformations undergone by the vasculature and the parenchyma, given a linear or volumetric elastic deformation model of the vessels and the parenchyma, driven by local optical flow features.

The method has been implemented through software developed within the team; for more details see sec. 6.4 -6.3 . Validation has been performed on porcine in-vivo data, acquired in accordance with UE norms, in collaboration with Pr. Mario GIMENEZ and Dr. Alain GARCIA, Pr. Federico Davrieux, and Pim Hendriks, Daan Kuppens and from IHU-Strasbourg.

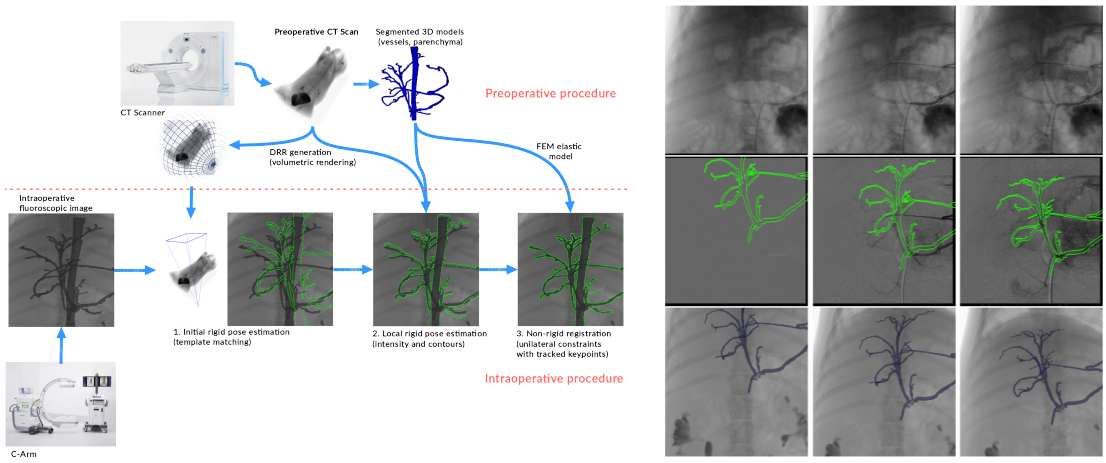

Figure 9. Pipeline of the system and results of the rigid registration system.

- *Authors:* Antoine Petit, Bruno Marques, Stéphane Cotin
- *Type:* Submission to IPCAI 2019

7.7. Capturing Deformations of Interacting Soft Objects Using RGB-D Data

We present a method for tracking multiple interacting deformable objects undergoing rigid motions, elastic deformations and contacts, using images and point cloud data provided by an RGB-D sensor (see Fig. 10). A joint registration framework is proposed, based on physical Finite Element Method (FEM) elastic and interaction models. It first relies on a visual segmentation of the considered objects in the RGB images.

The different segmented point clouds are then processed to estimate rigid transformations with on an ICP algorithm, and to determine geometrical point-to-point correspondences with the meshes. External forces resulting from these correspondences and between the current and the rigidly transformed mesh can then be derived. It provides both non-rigid and rigid data cues. Classical collision detection and response model is also integrated, giving contact forces between the objects. The deformations of the objects are estimated by solving a dynamic system balancing these external and contact forces with the internal or regularization forces computed through the FEM elastic model. This approach has been here tested on different scenarios involving two or three interacting deformable objects of various shapes, with promising results.

A case study in open surgery on the liver has also been investigated. Yet in this case a major improvement in the accuracy of the registration is provided by the integration of anatomical shape constraints, which are naturally hidden from the RGB-D camera, and that we account for through a registration with the pre-operative CT data. With a comparative study, we demonstrate the relevance of our method in a real-world application mimicking an open surgery scenario where the liver has to be tracked to provide an augmented reality view.

The method has been implemented through software developed within the team, especially the RGBDTracking plugin. 6.4 and 6.3 . Validation has been performed on porcine in-vivo data, acquired in accordance with UE norms, in collaboration with Pr. Mario GIMENEZ and Dr. Alain GARCIA, Pr. Federico Davrieux, and Pim Hendriks, Daan Kuppens and from IHU-Strasbourg.

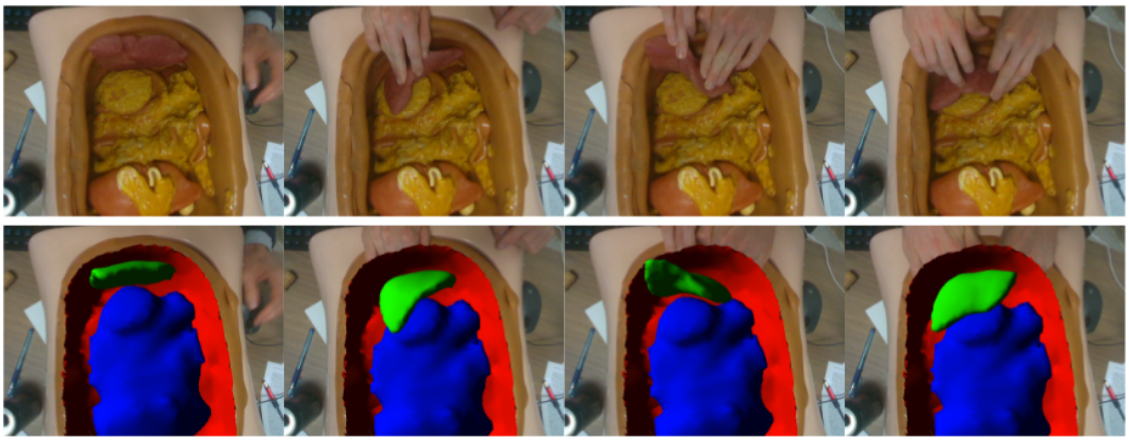

Figure 10. Pipeline of the system and results of the rigid registration system.

- *Authors:* Antoine Petit, Stéphane Cotin
- *Type:* Publications to IROS 2018 and ACCV Workshops 2018

MOCQUA Team

7. New Results

7.1. Completeness of the ZX-calculus

- Participants: Renaud Vilmart, Simon Perdrix, Emmanuel Jeandel

The ZX-Calculus is a powerful graphical language for quantum reasoning and quantum computing introduced by Bob Coecke and Ross Duncan [36]. The ZX-calculus has several applications in quantum information processing [37] (e.g. measurement-based quantum computing, quantum codes, foundations), and can be used through the interactive theorem prover Quantomatic. However, the main obstacle to wider use of the ZX-calculus was the absence of a *completeness* result for a *universal* fragment of quantum mechanics, in order to guarantee that any true property is provable using the ZX-calculus. We have introduced the first complete axiomatisation for a universal fragment of quantum mechanics. We also showed that a single additional rule makes the ZX-calculus complete for the whole pure qubit quantum mechanics. These results have been presented at LICS this year [16], [17] and will be presented at QIP'19, the main conference in quantum information processing.

7.2. Second-order entropy accumulation theorem

- Participants: Frédéric Dupuis

Device-independent cryptography is a way to use quantum mechanics to perform cryptographic tasks using equipment from an untrusted manufacturer. To prove the security of device-independent protocols, the main challenge is to show that a step-by-step procedure involving the untrusted device produces a certain of randomness even from the point of view of the manufacturer. The entropy accumulation theorem [38] provides a generic way to obtain such statements. However, while the bounds provided by this theorem are optimal in the first order (meaning the term that is linear in the number of steps in the process), the second-order sublinear term is bounded more crudely, in such a way that the bounds deteriorate significantly when the theorem is applied directly to protocols where parameter estimation is done by sampling a small fraction of the positions, as is done in most QKD protocols. In [25], we improve this second-order sublinear term and remedy this problem. This paper has been submitted to IEEE Transactions on Information Theory.

7.3. Mixed-state certification

- Participants: Frédéric Dupuis

Mixed-state certification consists of ensuring that a quantum state on n subsystems is close to n copies of a given mixed state, up to a small number of errors, by sampling a small fraction of the positions. While this task makes no sense classically (it effectively amounts to certifying that a string came from a particular probability distribution), it makes sense quantumly if we can ask someone (that we call a prover) to supply purifications of the sampled positions. However, such sampling procedures cannot be analyzed straightforwardly using standard sampling results, and care must be taken even when defining what success means. In [26], we introduced these concepts, and we showed that this sampling protocol offers secure certification in the presence of a possibly dishonest prover. We then applied this result to two-party quantum coin-tossing. This work was presented at QCrypt 2018 and TCC 2018 (and will appear in the proceedings of the latter).

7.4. Descriptive Set Theory

- Participants: Mathieu Hoyrup

Descriptive Set Theory (DST) aims at classifying sets and functions in terms of the complexity of describing them. It is closely related to logic and computation theory, where sets and functions can be described by logical formulas or computer programs. DST was originally developed on a restricted class of topological spaces, the Polish spaces, which does not cover important classes of spaces that are needed in Theoretical Computer Science, especially in programming semantics, notably (Scott) domains or spaces of higher-order (Kleene-Kreisel) functionals. We investigate DST on such spaces and show that it does not work as nicely as on usual spaces. The article [29] is currently submitted. This work has been presented during an invited talk at CiE 2018 [13].

7.5. Semicomputable geometry

- Participants: Mathieu Hoyrup

Semicomputability is a natural notion arising from logic and theoretical computer science. Termination of programs is not decidable but semidecidable. Semicomputability of subsets of the plane is an important notion. For instance whether the famous Mandelbrot set is computable is still an open problem, while its semicomputability is easy to prove. Intuitively, we can write a program that progressively fills out the complement of the set, but we do not know when the picture is complete. We studied semicomputability of much simpler sets, namely filled triangles. While this problem looks simple at first sight, it is considerably rich and raises many questions. What properties should the coordinates of the vertices of a triangle satisfy to make it semicomputable? How can we parametrize such triangles? What happens for other sets such as disks or general convex sets? We developed a thorough study of these problems in [15].

7.6. Resource bounded computation

- Participants: Emmanuel Hainry

Controlling resource consumption is a crucial aspect of programming. Resources such as time, space, intrication are limited, and helping the programmer to avoid overconsumption or pointing problematic code is an important endeavor. We introduced a type-system for an Object Oriented Programming Language (*à la* Java) that gives a guarantee of polynomial-time computability provided that the program halts [12]. This result has several interesting features as it works with complex object data-structures in a real-like programming language; checking the type system is polynomial time decidable; we provided a O bound hence giving an explicit worst case complexity bound.

7.7. Inductive reasoning

- Participants: Isabelle Gnaedig, Sofien Ben Ayed

We are interested in quantifying the power of axiomatic theories. For this purpose, induction is a key concept. We have investigated the different validity proofs of inductive reasoning, the equivalence of induction with the well-ordered principle and well-foundedness, the differences between first and second order forms of the induction principle, and the notion of ω-consistency, qualifying theories interpreting arithmetic for which proving a property for each value of standard integers does not imply that the property is always true. We have also studied the importance of the axiom of choice for induction, and analysed a recent interpretation of induction by Hardin and Taylor through the hat problem [22].

7.8. Cellular automata with stochastic evolutions

- Participants: Nazim Fatès, Irène Marcovici

In order to explore the computing abilities of simple stochastic cellular automata, we tackle the case of Alesia, a two-player zero-sum game which is quite similar to the rock-paper-scissors game. In this game, two players simultaneously move and do not know what the opponent plays at a given round. The simultaneity of the moves implies that there is no deterministic good strategy in this game, otherwise one would anticipate the moves of the opponent and easily win the game. We explored how to build a family of one-dimensional stochastic cellular automata to play this game by progressively increasing the complexity of the transitions. We showed the possibility to construct a family of rules with interesting results, including good performance when confronted to the Nash-equilibrium strategy [14].

The reversibility of classical cellular automata (CA) was examined for the case where the updates of the system are random. In this context, with B. Sethi and S. Das (IIT Karaghpur, India), we studied a particular form of reversibility: the possibility of returning infinitely often to the initial condition after a random number of time steps. This is the recurrence property of the system. We analyzed this property for the simple rules and described the communication graph of the system [21].

We also contributed to the diffusion of some already-established knowledge on the simulation of complex systems in Biology, more precisely in the case of the formation of swarms [19] and in the case of asynchronous cellular automata [20].

MULTISPEECH Project-Team

7. New Results

7.1. Explicit Modeling of Speech Production and Perception

Participants: Anne Bonneau, Vincent Colotte, Denis Jouvet, Yves Laprie, Slim Ouni, Agnès Piquard-Kipffer, Théo Biasutto-Lervat, Sara Dahmani, Ioannis Douros, Valérian Girard, Thomas Girod, Anastasiia Tsukanova.

7.1.1. Articulatory modeling

7.1.1.1. Articulatory models and synthesis

Since articulatory modeling, i.e. representing the geometry of the vocal tract with a small number of parameters, is a key issue in articulatory synthesis the improvement of the articulatory models remains an important objective. This year we put emphasis on thin articulators as the epiglottis and velum. Indeed, the delineation of those contours often leads to erroneous transverse dimensions (too thin or too thick contours) which generates some artificial swelling deformations. Before the determination of the deformation modes, the central lines of the velum and epiglottis are extracted in the images use to build the model. The deformation modes thus only concern the central line, which prevents artificial swelling factors to emerge from the factor analysis. A reconstruction algorithm has been developed to obtain the contour from the central line.

7.1.1.2. Acoustic simulations

One of the issues in articulatory synthesis is to assess the impact of the geometric simplifications that are made on the vocal tract so as to enable faster acoustic simulation and to decrease the number of parameters required to approximate the vocal tract shape. The other issue concerns the impact of the plane wave assumption. The idea consists of comparing the signal or spectrum synthesized via numerical acoustic simulation against the one measured on a real human subject. However, this requires that both geometric and corresponding acoustic data are available at the same time. This can be achieved with MRI data when the acquisition duration is sufficiently short to allow the speaker to phonate the sound during the whole acquisition. The MRI acquisition protocol has thus been optimized on the new Siemens Prisma MRI machine of Nancy hospital so as to reduce the acquisition time to 7 seconds, which makes it possible for the subject to produce a sound throughout the acquisition. The acoustic simulation was achieved by using the Matlab K-wave package, either from the entire 3D volume extracted from the MRI data, or from the 2D shape extracted from the mid-sagittal plane. Several simplifications have been carried out (with or without the epiglottis, with or without the velum...) so as to assess their acoustic impacts. These simulations only concern vowels because these sounds can be sustained by subjects and the MRI machine noise does not change the position of formant frequencies dramatically. This work has been carried out in cooperation with IADI laboratory.

7.1.1.3. Exploitation of dynamic articulatory data

The size of the dynamic database (recorded last year in the Max Planck Institute for Biophysical Chemistry in Göttingen), in the form of MRI films of the mid-sagittal plane acquired at 55 Hz, is about 200.000 images. Even if the long term objective is to exploit the whole database, efforts were dedicated to manual delineation of contours in some films with the idea of using those data to train a machine learning technique. Several students were trained, and in total more than 1000 images have been delineated. The corresponding films have been exploited to achieve articulatory copy synthesis by improved acoustic simulations developed last year.

7.1.1.4. Acoustic-to-articulatory inversion

Deriving articulatory dynamics from the acoustic speech signal is a recurrent topic in our team. This year, we have investigated whether it is possible to predict articulatory dynamics from phonetic information without having the acoustic speech signal. The input data may be considered as not sufficiently rich acoustically, as there is probably no explicit coarticulation information, but we expect that the phonetic sequence provides compact yet rich knowledge. We have experimented a recurrent neural network architecture, where we have trained the model with an electromagnetic articulography (EMA) corpus, and have obtained good performances similar to the state-of-the-art articulatory inversion from line spectral frequencies (LSF) features [21].

7.1.2. Expressive acoustic and visual synthesis

7.1.2.1. Expressive speech

A comparison between emotional speech and neutral speech has been carried on using a small corpus of acted speech. The analysis was focused on the way pronunciations and prosodic parameters are modified in emotional speech, compared to neutral style [20].

Experiments with deep learning-based approaches for expressive speech synthesis are described in 7.2.4.2 .

7.1.2.2. Expressive audiovisual synthesis and lipsync

This year, we have acquired audiovisual 3D corpus (using the optitrack system, using 8 cameras) for a set of emotions acted by a professional actress. We recorded 6 basic emotions: joy, fear, disgust, sadness, anger, surprise; in addition to neutral speech. The corpus contains 5000 utterances (2000 utterances for the neutral speech and 500 utterances per emotion). The visual and acoustic data have been processed, segmented and labeled spatially and temporally. An important aspect of the work was to study the evaluation of the quality of the animation of a 3D talking head where the animation is generated from the acquired 3D data. For this purpose, we studied the relevance of root mean square error (RMSE) measure which is classically used to evaluate the error of the prediction. Our preliminary results confirmed that RMSE can be irrelevant in our field, as we may not reach critical articulatory target, and we still obtain very low RMSE. Thus the audiovisual intelligibility of the system would be low. To improve the results, we have worked on improving the 3D model controls using better key-shapes and reduced redundant and confusing blendshapes.

The processed neutral-speech data have been used to train a deep neural network to predict from speech and linguistic information the trajectories of the animation controls of the talking head, which is the core of the lipsync system. We have also used this expressive-speech data to train a DNN-based TTS to synthesize expressive audiovisual speech from text. Currently, we are performing extensive testing and validation of the results.

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

7.1.3.1. Visual clues in speech perception and production

We continue our research focused on the importance of multimodal speech combining oral and visual clues. We investigated identification and production of morpho-syntactic skills in ten deaf children (severe with cochlear implant using French cued-speech LPC - *Langue française Parlée Complétée*) and ten age-matched children with typical development. Our goal was to examine the production of morpho-syntactic structures in auditory channel versus audiovisual speech. Five conditions were observed: audiovisual conditions with a 3D avatar speaking or coding oral language with LPC versus a human speaker with or without LPC and auditory channel. We used the 3D avatar coding set up in the ADT Handicom project. Statistical analysis and interpretation of results is ongoing.

7.1.3.2. Reading and related skills norms

We set-up standardized norms on the development of reading and related skills in French: EVALEC Primaire software (in collaboration with the LPC - *Laboratoire de Psychologie Cognitive*, UMR 7290, *Aix-Marseille Université*). This year, LPC collected new data at the end of grade 5 (about 100 children) and added them to those previously collected at the end of grades 1–4, about 100 children for each level [69]. EVALEC primaire software includes five tests focused on written word processing, recording both accuracy scores and processing time (time latency and vocal response duration for the reading aloud tests). EVALEC primaire software also includes tests of phonemic and syllabic awareness, phonological short-term memory, and rapid naming. These data would allow researchers and speech therapists to assess the reading and reading-related skills of dyslexic children as compared to average readers.

7.1.3.3. Analysis of non-native pronunciations

We have examined the effects of L1/L2 interferences at the segmental level, and of the lack of fluency at the sentence level, on the realizations of French final fricatives by German learners. Due to L1/L2 interference, German speakers tend to devoice French final fricatives. A well-known effect of the lack of L2 mastering is the decrease of the speech articulation rate, which lengthens the average duration of segments. In order to better apprehend the impact of categorization and fluency, we selected four series of consonants from the IFCASL corpus, i.e. voiced and unvoiced fricatives uttered by French native and German non-native speakers. The realizations of French unvoiced consonants uttered by German speakers are essentially dependent on fluency, whereas the realizations of voiced consonants by the same speakers are dependent on both fluency and categorization. We evaluated a set of acoustic cues related to the voicing distinction -including consonant duration and periodicity-, and submitted the data to a hierarchical clustering analysis. Results, discussed as a function of speaker's level and prosodic boundaries, confirmed the mutual importance of fluency and segmental categorization on non-native realizations [22].

Within the METAL project, work is on-going for integrating speech processing technology in an application to help learning foreign language and for experimenting it with middle and high school students learning German. This includes tutoring aspects using a talking head to show proper articulation of words and sentences; as well as using automatic tools derived from speech recognition technology, for analyzing student pronunciations. Preliminary experiments have shown the poor quality of speech signals recorded from groups of students in classrooms.

7.2. Statistical Modeling of Speech

Participants: Vincent Colotte, Antoine Deleforge, Dominique Fohr, Irène Illina, Denis Jouvet, Odile Mella, Romain Serizel, Emmanuel Vincent, Md Sahidullah, Guillaume Carbajal, Ken Déguernel, Diego Di Carlo, Adrien Dufraux, Raphaël Duroselle, Mathieu Fontaine, Nicolas Furnon, Amal Houidhek, Ajinkya Kulkarni, Nathan Libermann, Aditya Nugraha, Manuel Pariente, Laureline Perotin, Sunit Sivasankaran, Nicolas Turpault, Imene Zangar.

7.2.1. Source localization and separation

Emmanuel Vincent has co-edited a 500-page book on audio source separation and speech enhancement, which provides a unifying view of array processing, matrix factorization, deep learning and other methods, with application to speech and music [64]. We also contributed to five chapters in that book [60], [62], [59], [54], [61] and three chapters in another book [53], [56], [55].

7.2.1.1. Source localization

In multichannel scenarios, source localization and source separation are tightly related tasks. We introduced the real and imaginary parts of the acoustic intensity vector in each time-frequency bin as suitable input features for deep learning based speaker localization [37]. We analyzed the inner working of the neural network using a methodology called layerwise relevance propagation, which points the time-frequency bins on which the network relies to output a given location [68]. We defined a new task called text-informed speaker localization, which consists of localizing the speaker uttering a known word or sentence such as the wake-up word of a hands-free voice command system in a situation when other speakers are overlapping. We proposed a method to address this task, where a phonetic alignment is obtained, converted into an estimated time-frequency mask, and fed to a convolutional neural network together with interchannel phase difference features in order to localize the desired speaker [43]. We published a new dataset using a microphone array embedded in an unmanned aerial vehicle in [45], organized an international sound source localization challenge associated to this dataset and participated to the 2018 LOCATA sound source localization challenge. We published a book chapter on audio-motor integration, showing an application to sound source localization with robots [52].

7.2.1.2. Room acoustics modeling

In a given room, each possible position of the microphones and the sources corresponds to different room transfer functions. The goal of room acoustic modeling is to model the manifold formed by these transfer functions. Past studies have focused on learning a supervised mapping between the relative transfer function and the source location for localization purposes. We introduced the reverse task consisting of learning a mapping between the source location and the corresponding relative transfer function, which may be used as a prior on the relative transfer function for source separation purposes. We proposed a semi-supervised algorithm to learn this mapping in a situation when the location of each relative transfer function measurement is not precisely known [48]. We also started investigating the estimation and modeling of early acoustic echoes. In [39] we showed how their knowledge could improve performance of sound source separation algorithms. In [36] we proposed a new method to estimate them blindly from multichannel recordings with much higher precision than conventional blind channel identification methods.

7.2.1.3. Deep neural models for source separation and echo suppression

We pursued our research on the use of deep learning for multichannel source separation [5]. We introduced a method that exploits knowledge of the source locations in order to estimate multichannel Wiener filters for two or more sources [38]. We explored several variants of the multichannel Wiener filter, which turned out to result in better speech recognition performance on the CHiME-3 dataset [17]. We also used deep neural networks for reducing the residual nonlinear echo after linear acoustic echo cancellation [23] and started extending this approach to joint reverberation, echo, and noise reduction. Finally, we recently started exploring the case where the microphones composing a multichannel array are not distributed according to a predefined geometry and do not have a common sampling clock.

7.2.1.4. Alpha-stable modeling of audio signals

This year, our work on heavy tails distribution has witnessed a significant advance with the development of a multichannel model that is able to account for the inter-channel delays and time difference of arrivals in an alpha-stable framework, hence benefiting from the inherent robustness of such distributions. This work has been submitted to the IEEE transactions on Signal Processing by Mathieu Fontaine and is still under review. Its main applications are: i/ the separation of multichannel sources, for which we have demonstrated a superiority with respect to the multichannel Wiener filter in the oracle setting, and ii/ localizations of heavy tailed sources, where we worked on the theoretical foundations

7.2.1.5. Beyond Gaussian modeling of audio signals

The team has investigated a number of alternative probabilistic models to the symmetric local complex Gaussian (LCG) model for audio source separation. An important limit of LCG is that most signals of interest such as speech or music do not exhibit Gaussian distributions but heavier-tailed ones due to their important dynamic. In [31] we proposed a new sound source separation algorithm using heavy-tailed alpha stable priors for source signals. Experiments showed that it outperformed baseline Gaussian-based methods on under-determined speech or music mixtures. Another limitation of LCG is that it implies a zero-mean complex prior on source signals. This induces a bias towards low signal energies, in particular in under-determined settings. With the development of accurate magnitude spectrogram models for audio signals using deep neural networks, it becomes desirable to use probabilistic models enforcing stronger magnitude priors and better accounting for phases. In [35], we presented the BEADS (Bayesian Expansion Approximating the Donut Shape) model. The prior considered is a mixture of isotropic Gaussians regularly placed on a zero-centered complex circle. We showed it outperformed LCG on an informed source separation task.

7.2.1.6. Interference reduction

Our work on interference reduction focused this year in scaling our previous work to full-length recording. This has been achieved thanks to a new method we proposed, which estimates the interference reduction parameters based on random projections of the full length recordings [25]. This technique scales linearly with the duration of the recording, making it usable in real-world use-cases.

The book chapter we published on audio-motor integration, shows an application to ego-noise reduction for robots [52]. In the context of robotics, ego-noise refers to the acoustic noise produced in a robot's microphones by its own movement.

7.2.2. Acoustic modeling

7.2.2.1. Robust acoustic modeling

Achieving robust speech recognition in reverberant, noisy, multi-source conditions requires both speech enhancement and separation and robust acoustic modeling. In order to motivate further work by the community, we created the series of CHiME Speech Separation and Recognition Challenges in 2011 [1]. We oversaw the collection of a new dataset sponsored by Google, which considers a 'dinner party' scenario. Twenty parties of four people, who know each other well, were recorded in their own homes using 2 binaural in-ear microphones per participant and 6 distant Kinects, for a total duration of about 50 h. We organized the CHiME-5 Challenge based on these data [19]. We also participated in the collection of two French datasets for ambient assisted living applications as part of the voiceHome [11] and VOCADOM [51] projects.

7.2.2.2. Ambient sounds

We are constantly surrounded by sounds and we rely heavily on these sounds to obtain important information about what is happening around us. Our team has been involved in the community on ambient sound recognition for the past few years. In collaboration with Johannes Kepler University (Austria) and Carnegie Mellon University (USA), we co-organized a task on large-scale sound event detection as part of the Detection and Classification of Acoustic Scenes and Events (DCASE) 2018 Challenge [40]. It focused on the problem of learning from audio segments that are either weakly labeled or not labeled, targeting domestic applications. In this context, we work on semi-supervised sampling strategies to create triplets (a triplet is composed of the current sample, a so-called positive sample from the same class as the current sample and a negative sample from a different class) and studied their application to train triplet networks for audio tagging.

7.2.2.3. Speech/Non-speech detection

Automatic Speech Recognition (ASR) of multimedia content such as videos or multi-genre broadcasting requires a correct extraction of speech segments. We explored the efficiency of deep neural models for speech/non-speech segmentation. We used a bidirectional LSTM model to obtain speech/non-speech probabilities and a decision module (4-state automaton with safety margins). Compared to a Gaussian Mixture Model (GMM) based speech/non-speech segmenter, the results achieved on the MGB British Challenge data, show a reduction of the ASR word error rate (23.7% versus 29.4%). We have also trained models for the Arabic and French languages.

7.2.2.4. Transcription systems

Within the AMIS project, speech recognition systems have been developed for the transcription of videos in French, English and Arabic. They have been integrated with other components (such as translation and summarization) to allow for the summarization of videos in a target language [44], [29], [28].

7.2.2.5. Speaker recognition

Speaker recognition is the task of recognizing a person from its voice. The performances of speaker recognition systems severely degrade due to several practical challenges such as the limited amount of speech data, real-world noises and spoofing. We explored the efficiency of DNN-based distance metric learning methods for speaker recognition in short duration conditions. Currently, we are developing a neural network architecture that gives phone-invariant speaker embeddings for robust speaker recognition. We also participated in the NIST speaker recognition evaluation 2018 as a part of the I4U consortium. The speaker recognition technology is vulnerable to spoofing attacks where mimicked voice, synthetic speech, or playback voice is used to get illegitimate access. We are investigating whether technology-assisted speaker selection can help in improving mimicry attack [67]. In [24], we proposed an enhanced baseline system for replay spoofing detection with ASVspoof 2017 dataset. In [26], we demonstrated that playback speech enhanced with DNN-based speech enhancement method can severely degrade the speaker recognition and countermeasure performance as compared to the conventional replay attacks with voice samples from covert recording. We also proposed

a common feature and back-end fusion scheme for the integration of spoofing countermeasures and speaker recognition [47]. Currently, we are co-organizing the third edition of automatic speaker verification spoofing challenge (ASVspoof 2019) where our newly developed cost function [32] will be adopted for the performance assessment of integrated systems. In the context of multimodal authentication with the voice as a modality, we investigated the optimization of speech features for audio-visual synchrony detection [41].

7.2.2.6. Language identification

With respect to language identification, the current research activity focuses on lightly supervised or unsupervised domain adaptation. The goal is to adapt a language identification system optimized for a given transmission channel to a new transmission channel.

7.2.3. Language modeling

7.2.3.1. Out-of-vocabulary proper name retrieval

Despite recent progress in developing Large Vocabulary Continuous Speech Recognition Systems (LVCSR), these systems suffer from Out-Of-Vocabulary words (OOV). In many cases, the OOV words are Proper Nouns (PNs). The correct recognition of PNs is essential for broadcast news, audio indexing, etc. We addressed the problem of OOV PN retrieval in the context of broadcast news LVCSR. We focused on dynamic (document dependent) extension of LVCSR lexicon. To retrieve relevant OOV PNs, we proposed to use a very large multipurpose text corpus: Wikipedia. This corpus contains a huge number of PNs. These PNs are grouped in semantically similar classes using word embedding. We used a two-step approach: first, we selected OOV PN pertinent classes with a multi-class Deep Neural Network (DNN). Secondly, we ranked the OOVs of the selected classes. The experiments on French broadcast news show that a bi-directional Gated Recurrent Unit model outperforms other studied models. Speech recognition experiments demonstrate the effectiveness of the proposed methodology [18].

7.2.3.2. Updating speech recognition vocabularies

Within the AMIS project, the update of speech recognition vocabularies has been investigated using web data collected over a time period similar to that of the collected videos, for three languages: French, English and Arabic. Results have been analyzed globally, and also with respect to names only. This analysis has shown the poor coverage of the names by the baseline lexicons, and has also demonstrated the benefits of the updated lexicons, both in term of WER reduction and OOV rate reduction [14].

7.2.3.3. 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. Ken Déguernel defended his PhD on automatic music improvisation [10] and he proposed a polyphonic music improvisation approach that takes the structure of the musical piece at multiple time scales into account [12]. We also explored the ability of a conventional recurrent neural network with moving history to account for long-term dependencies in music melodies, and compared it with two new architectures with growing or parallel history [50].

7.2.3.4. Automatic detection of hate speech

Nowadays, Twitter, LinkedIn, Facebook and YouTube are very popular for communicating ideas, beliefs, feelings or any other form of information. At the same time, the dark side of these new technologies has led to an increase in hate speech or racism. Our work seeks to study hate speech in user-generated contents in France, which thus requires French resources. We plan to design a hate speech corpus and a lexicon in French; whereas such hate speech lexicons exist for other languages, no such tool can be found in French. We began, on English data, to develop a new methodology to automatically detect hate speech, based on machine learning and Neural Networks. Human detection of this material is unfeasible since the contents to be analyzed are huge. Current machine learning methods use only certain task specific features to model hate speech. We propose to develop an innovative approach to combine these pieces of information into a multi-feature approach so that the weaknesses of the individual features are compensated by the strengths of other features. We began a collaboration with the CREM laboratory in Metz and Saarland University.

7.2.4. Speech generation

7.2.4.1. Arabic speech synthesis

Work on Arabic speech synthesis was carried out within a CMCU PHC project with ENIT (École Nationale d'Ingénieurs de Tunis, Tunisia), using HMM and NN based approaches applied to Modern Standard Arabic language. Speech synthesis systems rely on a description of speech segments corresponding to phonemes, with a large set of features that represent phonetic, phonologic, linguistic and contextual aspects. When applied to Modern Standard Arabic, two specific phenomena have to be taken in account: vowel quantity and gemination. This year, we studied thoroughly the modeling of these phenomena. Results of objective and subjective evaluations showed that the use of a deep neural architecture in speech synthesis (more specifically in predicting the speech parameters) enhanced the accuracy of acoustic modelling so that the quality of generated speech is better than that of HMM-based speech synthesis [30], [13].

Deep neural network (DNN) approaches have been further investigated for the modeling of phoneme duration. According to the specific phenomena of the Arabic language, we proposed a class-specific modeling of the phoneme durations. An objective evaluation showed that the proposed approach leads to a more accurate modeling of the phoneme duration (compared to HMM-based or MERLIN DNN-based approaches) [49].

7.2.4.2. Expressive acoustic synthesis

Expressive speech synthesis using parametric approaches is constrained by the style of the speech corpus used. We carried out a preliminary study on developing expressive speech synthesis for a new speaker voice without requiring a specific recording of expressive speech by this new speaker. For that, we focused on deep neural network based layer adaptation for investigating the transfer the expressive characteristics to a new speaker for which only neutral speech data is available. Such transfer learning mechanism should accelerate the efforts towards exploiting existing expressive speech corpora. However, there is a trade-off between the knowledge transfer of expressivity characteristics and the retaining of the speaker's identity in the synthesized speech.

7.3. Uncertainty Estimation and Exploitation in Speech Processing

Participants: Irène Illina, Denis Jouvet, Emmanuel Vincent, Yassine Boudi, Baldwin Dumortier, Elodie Gauthier, Mathieu Hu, Lou Lee, Anne-Laure Piat-Marchand.

7.3.1. Uncertainty and acoustic modeling

7.3.1.1. Uncertainty in noise-robust speech and speaker recognition

In many real-world conditions, the target speech signal overlaps with noise and some distortion remains after speech enhancement. The framework of uncertainty decoding assumes that this distortion has a Gaussian distribution and seeks to estimate its covariance matrix and propagate it through the acoustic model for robust ASR [4]. We introduced new Gaussian mixture model-derived (GMMD) uncertainty features for robust DNN-based acoustic model training and decoding, which are computed as the difference between the closed-form GMM log-likelihoods obtained with vs. without uncertainty. We concatenated the GMMD features with conventional acoustic features and showed that they improve ASR performance on both the CHiME-2 and CHiME-3 datasets [15].

7.3.1.2. Uncertainty in other applications

Besides the above application, we finalized our exploration of uncertainty modeling for wind turbine control. Baldwin Dumortier defended his PhD thesis on this topic [9].

7.3.2. Uncertainty and phonetic segmentation

In the METAL project, experiments are planned to investigate further the use of speech technologies for foreign language learning in middle and high schools. Besides adapting acoustic models to teenager voices, current work investigates the reliability of speech technologies for analyzing student pronunciations, and for detecting miss-pronunciations. Also, besides making the pronunciation diagnostics more reliable, the aim is to elaborate robust strategies that will make it possible to handle sets of unreliable individual results, and still be able to provide a relevant feedback on recurrent miss-pronunciations.

7.3.3. Uncertainty and prosody

The analysis of prosodic correlates of discourse particles has continued. Some additional data has been annotated. The automatic word and phonetic segmentation of the discourse particles has been manually checked and corrected when necessary. Once more, this has shown that automatic segmentation is not perfect, especially on spontaneous speech recording in real conditions. For each discourse particle, prosodic characteristics of occurrences of each pragmatic function (conclusive, introductive, etc.) were automatically extracted. For each discourse particle and each pragmatic function, the most frequent F0 patterns were retained as the representative forms. Results show that a pragmatic function, common to several discourse particles, gives rise to a uniform prosodic marking [34].

NEUROSYS Project-Team

7. New Results

7.1. From the microscopic to the mesoscopic scale

Participants: Laure Buhry, Axel Hutt Amélie Aussel, Nathalie Azevedo Carvalho.
In collaboration with Radu Ranta (univ. Lorraine), Dominique Martinez (CNRS), Abderrahman Iggidr (Inria), Patrick Hénaff (univ. Lorraine), Beate Knauer and Motoharu Yoshida (Ruhr university) and LieJune Shiau (university of Houston)

7.1.1. Memory & sleep

We proposed a detailed anatomical and mathematical model of the hippocampal formation for the generation of sharp-wave ripples and theta-nested gamma oscillations [1], [7]. Indeed, the mechanisms underlying the broad variety of oscillatory rhythms measured in the hippocampus during the sleep-wake cycle are not yet fully understood. We proposed a computational model of the hippocampal formation based on a realistic topology and synaptic connectivity, and we analyzed the effect of different changes on the network, namely the variation of synaptic conductances, the variations of the CAN channel conductance and the variation of inputs. By using a detailed simulation of intracerebral recordings, we showed that this model is able to reproduce both the theta-nested gamma oscillations that are seen in awake brains and the sharp-wave ripple complexes measured during slow-wave sleep. The results of our simulations support the idea that the functional connectivity of the hippocampus, modulated by the sleep-wake variations in Acetylcholine concentration, is a key factor in controlling its rhythms. A presentation of this work received a best poster award at the 27th annual Computational Neuroscience Meeting, CNS'2018.

7.1.2. Parkinson's network

Using a Hodgkin and Huxley's model, we modeled pathological oscillations of Parkinson's disease in basal ganglia. Our hypothesis was that the pathological oscillations are generated by a MSN-GPeA-FSN circuit and then transferred to the STN by the GPeP. The normal state is represented by neurons of the MSN emitting at a frequency of $\sim 1Hz$ and the parkinsonian state is represented by MSN neurons that emit at a frequency of $\sim 15Hz$. Our results correspond to the experimental results of the rat. In the normal state, there is no visible synchronization, whereas in the parkinsonian state pathological synchronizations are formed at the level of the circuit. There is even a rhythm that is created, that is to say, the neurons of MSN emit first then those of the FSN and then those of the GPeA and so on. We performed large-scale simulations of 1.5 million neurons in the basal ganglia in rats using the Grid5000 (parallel computation platform).

7.2. From the Mesoscopic to the Macroscopic Scale

Participants: Laurent Bougrain, Axel Hutt, Sébastien Rimbert, Oleksii Avilov, Rahaf Al-Chwa.
In collaboration with Stéphanie Fleck (Univ. Lorraine) and Patrick Hénaff (univ. Lorraine)

7.2.1. Motor system

In collaboration with Stéphanie Fleck (Univ. Lorraine)
Kinesthetic motor imagery (KMI) tasks induce brain oscillations over specific regions of the primary motor cortex within the contralateral hemisphere of the body part involved in the process. This activity can be measured through the analysis of electroencephalographic (EEG) recordings and is particularly interesting for Brain-Computer Interface (BCI) applications.

7.2.1.1. Electroencephalographic modulations during an open- or closed-eyes motor task

There is fundamental knowledge that during the resting state cerebral activity recorded by electroencephalography (EEG) is strongly modulated by the eyes-closed condition compared to the eyes-open condition, especially in the occipital lobe. However, little research has demonstrated the influence of the eyes-closed condition on the motor cortex, particularly during a self-paced movement. This prompted the question: How does the motor cortex activity change between the eyes-closed and eyes-open conditions? To answer this question, we recorded EEG signals from 15 voluntary healthy subjects who performed a simple motor task (i.e., a voluntary isometric flexion of the right-hand index) under two conditions: eyes-closed and eyes-open. Our results confirmed strong modulation in the mu rhythm (7–13 Hz) with a large event-related desynchronisation. However, no significant differences have been observed in the beta band (15–30 Hz). Furthermore, evidence suggests that the eyes-closed condition influences the behaviour of subjects [5]. Our study gives greater insight into the motor cortex and could also be useful to improve brain-computer interface (BCI) based on motor imagery.

7.2.1.2. Can a Subjective Questionnaire be used as a Brain-Computer Interface performance Predictor?

Predicting a subject's ability to use a Brain Computer Interface (BCI) is one of the major issues in the BCI domain. Relevant applications of forecasting BCI performance include: the ability to adapt the BCI to the needs and expectations of the user; assessing the efficiency of BCI use in stroke rehabilitation; and finally, homogenizing a research population. A limited number of recent studies have proposed the use of subjective questionnaires, such as, the Motor Imagery Questionnaire Revised-Second Edition (MIQ-RS). However, further research is necessary to confirm the effectiveness of this type of subjective questionnaire as a BCI performance estimation tool. In this study we aim to answer the following questions: can the MIQ-RS be used to estimate the performance of an MI-based BCI? If not, can we identify different markers that could be used as performance estimators? To answer these questions, we recorded EEG signals from 35 voluntary healthy subjects during BCI use. The subjects previously had completed the MIQ-RS questionnaire. We conducted an offline analysis to assess the correlation between the questionnaire scores related to Kinesthetic and Motor imagery tasks and the performances of four classification methods. Our results show no significant correlation between BCI performance and the MIQ-RS scores. However, we reveal that BCI performance is correlated to habits and frequency of practicing manual activities [15] (accepted in Front. Hum. Neurosci. | doi: 10.3389/fnhum.2018.00529).

7.2.1.3. Median nerve stimulation based BCI: a new approach to detect intraoperative awareness during general anesthesia

Hundreds of millions of general anesthesia are performed each year on patients all over the world. Among these patients, 0.1-0.2% are victims of Accidental Awareness during General Anesthesia (AAGA), i.e. an unexpected awakening of the patient during a surgical procedure under general anesthesia. This terrifying experience may be very traumatic for the patient and should be avoided by the anesthesiologists. Out of all the techniques used to prevent these awakenings, there is currently no solution based on the EEG signal to detect this phenomenon efficiently. Since the first reflex for a patient during an AAGA is to move, a passive BCI based on the intention of movement is conceivable. However, the challenge of using such BCI is that the intention to move from the waking patient is not initiated by a trigger that could be used to guide a classifier. We proposed a solution based on Median Nerve Stimulation (MNS), which causes specific modulations in the motor cortex and can be altered by an intention of movement. We showed that MNS may provide a foundation for an innovative BCI that would allow the detection of an AAGA [17].

ORPAILLEUR Project-Team

7. New Results

7.1. Mining of Complex Data

Participants: Nacira Abbas, Guilherme Alves Da Silva, Alexandre Blansché, Lydia Boudjeloud-Assala, Quentin Brabant, Brieuc Conan-Guez, Miguel Couceiro, Adrien Coulet, Alain Gély, Laurine Huber, Nyoman Juniarta, Florence Le Ber, Joël Legrand, Pierre Monnin, Tatiana Makhalova, Amedeo Napoli, Abdelkader Ouali, François Pirot, Frédéric Pennerath, Justine Reynaud, Chedy Raïssi, Sébastien Da Silva, Yannick Toussaint.

Keywords: formal concept analysis, relational concept analysis, pattern structures, pattern mining, association rule, redescription mining, graph mining, sequence mining, biclustering, hybrid mining, meta-mining

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 [73]) allow building 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 [84] and can play an important role in text classification and text mining. 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 got several results in the discovery of approximate functional dependencies [8], the mining of RDF data and the and visualization of the discovered patterns [1], and redescription mining (detailed later). Moreover, we have also investigated the use of the MDL principle ("Minimum Description Length") for the selection of interesting and diverse patterns [37], [39].

In the framework of the CrossCult European Project about cultural heritage, we worked on the mining of visitor trajectories in a museum or a touristic site. We presented a theoretical and practical research work about the characterization of visitor trajectories and the mining of these trajectories as sequences [32], [33]. The mining process is based on two approaches in the framework of Formal Concept Analysis (FCA). We focused on different types of sequences and more precisely on subsequences without any constraint and frequent contiguous subsequences. In parallel, we introduced a similarity measure allowing us to build a hierarchical classification which is used for interpretation and characterization of the trajectories. In addition, for completing the research work on the characterization of trajectories, we also studied how biclustering may be applied to trajectory recommendation [31], [52].

7.1.2. Redescription Mining

Among the mining methods developed in the team is redescription mining. Redescription mining aims to find distinct common characterizations of the same objects and, vice versa, to identify sets of objects that admit multiple shared descriptions [82]. It is motivated by the idea that in scientific investigations data oftentimes have different nature. For instance, they might originate from distinct sources or be cast over separate terminologies. In order to gain insight into the phenomenon of interest, a natural task is to identify the correspondences that exist between these different aspects.

A practical example in biology consists in finding geographical areas that admit two characterizations, one in terms of their climatic profile and one in terms of the occupying species. Discovering such redescriptions can contribute to better our understanding of the influence of climate over species distribution. Besides biology, applications of redescription mining can be envisaged in medicine or sociology, among other fields.

This year, we used redescription mining for analyzing and mining RDF data with the objective of discovering definitions of concepts and as well disjunctions (incompatibilities) of concepts, for completing knowledge bases in a semi-automated way [49], [44].

7.1.3. Text Mining

In the context of the PractikPharma ANR Project, we study how cross-corpus training may guide the task of relationship extraction from texts, and especially, how large annotated corpora developed for alternative tasks may improve the performance of biomedical tasks, for which only a few annotated resources are available [34].

Transfer learning proposes to enhance machine learning performance on a problem, by reusing labeled data originally designed for a related problem. This is particularly relevant to the applications of deep learning in Natural Language Processing, because those usually require large annotated corpora that may not exist for the targeted domain, but exist for side domains. In a recent work, we experimented the extraction of relationships from biomedical texts with two deep learning models. The first model combines locally extracted features using a Multi Channel Convolutional Neural Network (MCCNN) model, while the second model exploits the syntactic structure of sentences using a Tree-LSTM (Long Short-Term Memory) architecture. The experiments show that the Tree-LSTM model benefits from a cross-corpus learning strategy, i.e. performances are improved when training data are enriched with off-target corpora, whereas it is not the case with MCCNN.

Indeed our approach leads to state of the art performances in four biomedical tasks for which only a few annotated resources are available (less than 400 manually annotated sentences) and even surpass state of the art performances in two of these four tasks. We particularly investigated how the syntactic structure of a sentence, which is domain independent, participates in the increase of performance when adding additional training data. This may have a particular impact in specialized domains in which training resources are scarce, because it means that these resources may be efficiently enriched with data from other domains for which large annotated corpora exist.

7.1.4. Mining subgroups as a single-player game

Discovering patterns that strongly distinguish one class label from another is a challenging data-mining task. The unsupervised discovery of such patterns would enable the construction of intelligible classifiers and to elicit interesting hypotheses from the data. Subgroup Discovery (SD) is one framework that formally defines this pattern mining task. However, SD still faces two major issues: (i) how to define appropriate quality measures to characterize the uniqueness of a pattern; (ii) how to select an accurate heuristic search technique when exhaustive enumeration of the pattern space is unfeasible. The first issue has been tackled by the Exceptional Model Mining (EMM) framework. This general framework aims to find patterns that cover tuples that locally induce a model that substantially differs from the model of the whole dataset. The second issue has been studied in SD and EMM mainly with the use of beam-search strategies and genetic algorithms for discovering a pattern set that is non-redundant, diverse and of high quality. Consequently,

In our current work [9], we proposed to formally define pattern mining as a single-player game, as in a puzzle, and to solve it with a Monte Carlo Tree Search (MCTS), a technique mainly used for artificial intelligence and planning problems. The exploitation/exploration trade-off and the power of random search of MCTS lead to an any-time mining approach, in which a solution is always available, and which tends towards an exhaustive search if given enough time and memory. Given a reasonable time and memory budget, MCTS quickly drives the search towards a diverse pattern set of high quality. MCTS does not need any knowledge of the pattern quality measure, and we show to what extent it is agnostic to the pattern language.

7.1.5. Consensus and Aggregation Functions

Aggregation and consensus theory study processes dealing with the problem of merging or fusing several objects, e.g., numerical or qualitative data, preferences or other relational structures, into a single or several objects of similar type and that best represents them in some way. Such processes are modeled by so-called aggregation or consensus functions [76], [78]. The need to aggregate objects in a meaningful way appeared naturally in classical topics such as mathematics, statistics, physics and computer science, but it became

increasingly emergent in applied areas such as social and decision sciences, artificial intelligence and machine learning, biology and medicine.

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. It also requires a representation formalism for knowledge, in our case decision rules and methods for discovering them. Typical approaches include rough-set and FCA approaches, that we aim to extend in order to increase expressivity, applicability and readability of results. Applications of these efforts already appeared and further are expected in the context of three multidisciplinary projects, namely the "Fight Heart Failure" (research project with the Faculty of Medicine in Nancy), the European H2020 "CrossCult" project, and the "ISIPA" (Interpolation, Sugeno Integral, Proportional Analogy) project.

In the context of the project RHU "Fighting Heart Failure" (that aims to identify and describe relevant bio-profiles of patients suffering from heart failure) we are dealing with biomedical data, highly complex and heterogeneous, that include, among other, sociodemographical aspects, biological and clinical features, drugs taken by the patients, etc. One of our main challenges is to define relevant aggregation operators on this heterogeneous patient data that lead to a clustering of the patients. Each cluster should correspond to a bio-profile, i.e. a subgroup of patients sharing the same form of the disease and thus the same diagnosis and medical care strategy. We are working on ways for comparing and clustering patients, namely, by defining multidimensional similarity measures on this complex and heterogeneous biomedical data. To this end, we recently proposed a novel approach, that we named "unsupervised extremely randomized trees" (UET) [27], that is inspired by the frameworks of unsupervised random forests (URF) [85] and of extremely randomized trees (ET) [75]. The empirical study of UET showed that it outperforms existing methods (such as URF) in running time, while giving better clustering. However, UET was implemented for numerical data only, and this is a drawback when dealing with biomedical data. We are now working on the adaptation of UET for heterogeneous data (both numerical and symbolic), possibly, with missing values.

In the context of the project ISIPA, we mainly focused on the utility-based preference model in which preferences are represented as an aggregation of preferences over different attributes, structured or not, both in the numerical and qualitative settings. In the latter case, the Sugeno integral is widely used in multiple criteria decision making and decision under uncertainty, for computing global evaluations of items based on local evaluations (utilities). The combination of a Sugeno integral with local utilities is called a Sugeno utility functional (SUF). A noteworthy property of SUFs is that they represent multi-threshold decision rules. However, not all sets of multi-threshold rules can be represented by a single SUF. We showed how to represent any set of multi-threshold rules as a combination of SUFs and studied their potential advantages as a compact representation of large sets of rules, as well as an intermediary step for extracting rules from empirical datasets [51]. For further results in the qualitative approach to decision making see, e.g., [10] [3]; and see also [24] for a survey chapter on new perspectives in ordinal evaluation.

7.2. Knowledge Discovery in Healthcare and Life Sciences

Participants: Miguel Couceiro, Adrien Coulet, Nicolas Jay, Joël Legrand, Pierre Monnin, Amedeo Napoli, Abdelkader Ouali, Chedy Raïssi, Malika Smaïl-Tabbone, Yannick Toussaint.

7.2.1. Ontology-based Clustering of Biological Data

Biomedical objects can be characterized by ontology annotations. For example, Gene Ontology annotations provide information on the functions of genes, while Human Phenotype Ontology (HPO) annotations provide information about phenotypes associated with diseases. It is usual to consider such annotations in the analysis of biomedical data, most of the time annotations from only one single ontology. However, complex objects such as diseases can be annotated at the same time w.r.t. different ontologies, making clear distinct dimensions. We are investigating how annotations from several ontologies may be cooperating in disease classification. In

particular, we classified Genetic Intellectual Disabilities (GID), on the basis of their HPO annotations and of GO annotations of genes known for being responsible for these diseases [43]. We used clustering algorithms based on semantic similarities and enabling to compare sets of annotations. This experiment illustrates the fact that considering several ontologies provides better results, while selecting the best set of ontologies to combine is dependent on the dataset and on the classification task.

7.2.2. Validation of Pharmacogenomic Knowledge

State of the art knowledge in pharmacogenomics is heterogeneous w.r.t. validation. A part is well validated, observed on a large population and already used in clinical practice, while a large majority of this knowledge is lacking validation and reproducibility, mainly because of scarce observation. Accordingly, validating state of the art knowledge in pharmacogenomics by mining Electronic Health Records (EHRs) is one objective of the ANR project "PractiKPharma" initiated in 2016 (http://practikpharma.loria.fr/).

To lead this validation, we define a minimal data schema for pharmacogenomic knowledge units (PGxO ontology), which is instantiated with data of various provenance (e.g. biomedical databases, literature and EHR). Such an instantiation produces a unique knowledge graph named PGxLOD (https://pgxlod.loria.fr/). We defined and applied a first set of reconciliation rules that compare and align whenever possible knowledge elements of various provenance. A journal article on the construction of PGxLOD and its use in knowledge comparison is currently under evaluation. We are continuing this effort by studying methods which enable a more flexible knowledge comparison.

In addition, we took part to the Biohackathon 2018 Paris (https://bh2018paris.info/) during which we worked on two tasks. Firstly we updated PGxLOD for improving its quality, completeness and interconnection with other resources. Secondly we mined PGxLOD and searched for explanations of the molecular mechanism of adverse drug responses. PGxLOD is under evaluation for being registered as a resource of the IBF (*French Institute for Bioinformatics*) and Elixir (an international organization that supports and structures bioinformatics efforts in Europe).

7.2.3. Mining Electronic Health Records

In the context of the Snowball Inria Associate Team, we developed an approach based on pattern structures to identify frequently associated ADRs (Adverse Drug Reactions) from patient data either in the form of EHR or ADR spontaneous reports. Pattern structures provide an expressive representation of ADR, taking into account the multiplicity of drugs and phenotypes involved in such reactions. Additionally, pattern structures allow considering diverse biomedical ontologies used to represent or annotate patient data, enabling a "semantic" comparison of ADRs. Up to now, this is one of the first research attempts considering such representations to mine rules between frequently associated ADRs. We illustrated the generality of the approach on two patient datasets, each of them linked to distinct biomedical ontologies. The first dataset corresponds to anonymized EHRs, extracted from "STRIDE", the EHR data warehouse of Stanford Hospital and Clinics. The second dataset is extracted from the U.S. FDA (for Food & Drug Administration) "Adverse Event Reporting System" (FAERS). Several significant association rules have been extracted, analyzed and may be used as a basis for a recommendation system.

In collaboration with Stanford University and the CHRU Nancy, we studied the use of Electronic Health Records to predict at first prescription the need for a patient to be prescribed with a reduced drug dose [4]. We particularly focused on drugs whose dosage is known to be sensitive and variable. We used data from the Stanford Hospital to construct cohorts of patients that either did or did not need a dose change for each considered drug. After feature selection, we trained Random Forest models which successfully predict whether a new patient will or not require a dose change after being prescribed one of 23 drugs among 22 drug classes. Several of these drugs are related to clinical guidelines that recommend dose reduction exclusively in the case of adverse reaction. For these cases, a reduction in dosage may be considered as a surrogate for an adverse reaction, which our system could help predicting and preventing.

7.3. Knowledge Engineering and Web of Data

Participants: Nicolas Jay, Florence Le Ber, Jean Lieber, Amedeo Napoli, Emmanuel Nauer, Justine Reynaud, Yannick Toussaint.

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

7.3.1. Current Trends in Case-Based Reasoning

Case-based reasoning (CBR) aims at solving a new problem, called the target problem, by exploiting past experiences (i.e. source cases) as well as other knowledge sources: domain knowledge, similarity knowledge and adaptation knowledge.

Two research works were carried out about how exploiting at the best the source cases. A first work addresses the exploitation of negative cases for adaptation knowledge discovery. Usually CBR exploits positive source cases consisting of a source problem and its solution that is known to be correct for the problem. However, negative cases, i.e. problem-solution pairs where the solution is an incorrect answer to the problem, which can be acquired when CBR process fails, are useful, especially for adaptation knowledge discovery. In [29], we propose an adaptation knowledge discovery approach exploiting both type of cases (positive and negatives cases), using closed itemsets built on variations between cases. Experiments show that exploiting negative cases in addition to positive ones improves the quality of the adaptation knowledge being extracted and, so, improves the results of the CBR system.

A second work addresses the issue of the selection of source cases used to solve a target problem. Three approaches have been studied to better exploit source cases: (1) approximation, which considers the use of one source case (the most similar to the target problem) to solve the target problem, (2) interpolation, which considers the use of two source cases (such as the target problem is between these two similar source problems), and (3) extrapolation, which considers the use of three source cases, linked to the target problem by an analogical proportion, where the analogical proportion handles both similarity and dissimilarity between cases. Experiments show that interpolation and extrapolation techniques are of interest for reusing cases, either in an independent or in a combined way [36], [47].

Using analogical proportion has also been used to find relevant pathology-gene pairs [28]. This first study to infer pathology-gene relation is based on the following hypothesis: if a target pathology is in analogy with three other pathologies for which associated genes are known, then it is plausible that the gene to be associated with the target pathology is in analogy with the genes associated to the three pathologies involved in the analogical proportion.

Another use of analogical proportion is its application to machine translation and is based on a similar principle: if four sentences form an analogical proportion in a language, then it is plausible that their translations in another language also form an analogical proportion. This was the idea developed by Yves Lepage (Waseda University), a few years ago. Now, a starting work on case-based machine translation aims at developing these ideas by incorporation other knowledge sources to the CBR system than the cases (domain knowledge, retrieval knowledge and adaptation knowledge) [35].

Another work on CBR is its application to medical coding. Cancer registries are important tools in the fight against cancer. At the heart of these registries is the data collection and coding process. Ruled by complex international standards and numerous best practices, operators are easily overwhelmed. In [54], [55], a system is presented to assist operators in the interpretation of best medical coding practices.

There has been another work on CBR related to an application in agronomy developed some time ago that has been synthesized in [60].

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, following the increase of data published in RDF (Resource Description Framework) format and the interest in machine processable data. The 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. In the

team, we are particularly interested in the completeness of the data viewed as their their potential to provide concept definitions in terms of necessary and sufficient conditions [69]. We have proposed a novel technique based on Formal Concept Analysis which classifies subsets of RDF data into a concept lattice [83]. This allows data exploration as well as the discovery of implication rules which are used to automatically detect possible completions of RDF data and to provide definitions. Moreover, this is 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 [44].

In the same way, FCA can be used to improve ontologies associated with the Web of data. Accordingly, we proposed a method to build a concept lattice from linked data and compare the structure of this lattice with an ontology used to type the considered data. The result of this comparison makes clear some alternative axioms to be proposed to ontology developers. We extended and reused this work in ontology alignment tasks [41].

PESTO Project-Team

7. New Results

7.1. Security protocols

7.1.1. Analysis of equivalence properties

Participants: Vincent Cheval, Véronique Cortier, Antoine Dallon, Ivan Gazeau, Steve Kremer, Joseph Lallemand, Itsaka Rakotonirina, Christophe Ringeissen.

Automatic tools based on symbolic models have been successful in analyzing security protocols. These tools are particularly well adapted for trace properties (e.g. secrecy or authentication). A wide range of security properties, such as anonymity properties in electronic voting and auctions, unlinkability in RFID protocols and mobile phone protocols, are however naturally expressed in terms of indistinguishability, which is not a trace property. Indistinguishability is naturally formalized as an observational or trace equivalence in cryptographic process calculi, such as the applied pi calculus. While several decision procedures have already been proposed for verifying equivalence properties the resulting tools are often rather limited, and lack efficiency.

Our results are centered around the development of several, complementary verification tools for verifying equivalence properties. These tools are complementary in terms of expressivity, precision and efficiency.

- The *Akiss* tool provides good expressivity as it supports a large number of cryptographic primitives (including the XOR primitive, extremely popular in low energy devices such as RFID tags) and protocols with else branches. It allows verification for a bounded number of protocol sessions. The tool is precise for a class of determinate processes, and can approximate equivalence for other protocols. The tool however suffers from efficiency problems when the number of sessions increases. The computation can be partially distributed on different cores. To overcome these efficiency problems of the *Akiss* tool, Gazeau and Kremer completely revisit the theory underlying *Akiss*. Rather than enumerating the possible traces, the new version directly reasons about partial ordered traces. A new implementation is also in progress and the first results seem extremely promising.
- The SAT-Equiv tool is based on a novel algorithm, based on graph planning and SAT-solving. The tool has a limited expressivity in that it allows only the most standard cryptographic primitives, requires protocols to be determinate and does not support protocols with else branches. The tool is however extremely efficient, allowing verification for a very large (but bounded) number of sessions (where most other tools have to stop after one or two sessions). Cortier and Dallon, in collaboration with Delaune (IRISA), have presented at ESORICS'18 [20] an extension of SAT-EQUIV to support protocols with phases and a large class of cryptographic primitives that encompasses standard primitives. This required to first show a small attack property: whenever two protocols are not in equivalence, there exists a well-typed witness of non equivalence. This result was initially proved for symmetric encryption only and now holds for a large class of primitives [37].
- The DEEPSEC tool, presented by Cheval, Kremer and Rakotonirina at S&P'18 [18], is a new tool that allows for user-defined cryptographic primitives that can be modelled as a subterm convergent rewrite system (slightly more restricted than AKISS), but supports the whole applied pi calculus, except for bounding the number of sessions. It is precise, in that it decides equivalence (without any approximations) and has good efficiency (slightly less than SAT-Equiv) for the class of determinate processes (where partial order reductions apply). Their work also settled the question of the exact complexity of deciding different equivalences - static equivalence, trace equivalence and bisimulation. In particular they were able to show that both deciding trace equivalence and bisimulation in the case of cryptographic primitives modelled by subterm convergent rewrite systems are co-NEXP complete problems – this is a strong, new insight, solving a longstanding open question about the complexity of this problem. The DEEPSEC tool also implements state of the art partial order reductions and the verification can be distributed on different cores on a single machine and also on clusters of machines, as detailed in a CAV'18 tool paper [19].

- Unlike the above tools, the TYPE-EQ tool supports verification of both a bounded and unbounded number of protocol sessions (and a mix of them). It is based on a novel approach for equivalence properties. Instead of *deciding* equivalence like for the previous approaches, the tool uses a type system which is sound w.r.t. equivalence. Regarding precision, the tool is not complete, i.e. it may provide false attacks. It induces a significant speedup compared to previous tools for a bounded number of sessions and compares similarly to ProVerif [47] for an unbounded number of sessions. In collaboration with Maffei and Grimm, Lallemand and Cortier [23] extend this approach to all standard primitives and improve its precision, allowing to branch on secrets.

From a more foundational point of view, Ringeissen, in collaboration with Erbatur (LMU, Germany) and Marshall (Univ Mary Washington, USA), study decision procedures for two knowledge problems critical to the verification of security protocols, namely the intruder deduction and the static equivalence problems. These problems can be related to particular forms of context matching and context unification. Both problems are defined with respect to an equational theory and are known to be decidable when the equational theory is given by a subterm convergent term rewrite system. In a paper presented at UNIF'18 [33] they investigate the case of a subterm convergent equational term rewrite system defined modulo an equational theory, like Commutativity or Associativity-Commutativity. They show that for certain classes of such equational theories, namely the shallow classes, the two knowledge problems remain decidable.

7.1.2. Verification of protocols with global states

Participants: Vincent Cheval, Véronique Cortier, Jannik Dreier, Mathieu Turuani.

One known challenge when analysing security protocols for an unbounded number of sessions is the case of protocols with global states such as counters, tables, or more generally, memory cells. The popular tool ProVerif [47] fails to analyse such protocols, due to its internal abstraction. Cheval, Cortier, and Turuani have devised a generic transformation of the security properties queried to ProVerif. In a paper presented at CSF'18 [17], they proved the soundness of the transformation and implement it into a front-end GSVerif. Their experiments show that GSVerif (combined with ProVerif) outperforms the few existing tools, both in terms of efficiency and protocol coverage. GSVerif was successfully applied to a dozen of protocols of the literature, yielding the first fully automatic proof of a security API and a payment protocol of the literature.

The *TAMARIN* prover is a state-of-the-art verification tool for cryptographic protocols in the symbolic model. Dreier, in collaboration with Hirschi, Sasse (ETH Zurich), and Radomirovic (Dundee), improved the underlying theory and the tool to deal with an equational theory modeling XOR operations. Exclusive-or (XOR) operations are common in cryptographic protocols, in particular in RFID protocols and electronic payment protocols. Although there are numerous applications, due to the inherent complexity of faithful models of XOR, there is only limited tool support for the verification of cryptographic protocols using XOR. This makes *TAMARIN* the first tool to support simultaneously this large set of equational theories, protocols with global mutable state, an unbounded number of sessions, and complex security properties including observational equivalence. We demonstrated the effectiveness of our approach by analyzing several protocols that rely on XOR, in particular multiple RFID-protocols, where we can identify attacks as well as provide proofs. These results were presented at CSF'18 [29].

7.1.3. Analysis of deployed protocols

Participants: Jannik Dreier, Charlie Jacomme, Steve Kremer.

7.1.3.1. Multi-factor authentication.

Passwords are still the most widespread means for authenticating users, even though they have been shown to create huge security problems. This motivated the use of additional authentication mechanisms used in so-called multi-factor authentication protocols. In a paper, published at CSF'18 [30] Jacomme and Kremer define a detailed threat model for this kind of protocols: while in classical protocol analysis attackers control the communication network, the idea is to take into account that many communications are performed over TLS channels, that computers may be infected by different kinds of malwares, that attackers could perform phishing, and that humans may omit some actions. This model has been formalized in the applied pi calculus

and perform an extensive analysis and comparison of several widely used protocols — variants of Google 2 step and FIDO U2F. The analysis is completely automated, generating systematically all combinations of threat scenarios for each of the protocols and using the ProVerif tool [47] for automated protocol analysis. Even though threat scenarios are eliminated as soon as results are implied by weaker scenarios, the analysis required over 6 000 calls to ProVerif, yet finishes in only a few minutes. Their analysis highlights weaknesses and strengths of the different protocols, and allows them to suggest several small modifications of the existing protocols which are easy to implement, yet improve their security in several threat scenarios.

7.1.3.2. 5G Authentication.

Mobile communication networks connect much of the world's population. The security of users' calls, SMSs, and mobile data depends on the guarantees provided by the Authenticated Key Exchange protocols used. For the next-generation network (5G), the 3GPP group has standardized the 5G AKA protocol for this purpose. We provided the first comprehensive formal model of a protocol from the AKA family: 5G AKA. We also extracted precise requirements from the 3GPP standards defining 5G and we identified missing security goals. Using the security protocol verification tool Tamarin and its recent extension to support XOR, we conducted a full, systematic, security evaluation of the model with respect to the 5G security goals. Our automated analysis identifies the minimal security assumptions required for each security goal and we found that some critical security goals are not met, except under additional assumptions missing from the standard. Finally, we made explicit recommendations with provably secure fixes for the attacks and weaknesses we found. These results were presented at CCS'18 [13].

7.1.3.3. Authentication Methods with PIN Codes.

Touch screens have become ubiquitous in the past few years, like for instance in smartphones and tablets. These devices are often the entry door to numerous information systems, hence having a secure and practical authentication mechanism is crucial. In this work, we examined the complexity of different authentication methods specifically designed for such devices. We studied the common technology to authenticate a user using a Personal Identifier Number code (PIN code). Entering the code is a critical moment where there are several possibilities for an attacker to discover the secret. We considered three attack models: a Bruteforce Attack (BA) model, a Smudge Attack (SA) model, and an Observation Attack (OA) model where the attacker sees the user logging in on his device. The aim of the intruder is to learn the secret code. Our goal is to propose alternative methods to enter a PIN code. We compared such different methods in terms of security. Some methods require more intentional resources than other, this is why we performed a psychological study on the different methods to evaluate the users' perception of the different methods and their usage. This work was presented at RCIS'18 [16].

7.1.4. Protocol design

Participant: Jannik Dreier.

7.1.4.1. A Cryptographer's Conspiracy Santa.

In Conspiracy Santa, a variant of Secret Santa, a group of people offer each other Christmas gifts, where each member of the group receives a gift from the other members of the group. To that end, the members of the group form conspiracies, to decide on appropriate gifts, and usually divide the cost of the gift among all participants of the conspiracy. This requires to settle the shared expenses per conspiracy, so Conspiracy Santa can actually be seen as an aggregation of several shared expenses problems. In this work, we showed that the problem of finding a minimal number of transactions when settling shared expenses is NP-complete. Still, there exists good greedy approximations. Second, we presented a greedy distributed secure solution to Conspiracy Santa. This solution allows a group of people to share the expenses for the gifts in such a way that no participant will learn the price of his/her gift, but at the same time notably reduces the number of transactions with respect to a naive aggregation. Furthermore, our solution does not require a trusted third party, and can either be implemented physically (the participants are in the same room and exchange money) or, virtually, using a cryptocurrency. This work was presented at FUN'18 [14].

7.1.4.2. A Physical Zero-Knowledge Proof for Makaro.

Makaro is a logic game similar to Sudoku. In Makaro, a grid has to be filled with numbers such that: given areas contain all the numbers up to the number of cells in the area, no adjacent numbers are equal, and some cells provide restrictions on the largest adjacent number. In this work we proposed a proven secure physical algorithm, only relying on cards, to realize a zero-knowledge proof of knowledge for Makaro. It allows a player to show that he/she knows a solution without revealing it. This work was presented at SSS'18 [15].

7.2. E-voting

7.2.1. Definitions for e-voting

Participants: Sergiu Bursuc, Véronique Cortier, Steve Kremer, Joseph Lallemand.

Electronic voting typically aims at two main security goals: vote privacy and verifiability. Verifiability typically includes individual verifiability (a voter can check that his/her ballot is counted); universal verifiability (anyone can check that the result corresponds to the published ballots); and eligibility verifiability (only legitimate voters may vote). Cortier and Lallemand have shown that privacy actually implies individual verifiability. In other words, systems without individual verifiability cannot achieve privacy (under the same trust assumptions). To demonstrate the generality of the result, they show this implication in two different settings, namely cryptographic and symbolic models, for standard notions of privacy and individual verifiability. This also highlights limitations in existing privacy definitions in cryptographic settings. This work has been presented at CCS'18 [24].

Some modern e-voting systems take into account that the platform used for voting may be corrupted, e.g. infected by malware, yet aiming to ensure privacy and integrity of votes even in that case. Bursuc and Kremer, in collaboration with Dragan (Univ of Surrey) propose a new definition of vote privacy, formalized in the cryptographic model as a computational indistinguishability game. The definition captures both known and novel attacks against several voting schemes, and they propose a scheme that is provably secure in this setting. Moreover the proof is formalized and machine-checked in the EasyCrypt theorem prover [45]. This result is currently under submission for publication.

7.2.2. Analysis of e-voting protocols

Participants: Véronique Cortier, Mathieu Turuani.

Belenios is a voting platform designed by our team in collaboration with the Caramba research group at Inria Nancy. Cortier, in collaboration with Warinschi (Univ Bristol), Dragan and Dupressoir (Univ of Surrey), has developed a machine-checked security proof of both privacy and verifiability of Belenios, in the computational model. For this, a novel framework has been developed for proving strong verifiability in EasyCrypt. In the process, several aspects of the pen-and-paper proof of Belenios have been clarified, such as how to deal with revote policies. The framework and the security proofs have been presented at CSF'18 [21].

Turuani and Cortier, in collaboration with Galindo (Univ Birmingham), have analysed the e-voting protocol developed by the Scytl company and planned to be deployed in Switzerland. The formal analysis of both privacy and individual verifiability has been conducted in ProVerif. It required the development of a crafty encoding of the security properties in order to avoid the limitations of ProVerif in the presence of global states (here, no revoting). This first encoding yielded the preliminary ideas for the GSVerif tool mentioned in the previous section. Such a formal analysis is required by the Swiss Chancellerie and has been presented at EuroSP'18 [22].

7.2.3. Design of e-voting protocols

Participants: Véronique Cortier, Alicia Filipiak, Joseph Lallemand.

Most existing voting systems either assume trust in the voting device or in the voting server. Filipiak, Lallemand, and Cortier proposed a novel Internet voting scheme, BeleniosVS, that achieves both privacy and verifiability against a dishonest voting server as well as a dishonest voting device. In particular, a voter does not leak her vote to her voting device and she can check that her ballot on the bulletin board does correspond to her intended vote. Additionally, our scheme guarantees receipt-freeness against an external adversary. A formal proof of privacy, receipt-freeness, and verifiability has been established using the tool ProVerif, covering a hundred cases of threat scenarios. Proving verifiability required the identification of a set of sufficient conditions, that can be handled by ProVerif [47]. This contribution is of independent interest. This work is part of the PhD thesis [10] of Alicia Filipiak, defended in March 2018. A conference paper is under submission.

7.3. Privacy

7.3.1. Privacy Protection in Social Networks

Participants: Younes Abid, Bizhan Alipour, Sourya Joyee De, Abdessamad Imine, Michaël Rusinowitch.

To increase awareness about privacy threats, we have designed a tool, SONSAI, for Facebook users to audit their own profiles. SONSAI predicts values of sensitive attributes by machine learning and identifies user public attributes that have guided the learning algorithm towards these sensitive attribute values. The tool is designed to perform reasonably with the limited resources of a personal computer, by collecting and processing only a small relevant part of the network data [31], [32]. We also show how SONSAI is fully interfaced with Facebook along different scenarios. In each case a dataset was built from real profiles collected in the user's neighbourhood network. The whole analysis process is performed online, mostly automatically and with an accuracy of 0.79 when inferring political orientation. More details on the inference of other sensitive attributes are given in [8]. We are now investigating potential privacy attacks based on other data types such as posts, comments and images.

Online social network profiles help users to build new friendships as well as reviving and enhancing existing ones. However, users can become the victims of privacy harms such as identity theft, stalking or discrimination due to the personal data revealed in these profiles. So they have to carefully select the privacy settings for their profile attributes, keeping in mind this trade-off between privacy and social benefit. To aid in this decision process, we have developed a user-friendly model based on Integer Programming [27]. Our model provides a social network user with easy-to-implement suggestions about the privacy settings of his profile attributes such that he can achieve the maximum social benefit while protecting himself from all or at least some major privacy risks. We have tested our approach on user profiles with varying vicinities (i.e. the list of friends) and social benefit requirements [25].

Users' interactions must consider both privacy risks and social benefits, a view supported by the EU General Data Protection Regulation (GDPR). In addition, the GDPR recognizes user consent as a legitimate ground for data processing. In [26], we analyze the present status of user consent in online social networks and we observe that evaluating the privacy risks of user consents to data processing activities can be an effective way to help users in their decision to give or refuse consent.

7.3.2. Compressed and Verifiable Filtering Rules in Software-defined Networking

Participants: Ahmad Abboud, Michaël Rusinowitch.

In a joint project with the Resist research group at Inria Nancy and the Cynapsys/Numeryx companies, we are working on the design, implementation and evaluation of a double-mask technique for building compressed and verifiable filtering rules in Software Defined Networks with the possibility of distributing the workload processing among several packet filtering devices operating in parallel.

RESIST Team

7. New Results

7.1. Monitoring

7.1.1. HTTPS traffic monitoring

Participants: Jérôme François [contact], Pierre-Olivier Brissaud, Olivier Bettan [Thales], Isabelle Chrisment, Thibault Cholez.

While privacy is empowered by encrypted communications such as through the HTTPS protocol, it is also legitimated to allow network monitoring of HTTPS traffic. To be compliant with privacy, we proposed a transparent and passive technique that only detects if an HTTPS request is related to a previously defined action [7]. Our technique is able to detect forbidden searches over a web service such as Google Images. It differs from related work that either focuses on detecting the type of traffic or the used web service. To achieve a high accuracy, our technique relies on learning stage where keywords to be monitored are crawled before we leverage KDE (Kernel Density Estimation). KDE allows us to construct a signature summarizing the sizes of the loaded objects on a page, which strongly depend on the user action or search.

7.1.2. Monitoring Programmable Networks

Participants: Jérôme François [contact], Olivier Festor, Paul Chaignon [Orange Labs], Kahina Lazri [Orange Labs], Thibault Delmas [Orange Labs].

SDN-based monitoring allows us to gather more valuable indicators by specifying or programming the monitoring with a fine granularity. We proposed to use eBPF (extended Berkeley Packet Filter) to apply fine-grained filtering in comparison to OpenFlow. It brings safety guarantees regarding program execution and allows stateful programs. In order to limit the impact on the throughput, we integrated our solution within the regular packet processing pipeline of Open vSwitch, a major software switch for OpenFlow, by extending the cache mechanisms [8].

7.1.3. Predictive Security Monitoring for Large-Scale Internet-of-Things

Participants: Jérôme François [contact], Rémi Badonnel, Abdelkader Lahmadi, Isabelle Chrisment, Adrien Hemmer.

The Internet-of-Things has become a reality with numerous protocols, platforms and devices being developed and used to support the growing deployment of smart services. Providing new services requires the development of new functionalities, and the elaboration of complex systems that are naturally a source of potential threats. Real cases recently demonstrated that the IoT can be affected by naïve weaknesses. Therefore, security is of paramount importance. In the last decade, many IoT architectures have been proposed. However, security cannot be guaranteed without failure or by-design. In that context, we are currently investigating predictive security monitoring strategies for large-scale Internet-of-Things. In particular, we are considering the building of behavioral models characterizing such complex networks. The objective is to support both the detection of malicious activities, as well as the selection of security counter-measures.

7.1.4. Quality of Experience Monitoring

Participants: Isabelle Chrisment [contact], Thibault Cholez, Antoine Chemardin, Vassili Rivron [University of Caen], Lakhdar Meftah [University of Lille].

We have pursued our work on smartphone usage monitoring with the SPIRALS team (Inria/Université de Lille) and more specifically on proposing new methods to help measure the QoE and to protect the user's privacy when collecting such data.

In the context of the BottleNet project, to build an adequate instrumented investigation system (mobile applications combining measurements and questionnaires), we decomposed, with a group of students, the network quality concept and the perception of the services in several different approaches. These students worked on bibliographic research, on the smartphone usage and on the perception of the Internet. Structured debates on social issues associated with mobile connectivity were organized. The following topics were dealt: Quality of Service/Quality of Experience; rhythms of life and routines; privacy: diversity of practices and ethical issues; advertising and free: volume, exposure, perception, third-party and cost; quantified self-*: relation to self-quantification; online cultural consumption; information practices on mobile; communication practices.

In the context of the IPL BetterNet project, we continued to work on federating Inria's monitoring tools (APISENSE®, Fathom, Hostview, ACQUA) in a common measurement platform. A first test campaign has been performed with a small set of volunteer users to evaluate the full data collection system built from all these tools.

7.2. Experimentation

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

7.2.1. Grid'5000 design and evolutions

Participants: Florent Didier, Alexandre Merlin, Lucas Nussbaum [contact], Olivier Demengeon [SED], Teddy Valette [SED].

The team was again heavily involved in the evolutions and the governance of the Grid'5000 testbed.
Technical team management Since the beginning of 2017, Lucas Nussbaum serves as the *directeur technique* (CTO) of Grid'5000 in charge of managing the global technical team (9 FTE).
SILECS project We are also heavily involved in the ongoing SILECS project, that aims to create a new infrastructure on top of the foundations of Grid'5000 and FIT in order to meet the experimental research needs of the distributed computing and networking communities. Since 2018, SILECS has been listed as part of the French National Roadmap for Very Large Research Infrastructures (TGIR program).
Grid'5000/FIT school We had a central role in the organization of the Grid'5000/FIT school that took place in Sophia-Antipolis in April 2018, gathering 93 participants. Lucas Nussbaum delivered a keynote talk presenting Grid'5000 and its recent evolutions [28]. A successful evaluation of Grid'5000 by its Scientific Advisory Board also took place during the school.
Storage manager A contribution from the team was the design and development of a new storage access manager that allows secure access to NFS home directories, thus closing a widely-spread security vulnerability.

7.2.2. I/O emulation support in Distem

Participants: Alexandre Merlin, Olivier Dautricourt, Abdulqawi Saif, Lucas Nussbaum [contact].

Distem had a new release (version 1.3) at the beginning of 2018. This release mainly focused on bringing it up-to-date in terms of software quality (newer dependencies, added tests) and added some network emulation features that were previously missing.

The emulator was then featured in a tutorial during the Grid'5000/FIT school.

There is ongoing work on adding I/O emulation support in Distem, in order to experiment how Big Data solution can handle degraded situations. This is still pending completion and publication.

7.2.3. I/O access patterns analysis with eBPF

Participants: Abdulqawi Saif, Lucas Nussbaum [contact], Ye-Qiong Song.

We explored the relevance of an emerging instrumentation technology for the Linux kernel, eBPF, and used it to analyze I/O access patterns such as non-sequential accesses, which are particularly harmful on non-SSD drives. We designed a tool to help with such analysis, and applied it to two popular NoSQL databases, MongoDB and Cassandra, outlining severe performance problems [19] with MongoDB, where a workload that should have resulted in sequential accesses was in fact turned into lots of random accesses.

7.2.4. Experiment Monitoring

Participants: Abdulqawi Saif, Alexandre Merlin, Lucas Nussbaum [contact], Ye-Qiong Song.

Most computer experiments include a phase where metrics are gathered from and about various kinds of resources. This phase is often done via manual, non-reproducible and error-prone steps. We designed an experiment monitoring framework called MonEx, built on top of infrastructure monitoring solutions and supporting various monitoring approaches. MonEx fully integrates into the experiment workflow by encompassing all steps from data acquisition to producing publishable figures [18], [29].

7.2.5. Testbed federation and collaborations in the testbeds community

Participant: Lucas Nussbaum [contact].

The Fed4FIRE+ H2020 project started in January 2017 and will run until the end of September 2021. This project aims at consolidating the federation of testbeds in Europe of which Grid'5000 is a member. In 2018, we focused on various aspects related to experiment reproducibility.

We are also active in the GEFI initiative that aims at building links between the US testbeds community (GENI) and their European (FIRE), Japanese and Brazilian counterparts. We participated in the annual GEFI meeting where we chaired two sessions on *Experiment reproducibility* and *Networking experiments*, respectively, and gave one talk on Experiment data management, outlining the recent work that was done on Grid'5000 on disk reservation [27].

7.2.6. Blockchain experimentation

Participants: Jérôme François [Contact], Wazen Shbair [University of Luxembourg, Luxembourg], Radu State [University of Luxembourg, Luxembourg], Mathis Steichen [University of Luxembourg, Luxembourg].

The experimentation of distributed applications like blockchains needs a highly reconfigurable and controllable environment for fine-tuning blockchain and network parameters in different scenarios. Therefore, there might be significant manual operations which lead to human errors and make it hard to reproduce experiments. We proposed an easy to use orchestration framework over the Grid'5000 platform [23]. Our tool can fine-tune blockchain and network parameters before and between experiments. The proposed framework offers insights for private and consortium blockchain developers to identify performance bottlenecks and to assess the behavior of their applications in different circumstances.

7.2.7. NDN experimentation

Participants: Thibault Cholez [Contact], Xavier Marchal, Olivier Festor.

While ICN is a promising technology, we currently lack experiments carrying real user traffic. This also highlights the difficulty of making the link between the new NDN world and the current IP world. To address this issue, we designed and implemented an HTTP/NDN gateway (composed of ingress and egress gateways) that can transport the traffic of regular web users over an NDN island. Users just need to configure the ingress gateway as a standard web proxy that will be the entry point to the virtualized NDN island, and their traffic is seamlessly transported over NDN, thus benefiting from the good properties of the protocol to deliver content (request mutualization, caching, etc.). HTTP requests/responses are converted into NDN Interest/Data and the answer can either come from the island, or from the web through the egress gateway. Our first functional experimental results of an initial testbed deployment exhibit the capability of our global infrastructure to retrieve the top-1000 most popular web sites without difficulty [17]. This opens the way to wider and more realistic experiments of NDN with real traffic. In particular, the gateway was used to perform QoE experiments involving real users from Nancy and Troyes. They accessed many websites through the NDN network in a very satisfying way.

7.3. Analytics

7.3.1. CPS Security analytics

Participants: Abdelkader Lahmadi [contact], Mingxiao Ma, Isabelle Chrisment.

During 2018, we designed and evaluated a novel type of attack, named Measurement as Reference attack (MaR), on the cooperative control and communication layers in microgrids, where the attacker targets the communication links between distributed generators (DGs) and manipulates the reference voltage data exchanged by their controllers. We analyzed the control-theoretic and detectability properties of this attack to assess its impact on reference voltage synchronization at the different control layers of a microgrid. Results from numerical simulation are presented in [15] and demonstrate this attack, in particular the maximum voltage deviation and inaccurate reference voltage synchronization it causes in the microgrid.

7.3.2. Analysis of Internet-wide attacks

Participants: Abdelkader Lahmadi [contact], Giulia de Santis, Jérôme François, Olivier Festor.

Internet-wide scanners are heavily used for malicious activities. In [13], we developed models based on HMMs (Hidden Markov Models) and finite mixture models to identify network scanners from the packets received by a darknet. We used data collected by the darknet hosted in the High Security Lab of Inria Nancy - Grand Est to build these models by characterizing the spatial and temporal movements of the studied scanners (Zmap and Shodan). Our models are able to recognize the scanner with an accuracy of 95% when using spatial movements, and of 98% when using temporal movements.

Under the umbrella of the ThreatPredict project with the International University of Rabat, we have performed preliminary exploratory analysis of Inria darknet data that consists of examining time series of scan activities and the scanning behavior of different attackers [24]. We performed experiments on the clustering of darknet data to extract threat patterns including scanning and DDoS activities. We are still extending the technique with more features and developing Hololens based visualization of the obtained graphs. Based on our experience, traffic analysis faces a major challenge when using machine learning or data-mining techniques due to data which cannot be represented in a meaningful metric space. One major case is TCP or UDP ports. We thus proposed a new semantic based metric between port numbers that does not follow a regular numeric distance but relies on observed attacks of the past.

7.3.3. Cyber Threat Intelligence

Participants: Jérôme François, Abdelkader Lahmadi [contact], Quang Vinh Dang.

We are exploring and validating techniques for learning correlations between vulnerabilities and attack patterns from Cyber threat intelligence data sources including CVE (Common Vulnerabilities and Exposures), CAPEC (Common Attack Pattern Enumeration and Classification) and CWE (Common Weaknesses Enumeration) documents. While there already exist some relations between them, they have been defined manually and so are quite incomplete. Finding these relations is a cumbersome and tedious task and our objective is to guide or even automatically detect relations or correlations between documents. This will ease a better understanding and mitigation of threats. Our work relies on leveraging NLP (Natural Language Processing Techniques) with several techniques such as graph-based or recommendation-based mining. The first results show the ability of our technique to automatically discover missing relations between attack patterns and vulnerability descriptions in the context of SDN [12]. We also consider word and document embedding to identify correlations between them.

7.4. Orchestration

7.4.1. Programming of network functions

Participants: Thibault Cholez [contact], Diane Adjavon [Orange Labs], Anthony Anthony, Raouf Boutaba, Paul Chaignon, Shihabur Rahman Chowdhury, Olivier Festor, Jérôme François, Kahina Lazri [Orange Labs], Xavier Marchal.

NFV is a key technology for the successful deployment of new network protocol stacks like Named Data Networking (NDN). Instead of trying to oddly couple IP and new Information-Centric Networking protocols, one should rather deploy them in different network slices and ensure their isolation. We proposed a complete NFV architecture composed of several Virtual Network Functions (VNF) designed for NDN and orchestrated so that they can dynamically adapt the topology to react against issues such as an ongoing attack [25].

To push even further the possibilities of NFV, we applied the microservice architecture inherited from the software world to design atomic and flexible functions that must be combined to process NDN traffic. The proposed architecture, described in μNDN [16], includes seven orchestrated microservices. Some of them are components extracted from the monolithic and heavy-burden NDN router while others are new on-path functions that can perform specific processing on the traffic like a signature-verification module or a name-filtering module. The evaluation through two realistic scenarios proved the ability of our manager to dynamically scale-up bottleneck functions and mitigate ongoing attacks on the NDN network. We also refined our countermeasure against information leakage attacks in NDN [4].

In [8], we proposed to offload part of the processing of VNF to the programmable switches. The problem resides in guaranteeing a fair scheduling at the switch level assuming the required run-to-completion execution. We thus defined a token-based scheduling approach. In [6], we defined a new scheduler for VNFs that integrates a CPU cycle estimator and a heuristic to avoid wasting idle CPU cycles.

7.4.2. Software-defined security for clouds

Participants: Rémi Badonnel [contact], Olivier Festor, Maxime Compastié, He Ruan [Orange Labs].

We have pursued our work on a software-defined security framework for enabling the enforcement of security policies in distributed clouds. This framework aims at dynamically integrating and configuring security mechanisms for protecting cloud services that are distributed over multi-cloud and multi-tenant environments. In that context, we have described in [11] generation mechanisms for building protected cloud resources based on unikernels in an on-the-fly manner. These unikernels integrate security mechanisms at an early stage, and are characterized by highly-constrained configurations, in order to reduce the attack surface. A demonstration of this work has been showcased during the IFIP/IEEE NOMS 2018 international conference [10]. We have also investigated the exploitation of the TOSCA orchestration language to drive the generation of these unikernels. This language supports the specification of cloud services in the form of topologies and their orchestrations. The objective was to extend this language to both describe the generation of unikernel ressources, and specify different levels of security to be orchestrated. We have designed a framework to interpret this extended language, and to generate and configure protected resources according to these levels. We have evaluated the performance of generation mechanisms through extensive experiments. This generation can be performed in a proactive manner with respect to security levels, in accordance with elasticity and on-demand cloud properties.

7.4.3. Chaining of security functions

Participants: Rémi Badonnel [contact], Abdelkader Lahmadi, Stephan Merz, Nicolas Schnepf.

Software-defined networking offers new opportunities for protecting end users and their applications. It enables the elaboraboration of security chains that combines different security functions, such as firewalls, intrusion detection systems, and services for preventing data leakage. In that context, we have continued our efforts on the orchestration and verification of security chains, in collaboration with Stephan Merz from the VeriDis project-team at Inria Nancy. In particular, we have formalized and extended our approach for generating SDN policies to protect Android applications [21], [22]. We have introduced a system based on inference rules for automating the generation of such chains [20], taking into account both their networking behavior and the OS-level permissions that they request. By using first-order predicates for classifying network traffic observed in flow traces, the composition and factorization of security chains to be applied for several applications becomes straightforward. Our system infers a high-level representation of the security functions, which can be translated into an concrete implementation in the Pyretic language for programming software-defined networks. We showed that the generated chains satisfy several desirable properties such as the absence

of black holes or loops, shadowing freedom, and that they are consistent with the underlying security policy. We are currently working on optimizing and improving the parameterization of the security chains that are generated by our inference system.

SEMAGRAMME Project-Team

6. New Results

6.1. Syntax-Semantics Interface

Participants: Maxime Amblard, William Babonnaud, Philippe de Groote, Bruno Guillaume, Guy Perrier, Sylvain Pogodalla, Valentin Richard.

6.1.1. Abstract Categorial Grammars

Although Abstract Categorial Grammars have well established formal properties that make them suitable for language modeling, some missing features hinder their practical use. For instance, in order to have a compact description of grammatical properties such as number agreement between the subject and the verb of a sentence, a very common approach is to have syntactic descriptions augmented with feature value matrices. Having such a mechanism in Abstract Categorial Grammars requires a lot of attention in order to avoid impacting their computational properties (a previous approach using dependent types showed that, if too general, the problem may become intractable [64]). We have been working on theoretical approaches to this problem from different perspectives: looking for a computationally adequate type extension of the formalisms, and using the composition capabilities of the framework.

We also have been working on a unifying and general framework, provided by a categorical generalization of Abstract Categorial Grammars [50]. The goal is to get a unified approach to several semantic modeling, and to add numerical methods to the formalism.

6.1.2. Syntax-Semantics Interface as Graph Rewriting

In their book (English version: [22] and French version: [21]), Guillaume Bonfante (LORIA, Université de Lorraine), Bruno Guillaume and Guy Perrier devote two chapters to the usage of the Graph Rewriting formalism in the modeling of Syntax-Semantics Interface. Chapter 4 presents two existing semantics formalisms and shows how they can be encoded as graphs: Abstract Meaning Representation (AMR) [33] and Dependency Minimal Recursion Semantics (DMRS) [43], [42]. Chapter 5 described two Graph Rewriting Systems proposed by the authors to build semantics graphs in these two formalisms from syntactic dependencies.

6.1.3. Lexical Semantics

The lexicon model underlying Montague semantics is an enumerative model that would assign a meaning to each atomic expression. This model does not exhibit any interesting strucuture. In particular, polysemy problems are considered as homonymy phenomena: a word has as many lexical entries as it has senses, and the semantic relations that might exist between the different meanings of a same word are ignored. To overcome these problems, models of generative lexicons have been proposed in the literature. Implementing these generative models in the realm of the typed λ-calculus necessitates a calculus with notions of subtyping and type coercion. William Babonnaud is currently developing such a calculus.

6.2. Discourse Dynamics

Participants: Maxime Amblard, Timothée Bernard, Clément Beysson, Maria Boritchev, Philippe de Groote, Bruno Guillaume, Pierre Ludmann, Michel Musiol.

6.2.1. Dynamic Logic

We have revisited the type-theoretic dynamic logic introduced in [3]. We have shown how a slightly richer notion of continuation together with an appropriate notion of polarity results in a richer and more powerful framework. In particular, it allows new dynamic connectives and quantifiers to be defined in a systematic way. This work has been presented as an invited talk at the *LACompLing 2018* symposium [11].

6.2.2. Discourse Relations

A text as a whole must exhibit some coherence that makes it more than just a bag of sentences. This coherence hinges on discourse relations (DRs), that express the articulations between the different segments of the text. Typical DRs include relations of *Contrast*, *Consequence* or *Explanation*. The most direct and reliable way to express a DR is to use a discourse connective (e.g., *because*, *instead*, *for example*). These lexical items have specific syntactic, semantic and pragmatic properties, the study of which is the subject of Timothée Bernard's PhD thesis.

Some discourse connectives (typically, adverbial connectives such as *so* or *otherwise*) have only one syntactic argument. It then seems natural to use an anaphora mechanism to retrieve the other argument from the context. This proposal has been formalized in [12] by means of continuation-based type theoretic dynamic logic. In this model, the semantic arguments of a DR are considered to be abstract entities akin to Davidsonian events. This approach raises difficulties when the argument of DR is a negative sentence. Indeed, according to the standard analysis of negation in event semantics, a negative sentence does not introduce any specific event. In order to circumvent this problem, we have developped a logical theory of *negative events* [13], [17], [29].

6.2.3. Dynamic Generalized Quantifiers

Clement Beysson has continued his work on dynamic generalized quantifiers as denotations of the (French) determiners. In this context, he has studied several issues raised by the modeling of plural determiners. In particular, the opposition between distributive and collective interpretations suggests that intrinsically dynamic plural determiners should introduce plural discourse referents that stand for collection of entities. In order to formalize this notion, he has studied several theories of plurality: mereology, plural logic, and second-order logic.

6.2.4. Dialogue Modeling

Maxime Amblard and Maria Boritchev develop a dynamic approach of dialogue modelling. One of the main difference between discourse and dialogue is the interactions between the speakers. To do so, they introduce a formal approach to compositional processing of questions and answers. They address dialogue lexicality issues starting from the formal definitions of so-called Düsseldorf Frame Semantics given in [51]. They introduce a view of dialogues as compositions of negotiation phases that can be studied separately one from another while linked by a common dialogue context (accessible to all participants of a dialogue). They apply Inquisitive Semantics [39] in that context.

Maxime Amblard and Maria Boritchev works on the categorisation of questions and answers and apply some machine learning approches for automatic classification. They present the architecture of the model, especially how to handle these phenomena with logical representations in [14]. Their view is to narrow the problem of identifying incomprehension in dialogue to the one of finding logical incoherences in speech act combinations as the one we found in the SLAM project (ongoing project of the Sémagramme team on interviews with schizophrenics). They also start to build a new corpus - DinG (Discourse in Dialogue) - based on record and transcript plays to the settlers of Catan board game.

Maxime Amblard also started a cooperation with CLASP, especially with Robin Cooper, Ellen Breitholtz and Chris Howes. They work on the synchronisation of the representation of dialogue modelling with the previous proposals and Type-Theoretic-Records (TTR) [41]. They apply the solution on extracts from two corpora where patients with schizophrenia are involved.

6.2.5. Pathological Discourse Modelling

Michel Musiol obtained a part-time delegation in the Semagramme team. This proximity makes possible to set up a more active dialogue on the issue of pathological discourse modeling. He has worked on the development of the possibility of testing his conjectures on the cognitive and psychopathological profile of the interlocutors, in addition to information provided by the model of ruptures and incongruities in pathological discourse. This methodological system makes it possible to discuss, or even evaluate, the heuristic potential of the computational models developed on the basis of empirical facts.

Moreover, the diagnostic tools used today by the professional community (clinical and psychiatric) are of limited expertise for the effective identification of the signs of the pathology for at least two reasons: on the one hand, they are much too imprecise on the side of the recognition of Language Impairment and Thought Disorder (no underlying linguistic and psycholinguistic theories); on the other hand, they do not take into account (either theoretically or technically) the discursive structure within which these disorders are expressed. The objective of this research program is therefore also to anticipate the development of diagnostic tools for the psychiatric and psychological community.

As part of the work carried out in the SLAM project, Maxime Amblard, Michel Musiol and Manuel Rebuschi (Archives Henri-Poincaré, Université de Lorraine) continue to work on modelling interactions with schizophrenic patients. The project has progressed on three different operational levels: building new resources, editing a volume (Springer) on the SLAM project in 2019 and improving the representation model.

An agreement is being deployed with the psychiatric hospital of Aix-en-Provence. The on-site staff administered a test protocol to the entire test group of 60 people. Transcripts are in progress, which will provide a significant amount of data to work on for the project. Thanks to the involvement of a medical staff, the recovery of new data appears well advanced. In the same perspective, contacts are being made with the Psychotherapeutic Centre in Nancy.

In addition, Maxime Amblard carried out a one-week international mobility at CLASP thanks to a mobility grant from the French Embassy in Sweden. Discussions were initiated with these colleagues for the development of projects using formal semantic models for the analysis of interaction with schizophrenic patients.

6.3. Common Basic Resources

Participants: Maxime Amblard, Clément Beysson, Philippe de Groote, Bruno Guillaume, Maxime Guillaume, Guy Perrier, Sylvain Pogodalla, Nicolas Lefebvre.

6.3.1. Application of Graph Rewriting to Natural Language Processing

Guillaume Bonfante, Bruno Guillaume and Guy Perrier collected their work on the application of graph rewriting to Natural Language Processing (NLP) in a book written in French [21] and translated to English [22] by the editor. This book shows how graph rewriting can be used as a computational model adapted to NLP. Currently, there is no standard model for graph rewriting and, as such, the authors have conceived one that is specifically adapted to NLP, proposing their own implementation: the GREW system. In addition to the application to Syntax-Semantic Interface mentioned above, the book presents applications in syntactic parsing and in syntactic corpus conversion.

In [5], Guillaume Bonfante and Bruno Guillaume describe some mathematical properties of the Graph Rewriting framework used in GREW. The previous experiments on NLP tasks have shown that Graph Rewriting applications to Natural Language Processing do not require the full computational power of the general Graph Rewriting setting. The most important observation is that all graph vertices in the final structures are in some sense "predictable" from the input data and so, it is possible to consider the framework of Non-size increasing Graph Rewriting. The paper concerns the theoretical aspect of termination with respect to this calculus. It is shown that uniform termination is undecidable and that non-uniform termination is decidable. We define termination techniques based on weight, we prove the termination of weighted rewriting systems and we give complexity bounds on derivation lengths for these rewriting systems.

6.3.2. Building Linguistics Resources with Crowdsourcing

In the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions, Karën Fort (Sorbonne Université), Bruno Guillaume, Matthieu Constant (ATILF, Nancy), Nicolas Lefebvre and Yann-Alan Pilatte (Sorbonne Université) presented the results obtained in crowdsourcing French speakers' intuition concerning multi-word expressions (MWEs) [15]. They developed a slightly gamified crowdsourcing platform, part of which is designed to test users' ability to identify MWEs with no prior training. The participants perform relatively well at the task, with a recall reaching 65% for MWEs that do not behave as function words.

6.3.3. Corpus Annotation

Kim Gerdes (Sorbonne nouvelle, Paris 3), Bruno Guillaume, Sylvain Kahane (Université Paris Nanterre) and Guy Perrier proposed a surface-syntactic annotation scheme called Surface Universal Dependencies (SUD) that is near-isomorphic to the Universal Dependencies (UD) annotation scheme. The SUD scheme follows distributional criteria for defining the dependency tree structure and the naming of the syntactic functions [16]. Rule-based graph transformation grammars allow for a bi-directional transformation of UD into SUD. The back-and-forth transformation can serve as an error-mining tool to assure the intra-language and inter-language coherence of the UD treebanks. The UD corpora are available on gitlab.inria.fr.

Bruno Guillaume and Guy Perrier used the GREW system for the development of the French part of the Universal Dependencies project (UD) [32]. They focused in particular on correcting the annotation of two French corpora, *UD_French-GSD* and *UD_French-Sequoia*. For the correction, they first used the tool Grew-match (based on the pattern matching part of GREW) to detect error patterns, but also the GREW rewriting rule system to transform the annotation from one format to another one [19]. Version 2.3 of the UD corpora was released on 15 November 2018.

6.3.4. FR-Fracas

Maxime Amblard, Clement Beysson, Philippe de Groote, Bruno Guillaume and Sylvain Pogodalla continue their work on the FR-Fracas project. There are two major levels of processing that are significant in the use of a computational semantics framework: semantic composition, for the construction of meanings, and inference, either to exploit those meanings, or to assist the determination of contextually sensitive aspects of meanings. FraCas is an inference test suite for evaluating the inferential competence of different NLP systems and semantic theories. Providing an implementation of the inference level was beyond the scope of FraCaS, but the test suite nevertheless provides an overview of a useful and theory- and system-independent semantic tool [40].

There currently exists a multilingual version of the resource for Farsi, German, Greek, and Mandarin. Sémagramme completed the translation into French of the test suite. All translations were subject to a bidding phase by two project members. Then the cases that were identified as difficult were discussed by all project members. An adjudication step finally ensured the quality of the translation. In order to evaluate the inference mechanism triggered by the translated sentences, a web interface is being developed.

6.3.5. Large Coverage Abstract Categorial Grammars

Maxime Amblard, Maxime Guillaume, and Sylvain Pogodalla have worked on the automatic translation of large coverage Tree-Adjoining grammars into Abstract Categorial Grammars. On the theoretical side, this work hinges on the encoding proposed by Philippe de Groote and Sylvain Pogodalla [69], [63]. On the implementation side, the starting point are TAG grammars generated from meta-grammars by XMG [44], [61]. This generates Abstract Categorial grammars containing about 23 000 entries, and was used as a test bed for the ACGtk toolkit, some parts of which have been rewritten to scale up.

SPHINX Project-Team

6. New Results

6.1. Inverse problems for heterogeneous systems

Participants: David Dos Santos Ferreira, Karim Ramdani, Julie Valein, Alexandre Munnier, Jean-Claude Vivalda.

- In [32], we deal with a problem of observability for waves propagating in two environments with different speeds of propagation. We give an explicit construction of the regions of observability in the two-dimensional case. This allows us to determine in which locations we have to make some measurements in order to obtain the solution within the domain.
- In [33], we deal with the observability of the 1-D wave equation. The semi discretization of the waves problem leads to some uniform observability problems. This is due to the bad approximation of the high frequencies of discrete solutions. Some remedies are known, which involve finite element methods. In this paper, we give three methods allowing to retrieve the uniform observability when the approximations are made with a Galerkin method.
- In [15], Ramdani *et al.* proposed an algorithm for estimating from partial measurements the population for a linear age-structured population diffusion model. In this work, the physical parameters of the model were assumed to be known. The authors investigate the inverse problem of simultaneously estimating the population and the spatial diffusion coefficient for an age-structured population model. The measurement used is the time evolution of the population on a subdomain in space and age. The proposed method is based on the generalization to the infinite dimensional setting of an adaptive observer originally proposed for finite dimensional systems.
- In [13], Munnier and Ramdani proposed an explicit reconstruction formula for a two-dimensional cavity inverse problem. The proposed method was limited to the case of a single cavity due to the use of conformal mappings. In [13], Munnier and Ramdani consider the case of a finite number of cavities and aim to recover the location and the shape of the cavities from the knowledge of the Dirichlet-to-Neumann (DtN) map of the problem. The proposed reconstruction method is non iterative and uses two main ingredients. First, the authors show how to compute so-called generalized Pólia-Szegö tensors (GPST) of the cavities from the DtN of the cavities. Secondly, the authors shows that the obtained shape from GPST inverse problem can be transformed into a shape from moments problem, for some particular configurations. However, numerical results suggest that the reconstruction method is efficient for arbitrary geometries.
- In [2], we show that, generically, a (finite dimensional) sampled system is observable provided that the number of outputs is at least equal to the number of inputs plus 2. This work complements some previous works on the subject.
- In [18], we design a state observer for a coupled two dimensional partial differential equations (PDEs) system used to describe the heat transfer in a membrane distillation system for water desalination.

  In [23], we deal with uniqueness and stability issues for the inverse spectral problem of recovering the magnetic field and the electric potential in a Riemannian manifold from some asymptotic knowledge of the boundary spectral data of the corresponding Schrödinger operator under Dirichlet boundary conditions.

6.2. Control and stabilization of heterogeneous systems

Participants: Thomas Chambrion, David Dos Santos Ferreira, Takéo Takahashi, Julie Valein.

- In [8], we find, thanks to a a semiclassical approach, L^p estimates for the resolvants of the damped wave operator given on compact manifolds whose dimension is greater than 2.
- In [27], we have proved a "Ball-Marsedn-Slemrod" obstruction to the bi-linear controllability of the Klein-Gordon equation. With different methods, we obtained comparable results for the Gross-Pitaevskii equation in [28].
- In [7], we study the local exponential stability of the nonlinear Korteweg-de Vries equation with boundary time-delay feedback by using two different methods: a Lyapunov functional approach (with an estimation on the decay rate, but with a restrictive assumption on the length of the spatial domain) and an observability inequality approach (for any non critical lengths).
- In [12], we study the local controllability to trajectories of a Burgers equation with nonlocal viscosity. By linearization we are led to an equation with a non local term whose controllability properties are analyzed by using Fourier decomposition and biorthogonal techniques. Once the existence of controls is proved and the dependence of their norms with respect to the time is established for the linearized model, a fixed point method allows us to deduce the result for the nonlinear initial problem.
- In [26], we establish a Lebeau-Robbiano spectral inequality for a degenerated one dimensional elliptic operator and show how it can be used to impulse control and finite time stabilization for a degenerated parabolic equation.
- In [25], We prove a Carleman estimate in a neighborhood of a multi-interface, under compatibility assumptions between the Carleman weight, the operators at the multi-interface, and the elliptic operators in the interior and the usual sub-ellipticity condition. We derive some properties of unique prolongation, control of the heat equation, and satblization of the related damped waves equation.

6.3. Numerical analysis and simulation of heterogeneous systems

Participant: Xavier Antoine.

- In [10], we design some accurate artificial boundary conditions for the semi-discretized linear Schrödinger and heat equations in rectangular domains. We show the accuracy of the method thanks to simulations
- In [5], we design fast numerical and highly accurate methods for the computation of steady states and the dynamics of time or space-fractional Schrödinger equations.
- In [1], we design a numerical model of diffusion for the study of the properties of noble gases originating from volcanic eruptions.
- In [4], the deal with a multilevel Schwarz Waveform Relaxation (SWR) Domain Decomposition Method (DDM) for the Non Linear Schrödinger Equation (NLSE).
- In [6], we design a fast and pseudo spectral preconditioned conjugated gradient method for the computation of the steady states related to the Gross-Pitaevskii equation with non local dipolar interaction.
- In [3], we deal with fractional microlocal analysis for the obtention of asymptotic estimates for the convergence of Schwarz Waveform Relaxation (SWR) domain decomposition method; this study is done is the two dimensional quantum case.
- In [11], we design new methods of very high order for the computation of diffracted fields; these methods rely on a B-splines finite element method and are related to the isogeometric analysis.
- In [17], we deal with the numerical analysis of fast and accurate schemes for solving one-dimensional time-fractional nonlinear Schrödinger equations set with artificial boundaries.

- In [35], we obtain a close approximation of the optimal parameters for the convergence of domain decomposition methods for the Schrödinger equation.
- In [19], we compute an explicit approximation of the optimal parameters for the convergence of domain decomposition methods for the Schrödinger equation.
- In [21], we introduce an original method in order to integrate PML in a pseudospectral method for the computation of the dynamics of the Dirac equation. Some applications to lasers are given.
- In [20], we deal with the asymptotic analysis of the rate of convergence of the classical and quasi-optimal Schwarz waveform relaxation (SWR) method for solving the linear Schrödinger equation.

6.4. Fluid-Structure Interaction

Participants: Julien Lequeurre, Jean-François Scheid.

In [16], we deal with shape optimization problem for a Stokes/elasticity system. The aim is to find the optimal shape of an elastic structure which minimizes an energy type functional. Results are obtained for a simplified free-boundary one-dimensional problem.

In [34], we design a hilbertian framework for the analysis of the planar Navier-Stokes (NS) equations either in vorticity or in stream function formulation. The fluid is assumed to occupy a bounded possibly multiply connected domain. The velocity field satisfies either homogeneous (no-slip boundary conditions) or prescribed Dirichlet boundary conditions. We prove that the proposed approach is equivalent to the classical one (stated in primitive variables, i.e. velocity and pressure fields) for strong and weak solutions. In particular . In particular, in both cases, we retrieve the pressure from the vorticity or the current function.

TONUS Team

7. New Results

7.1. Palindromic BGK methods

Since two years we work on implicit relaxation methods to solve hyperbolic PDE without CFL and without matrices to invert. The Palindromic BGK method allows to approximate a hyperbolic system by a larger set of transport equations coupled by a nonlinear source term which relaxes the variables on an equilibrium. Using a splitting scheme, we can solve these transport equations in parallel and solve the local relaxation in a second step. The high-order extension is obtained by a symmetric modified Strang splitting and composition methods.

7.1.1. Boundary condition for Palindromic BGK method

Participants: Florence Drui, Emmanuel Franck, Philippe Helluy, Laurent Navoret.
One of the drawbacks to the Palindromic BGK model is the treatment of the boundary conditions. Indeed the BGK scheme admits more variables than the original one and the boundary conditions for these additional variables are not defined. The classical choice is to impose the equilibrium at the boundary. In this case we obtain instabilities and only the first order convergence. After an analysis of the symmetric modified Strang splitting method, we have identified the dynamic for the non-physical variables and proposed boundary conditions compatible with this dynamic. We obtain stable and second order boundary conditions.

7.1.2. Palindromic BGK scheme for Low-Mach models

Participants: Clémentine Courtès (IRMA), Emmanuel Franck, Philippe Helluy, Laurent Navoret.
Another drawback of the method is the application for "two-scale" problems like Low-Mach flows. Indeed, in this case the BGK representation used generate an large error on the slow scale which is homogeneous to the fast scale. Consequently the slow scale is not well resolved. This problem comes from the fact that the BGK approximation uses a linearization with a constant fast scale to approximate all the systems. We have proposed a new method where we also introduce a slow scale in the BGK approximation. Using this, we obtain accurate results for the Euler equation in the low-Mach regime in 1D. The method gives interesting results also for other applications. In the future we must extend the method in 2D.

7.1.3. Palindromic BGK scheme for diffusion models

Participants: Laura Mendoza, Emmanuel Franck, Laurent Navoret.
In MHD simulations for ITER, we must also discretize with an implicit scheme the anisotropic diffusion. Firstly, we have proposed to extend the previous Palindromic BGK method to the parabolic problems. For that we must use a different Palindromic BGK model with specific parameters. We obtain a second order scheme without CFL for the Heat equation in 1D and 2D. In the future we will consider the high-order schemes and the extension to the anisotropic case.

7.1.4. Semi-Lagrangian on complex geometries for Palindromic BGK scheme

Participants: Laura Mendoza, Emmanuel Franck, Philippe Helluy.
To apply the Palindromic BGK method we must have an advection solver without CFL. In the code Slappy we propose a 3D high-order Semi-Lagrangian solver able to treat blocks-structured meshes with overlapping and non-conformity. This allows to treat complex geometries easily. The solver is written in PYOpenCL and can be used on GPU. In the code the relaxation step is also implemented, which allows to use the Palindromic BGK method on some PDE (Euler, Diffusion etc).

7.1.5. Lattice Boltzmann scheme with PyOpenCL

Participants: Florence Drui, Emmanuel Franck, Philippe Helluy.

In the same idea, another code has been developed to treat hyperbolic systems with the BGK approach. In this case the transport is exact and consequently the method is equivalent to the Lattice Boltzmann scheme. The parallel part is similar and also based on PyOpenCL. This version is less accurate than the previous code, can be used only in Cartesian grids but is more stable and can run more complex problems. The main result is the simulation of 2D resistive MHD instabilities which have the same structure than Tokamak instabilities.

7.2. Numerical methods for Euler/MHD models

7.2.1. Splitting scheme in JOREK code

Participants: Emmanuel Franck
The Jorek code is the main European code for the simulation of Tokamak instabilities. The inversion of the full matrix is based on a Block Jacobi preconditioning which is not efficient in some cases and very greedy in memory. To solve the problem we investigate on splitting scheme which will allow to solve some simple subsystems separately. The splitting scheme have been tested on the first MHD model on JOREK in the quasi-linear case. In this regime the splitting gives good results since the accuracy is close to the original full implicit solver. The nonlinear case is currently studied.

7.2.2. Compatible finite element for MHD

Participants: Emmanuel Franck, Eric Sonnendruecker (IPP), Mustaga Gaja (iPP)
The works on the compatible finite elements for MHD is continued. This method allows to preserve the energy balance or the divergence free constrains with high-order finite element on complex geometries. This method is coupled with a splitting between the different physical parts and a nonlinear solver. The method gives expected results for Maxwell and Acoustic and also gives good results for the nonlinear acoustic part of the MHD model. The magnetic and convective parts of the MHD model are currently studied.

7.2.3. Semi implicit for relaxation model in low-Mach regime

Participants: Emmanuel Franck, Laurent Navoret.
To apply the previous method, we must invert a nonlinear problem. A parabolization method allows to reduce the dimension of the implicit problem. However the problem is still nonlinear and ill-conditioned for strong gradient of the physical quantities. To avoid this, we propose a new relaxation method for the Euler equations (to begin) which allows to linearize the acoustic part preserving the low-Mach limit (which is the relevant regime for our application). This relaxation method allows to obtain a well-conditioned and linear implicit part. The method is validated in 1D/2D in a finite volumes context and will be extended to the high-order scheme and MHD model.

7.3. Eulerian method for Vlasov equation

7.3.1. Recurrence phenomenon for finite element grid based Vlasov solver

Participants: Michel Mehrenberger, Laurent Navoret, Nhung Pham (IRMA)

In this work, we focus on one difficulty arising in the numerical simulation of the Vlasov-Poisson system: when using a regular grid-based solver with periodic boundary conditions, perturbations present at the initial time artificially reappear at a later time. For regular finite-element mesh in velocity, we show that this recurrence time is actually linked to the spectral accuracy of the velocity quadrature when computing the charge density. In particular, choosing trigonometric quadrature weights optimally defers in time the occurrence of the recurrence phenomenon. Numerical results using both the Semi-Lagrangian Discontinuous Galerkin and the Finite Element / Semi-Lagrangian methods have been carried out and confirm the analysis.

7.3.2. Numerical scheme for sheath equilibria

Participants: Mehdi Badsi (Nantes University), Michel Mehrenberger, Laurent Navoret

We are interested in developing a numerical method for capturing stationary sheaths that a plasma forms in contact with a metallic wall. This work is based on a bi-species (ion/electron) Vlasov-Ampère model proposed in [18]. The main question addressed in this work is to know if classical numerical schemes can preserve stationary solutions with boundary conditions, since these solutions are not a priori conserved at the discrete level. In the context of high-order semi-Lagrangian method, due to their large stencil, interpolation near the boundary of the domain also requires a specific treatment. As expected, we numerically observe that the preservation of the equilibria is very sensitive to the prescribed boundary conditions and high order schemes are mandatory to maintain the preservation of the energy in large times.

7.3.3. Realistic geometry for Gysela

Participants: N. Bouzat, C. Bressan, V. Grandgirard, G. Latu, M. Mehrenberger

In magnetically confined plasmas used in Tokamak, turbulence is responsible for specific transport that limits the performance of this kind of reactors. Gyrokinetic simulations are able to capture ion and electron turbulence that give rise to heat losses, but also require state-of-the-art HPC techniques to handle computation costs. Such simulations are a major tool to establish good operating regime in Tokamak such as ITER, which is currently being built. Some of the key issues to address more realistic gyrokinetic simulations are: efficient and robust numerical schemes, accurate geometric description, good parallelization algorithms. The framework of this work is the Semi-Lagrangian setting for solving the gyrokinetic Vlasov equation and the Gysela code. In this paper, a new variant for the interpolation method is proposed that can handle the mesh singularity in the poloidal plane at r = 0 (polar system is used for the moment in Gysela). A non-uniform meshing of the poloidal plane is proposed instead of uniform one in order to save memory and computations. The interpolation method, the gyroaverage operator, and the Poisson solver are revised in order to cope with non-uniform meshes. A mapping that establishes a bijection from polar coordinates to more realistic plasma shape is used to improve realism. Convergence studies are provided to establish the validity and robustness of our new approach.

7.3.4. Parallel computing for kinetic solvers

Participants: Ksander Ejjaaouani, Olivier Aumage, Julien Bigot, Michel Mehrenberger

Existing programming models tend to tightly interleave algorithms and optimizations in HPC simulation codes. This requires scientists to become experts in both the simulated domain and the optimization process and makes the code difficult to maintain and port to new architectures. This paper proposes the InKS programming model that decouples these two concerns with distinct languages for each. The simulation algorithm is expressed in the InKS pia language with no concern for machine-specific optimizations. Optimizations are expressed using both a family of dedicated optimizations DSLs (InKS O) and plain C++. InKS O relies on the InKS pia source to assist developers with common optimizations while C++ is used for less common ones. Our evaluation demonstrates the soundness of the approach by using it on synthetic benchmarks and the Vlasov-Poisson equation. It shows that InKS offers separation of concerns at no performance cost.

7.4. PIC method for Vlasov equation

7.4.1. Parallel computing for PIC method

Participants: Y. Barsamian, A. Chargueraud, S. Hirstoaga, M. Mehrenberger

Particle-in-Cell (PIC) codes are widely used for plasma simulations. On recent multi-core hardware, performance of these codes is often limited by memory bandwidth. We describe a multi-core PIC algorithm that achieves close-to-minimal number of memory transfers with the main memory, while at the same time exploiting SIMD instructions for numerical computations and exhibiting a high degree of OpenMP-level parallelism [6]. Our algorithm keeps particles sorted by cell at every time step, and represents particles from a same cell using a linked list of fixed-capacity arrays, called chunks. Chunks support either sequential or atomic insertions, the latter being used to handle fast-moving particles. To validate our code, called Pic-Vert, we consider a 3d electrostatic Landau-damping simulation as well as a 2d3v transverse instability of magnetized electron holes. Performance results on a 24-core Intel Sky-lake hardware confirm the effectiveness of our algorithm, in particular its high throughput and its ability to cope with fast moving particles.

7.4.2. Two species Vlasov solver

Participants: Y. Barsamian, J. Bernier, S. Hirstoaga, M. Mehrenberger

Thanks to a classical first order dispersion analysis, we are able to check the validity of 1Dx1D two-species Vlasov-Poisson simulations. The extension to second order is performed and shown to be relevant for explaining further details. In order to validate multidimensional effects, we propose a 2Dx2D single species test problem that has true 2D effects coming from the sole second order dispersion analysis. Finally, we perform, in the same code, full 2Dx2D nonlinear two-species simulations with mass ratio around 0.01, and consider the mixing of semi-Lagrangian and Particle-in-Cell methods.

7.5. Other works

7.5.1. Tomography

Participants: Laura Mendoza Virtually all magnetic fusion devices resort to tomography diagnostics for a variety of plasma emissions. All those diagnosis have a lot in common: the plasma is transparent to the observed quantity, such that the signal on a detector is derived from a spatial integration of the local emission. Solving the direct problem (i.e. from simulated emissivity to signals) requires modeling the diagnostic geometry and is used for physics code validation or diagnostic design. Solving the inverse problem (i.e. from experimental signals to reconstruct 2D emissivity) is useful for data interpretation and requires not only geometry modeling but also decomposing the unknown emissivity into basis functions and inversion-regularization routines. In this context, a python library, ToFu, solves the direct and inverse problems for synthetic diagnostics. The project objective for the second part of 2018 is to develop and optimize the existing geometry module in ToFu, with a special focus on the ray-tracing algorithms.

7.5.2. Discontinuous Galerkin solver

Participants: Philippe Helluy, Bruno Weber We have implemented and validated new optimizations in our Discontinuous Galerkin (DG) codes CLAC and SCHNAPS. In CLAC, Bruno Weber, our CIFRE PhD in the AxesSim company, has implemented a local time-step method and optimizations in order to run efficiently the OpenCL kernels both on CPU and GPU. This allows to run a huge electromagnetic simulation of a Bluetooth antenna in interaction with a full volumic human model. The simulation was run on the supercomputer Piz Daint (3rd at the "top 500" ranking in 2017). The computing hours were awarded through a PRACE call dedicated to small companies. In SCHNAPS we were able to assess the efficiency of the StarPU runtime for distributing the computational tasks efficiently on hybrid computers.

7.5.3. Finite volume methods for complex hyperbolic system

Participants: Philippe Helluy, Lucie Quibel In the thesis of Lucie Quibel (started in November 2017), we study numerical methods for solving compressible fluids with complex equation of states. The objective is to simulate liquid-vapor flows that occur in nuclear plants. The pressure behavior of the liquid-vapor mixture is very complex and obtained through measurements and tabulated laws. This sometimes prevent the system from being hyperbolic and leads to instabilities. We are trying to construct simpler but realistic laws that preserve the convexity structure and the scheme robustness.

TOSCA Project-Team

5. New Results

5.1. Probabilistic numerical methods, stochastic modelling and applications

Participants: Mireille Bossy, Nicolas Champagnat, Quentin Cormier, Madalina Deaconu, Olivier Faugeras, Coralie Fritsch, Pascal Helson, Antoine Lejay, Radu Maftei, Victor Martin Lac, Hector Olivero-Quinteros, Émilie Soret, Denis Talay, Etienne Tanré, Milica Tomasevic, Denis Villemonais.

5.1.1. Published works and preprints

- M. Bossy, J. Fontbona (Universidad de Chile, Chile) and H. Olivero-Quinteros (CIMFAV, Valparaíso, Chile) analysed mathematical model for the collective behavior of a fully connected network of finitely many neurons. They obtained that the whole system synchronize, up to some error controlled by the channels noise level. The associated nonlinear McKean-Vlasov equation concentrates, as time goes to infinity, around the dynamics of a single Hodgkin-Huxley neuron with a chemical neurotransmitter channel [42].
- M. Bossy, A. Dupré, P. Drobinski, L. Violeau and C. Briard (Zephyr ENR) obtained advances in stochastic Lagrangian approach for atmospheric boundary layer simulation, on the analysis of an optimal rate of convergence for the particle approximation method, and on validation case with the simulation of a Zephyr ENR wind farm site of six turbines [36].
- M. Di Iorio (Marine Energy Research and Innovation Center, Santiago, Chile), M. Bossy, C. Mokrani (Marine Energy Research and Innovation Center, Santiago, Chile), and A. Rousseau obtained advances in stochastic Lagrangian approaches for the simulation of hydrokinetic turbines immersed in complex topography [33], [50].
- Together with M. Andrade-Restrepo (Univ. Paris Diderot) and R. Ferrière (Univ. Arizona and École Normale Supérieure), N. Champagnat studied deterministic and stochastic spatial eco-evolutionary dynamics along environmental gradients. This work focuses on numerical and analytical analysis of the clustering phenomenon in the population, and on the patterns of invasion fronts [40].
- N. Champagnat and J. Claisse (Ecole Polytechnique) studied 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 quasi-stationary distribution of the processes controled by Markov controls [16].
- N. Champagnat and B. Henry (Univ. Lille 1) studied a probabilistic approach for the Hamilton-Jacobi limit of non-local reaction-diffusion models of adaptive dynamics when mutations are small. They used a Feynman-Kac interpretation of the partial differential equation and large deviation estimates to obtain a variational characterization of the limit. They also studied in detail the case of finite phenotype space with exponentially rare mutations, where they were able to obtain uniqueness of the limit [17].
- N. Champagnat and D. Villemonais solved a general conjecture on the Fleming-Viot particle systems approximating quasi-stationary distributions (QSD): in cases where several quasi-stationary distributions exist, it is expected that the stationary distribution of the Fleming-Viot processes approaches a particular QSD, called minimal QSD. They proved that this holds true for general absorbed Markov processes with soft obstacles [48].
- N. Champagnat, K. Coulibaly-Pasquier (Univ. Lorraine) and D. Villemonais obtained general criteria for existence, uniqueness and exponential convergence in total variation to QSD for multi-dimensional diffusions in a domain absorbed at its boundary [37]. These results improve and simplify the existing results and methods.

- N. Champagnat and D. Villemonais obtained contraction properties in total variation of general penalized processes, including time-inhomogeneous Markov processes with absorption and Markov processes in varying environments [20]. Their method allows to improve significantly the former results of [62], [63].
- N. Champagnat and D. Villemonais studied with R. Schott (Univ. Lorraine) models of deadlocks in distributed systems. They use the approach developped recently by the first two authors to study quasi-stationary distributions in order to characterize and compute numerically the asymptotic behaviour of the deadlock time and the behaviour of the system before deadlock, both for discrete and for diffusion models [47].
- A. Lejay and A. Brault have followed their work on rough flow, which provides an unified framework to deal with the theory of rough paths from the points of view of flows. In particular, they have shown existence of flows even when the associated rough differential equations have multiple solutions [44], [45].
- A. Lejay and P. Pigato have provided an estimator of the diffusion and drift coefficients when they are discontinuous at a threshold. These estimators have been applied to financial data and exhibit leverage as well as mean-reversion effects on S&P 500 stocks' prices [57], [30]
- A. Lejay, L. Lenôtre and G. Pichot have proposed a new Monte Carlo method based on random exponential time steps to deal with discontinuous diffusions coefficients and drift [35], [56]
- A. Lejay, S. Haraketi and E. Haoula have shown how to construct a diffusion on the Sierpinski gasket lifted to the Heisenberg group [53].
- J. Bion-Nadal (Ecole Polytechnique) and D. Talay have pursued their work on a Wasserstein-type distance on the set of the probability distributions of strong solutions to stochastic differential equations. This new distance is defined by restricting the set of possible coupling measures and can be expressed in terms of the solution to a stochastic control problem, which allows one to deduce a priori estimates or to obtain numerical evaluations: cf. [41]. This solution is now shown to exist and be smooth even in cases where the infinitesimal generators of the considered diffusion processes are not strongly elliptic.

  A notable application concerns the following modeling issue: given an exact diffusion model, how to select a simplified diffusion model within a class of admissible models under the constraint that the probability distribution of the exact model is preserved as much as possible? The objective being to select a model minimizing the above distance to a target model, the construction and analysis of an efficient stochastic algorithm are being in progress.
- In [60] D. Talay and M. Tomasevic have developed and analysed a new type of stochastic interpretation of the one-dimensional parabolic-parabolic Keller-Segel systems. It involves an original type of McKean-Vlasov interaction kernel. At the particle level, each particle interacts with all the past of each other particle. At the mean-field level studied here, the McKean-Vlasov limit process interacts with all the past time marginals of its probability distribution. In [12] M. Tomasevic has proven that the two-dimensional parabolic-parabolic Keller-Segel system in the whole Euclidean space and the corresponding McKean-Vlasov stochastic differential equation are well-posed under some explicit conditions on the parameters of the model.
- D. Talay and M. Tomasevic are studying the well-posedness and the propagation of chaos of the particle system related to the two-dimensional parabolic-parabolic Keller-Segel system. The singularity of the interaction kernel being more critical than in the one-dimensional case, the preceding analysis [26] cannot be extended and a fully new methodology needs to be developed.
- V. Martin Lac, D. Talay and M. Tomasevic have worked on theoretical and algorithmic questions related to the simulation of the Keller–Segel particle systems. A preliminary version of a library has been developed.
- H. Olivero (Inria, now University of Valparaiso, Chile) and D. Talay have constructed and analysed an hypothesis test which helps to detect when the probability distribution of complex stochastic

simulations has an heavy tail and thus possibly an infinite variance. This issue is notably important when simulating particle systems with complex and singular McKean-Vlasov interaction kernels whick make it extremely difficult to get a priori estimates on the probability laws of the mean-field limit, the related particle system, and their numerical approximations. In such situations the standard limit theorems do not lead to effective tests. In the simple case of independent and identically distributed sequences the procedure developed this year and its convergence analysis are based on deep tools coming from the statistics of semimartingales.

- V. Martin Lac, H. Olivero-Quinteros and D. Talay have worked on theoretical and algorithmic questions related to the simulation of large particle systems under singular interactions and to critical numerical issues related to the simulation of independent random variables with heavy tails. A preliminary version of a library has been developed.
- C. Graham (École Polytechnique) and D. Talay are ending and polishing the second volume of their series on Mathematical Foundation of Stochastic Simulation to be published by Springer.
- P-E. Jabin (University of Maryland) and D. Talay have ended their work on a mean-field game and shown the convergence of the joint density function of the controlled particle system. The construction of the limit has required the construction of suitable Sobolev spaces on sets of probability measures on Polish spaces.
- E. Tanré and Pierre Guiraud (Univ. of Valparaíso) have worked on the synchronization in a model of network of noisy biological neurons. Using a large deviation principle, they prove the stability of the synchronized state under stochastic perturbations. They also give a lower bound on the probability of synchronization for networks which are not initially synchronized. This bound shows the robustness of the emergence of synchronization in presence of small stochastic perturbations [25].
- E. Tanré, P. Grazieschi (Univ. Warwick), M. Leocata (Univ. Pisa), C. Mascart (Univ. Côte d'Azur), J. Chevallier (Univ. of Grenoble) and F. Delarue (Univ. Côte d'Azur) have extended the previous work [9] to sparse networks of interacting neurons. They have obtained a precise description of the limit behavior of the mean field limit according to the probability of (random) interactions between two individual LIF neurons [52].
- E. Tanré has worked with Nicolas Fournier (Sorbonne Université) and Romain Veltz (MATHNEURO Inria team) on a network of spiking networks with propagation of spikes along the dendrites. Consider a large number n of neurons randomly connected. When a neuron spikes at some rate depending on its electric potential, its membrane potential is set to a minimum value v_{min}, and this makes start, after a small delay, two fronts on the dendrites of all the neurons to which it is connected. Fronts move at constant speed. When two fronts (on the dendrite of the same neuron) collide, they annihilate. When a front hits the soma of a neuron, its potential is increased by a small value w_n. Between jumps, the potentials of the neurons are assumed to drift in $[v_{min}, \infty)$, according to some well-posed ODE. They prove the existence and uniqueness of a heuristically derived mean-field limit of the system when $n \to \infty$ [51].
- E. Tanré has worked with Patricio Orio (CINV, Chile) and Alexandre Richard (Centrale-Supelec) on the modelling and measurement of long-range dependence in neuronal spike trains. They exhibit evidence of memory effect in genuine neuronal data and compared a fractional integrate-and-fire model with the existing Markovian models [31].
- Q. Cormier and E. Tanré studied with Romain Veltz (team MATHNEURO) the long time behavior of a McKean-Vlasov SDE modeling a large assembly of neurons. A convergence to the unique (in this case) invariant measure is obtained assuming that the interactions between the neurons are weak enough. The key quantity in this model is the "firing rate": it gives the average number of jumps per unit of times of the solution of the SDE. They derive a non-linear Voltera equation satisfied by this rate. They used methods from integral equation to control finely the long time behavior of this firing rate [49].

- D. Villemonais collaborates with the Gerontology Service of CHRU Nancy on statistics of time evolution of telomere lengths in human blood cells. This is a collaboration with Anne Gégout Petit (IECL, Inria BIGS), Simon Toupance (CHRU Nancy), Eliane Albuisson (CHRU Nancy), Athanasios Benetos (CHRU Nancy), Daphnée Germain (Ecole des Mines de Nancy). They proposed in [32] a telomeric signature for human beings, stable along age evolution. Lionel Lenôtre works as a post-doc on this topic within the project GEENAGE of LUE.
- D. Villemonais studied with C. Coron (Univ. Paris-Saclay) and S. Méléard (École Polytechnique) the extinction probability before fixation for multi-dimensional models of Wright-Fisher type with mutations [21].
- In collaboration with E. Horton and A. Kyprianou (University of Bath), D. Villemonais studied the large-time asymptotic behaviour of the neutron transport equation in a three-dimensional domain [55]. This work is motivated by the simulation of the flow of particles in a nuclear tank.
- D. Villemonais studied with C. Mailler (University of Bath) the asymptotic behaviour of generalized measure-valued Polya urn models taking values in non-compact sets, using techniques from the theory of stochastic algorithms [58].

5.1.2. Other works in progress

- N. Champagnat, C. Fritsch and S. Billiard (Univ. Lille) are working on food web modeling.
- N. Champagnat and D. Villemonais are working with M. Benaïm (Univ. Neuchatel) on the convergence of stochastic algorithms to the quasi-stationary distribution of diffusion processes absorbed at the boundary of a domain.
- N. Champagnat is working with S. Méléard (École Polytechnique) and C. Tran Viet (Univ. Lille 1) on evolutionary models of bacteria with horizontal transfer. They study a scaling of parameters taking into account the influence of negligible but non-extinct populations, allowing to study specific phenomena observed in these models (re-emergence of traits, cyclic evolutionary dynamics and evolutionary suicide).
- Q. Cormier is investigating new methods to explore the long time behavior of the McKean-Vlasov SDE of [49], to go beyond the weak interactions case. The long time behavior of such McKean-Vlasov equations can be intricate as there can be multiple invariant measures or stable oscillations of the law of the process. The objective of this work is to develop (numerical and theoretical) methods to check the local stability of a given invariant measure of this non-linear SDE.
- C. Fritsch is working with A. Gégout-Petit (Univ. Lorraine and EPI BIGS), B. Marçais (INRA, Nancy) and M. Grosdidier (INRA, Avignon) on a statistical analysis of a Chalara Fraxinea model [34].
- C. Fritsch is working with Marianne Clausel (Univ. Lorraine) and Julien Trombini (Two-I) on the modeling of emotions spreading in a crowd.
- A. Lejay and A. Brault (U. Paris Descartes) continue their work to extend the framework of rough flows.
- O. Faugeras (MATHNEURO Inria Research Team), É. Soret (joint postdoc with MATHNEURO Inria Research Team) and É. Tanré are working on Mean-Field description of thermodynamics limits of large population of neurons with random interactions. They study the asymptotic behaviour for an asymmetric neuronal dynamics in a network of linear Hopfield neurons. They obtain the convergence in law of each component to a Gaussian process. The limit object is not a Markov process.
- P. Helson, E. Tanré and R. Veltz (MATHNEURO Inria team), are working on a neural network model of memory. The aim is to propose a new retrieval criterion and its mathematical analysis.
- E. Tanré has worked with Alexandre Richard (Centrale-Supelec) and Soledad Torres (Universidad de Valparaíso, Chile) on a one-dimensional fractional SDE reflected on the line. The existence and uniqueness of this process is known in the case where the Hurst parameter H of the noise (fBM) is larger than 0.5. They have proved the existence of a penalization scheme (suited to numerical approximation) to approach this object.

VERIDIS Project-Team

7. New Results

7.1. Automated and Interactive Theorem Proving

Participants: Jasmin Christian Blanchette, Martin Bromberger, Daniel El Ouraoui, Mathias Fleury, Pascal Fontaine, Stephan Merz, Hans-Jörg Schurr, Sorin Stratulat, Thomas Sturm, Andreas Teucke, Sophie Tourret, Marco Voigt, Uwe Waldmann, Christoph Weidenbach.

7.1.1. Extension of the Superposition Calculus with λ-free Higher-Order Terms and (Co)datatypes

Joint work with Alexander Bentkamp (VU Amsterdam), Simon Cruanes (Aesthetic Integration), Nicolas Peltier (IMAG Grenoble), and Simon Robillard (Chalmers Gothenburg).

Superposition is a highly successful calculus for reasoning about first-order logic with equality. As a stepping stone towards extending the calculus to full higher-order logic, Bentkamp et al. [19] designed a graceful generalization of the calculus to a fragment devoid of λ-abstractions, but with partial application and application of variables, two crucial higher-order features. This builds on the work on term orders, namely the recursive path order [57] and the Knuth-Bendix order [55]. We implemented the calculi in Simon Cruanes's Zipperposition prover and evaluated them on TPTP benchmarks. The performance is substantially better than with the traditional, encoding-based approach. The new superposition-like calculus serves as a stepping stone towards complete, efficient automatic theorem provers for full higher-order logic.

Another extension of superposition, by Blanchette et al. [21], concerns the native support for inductive and coinductive datatypes. The ability to reason about datatypes has many applications in program verification, formalization of the metatheory of programming languages, and even formalization of mathematics.

Both lines of work aim at bridging the gap between automatic and interactive theorem provers, by increasing the expressiveness and efficiency of best-of-breed automatic first-order provers based on the superposition calculus.

7.1.2. IsaFoL: Isabelle Formalization of Logic

Joint work with Alexander Bentkamp (VU Amsterdam), Andreas Halkjær From (DTU Copenhagen), Alexander Birch Jensen (DTU Copenhagen), Peter Lammich (TU München), John Bruntse Larsen (DTU Copenhagen), Julius Michaelis (TU München), Tobias Nipkow (TU München), Nicolas Peltier (IMAG Grenoble), Simon Robillard (Chalmers Gothenburg), Anders Schlichtkrull (DTU Copenhagen), Dmitriy Traytel (ETH Zürich), Jørgen Villadsen (DTU Copenhagen), and Petar Vukmirović (VU Amsterdam).

Researchers in automated reasoning spend a significant portion of their work time specifying logical calculi and proving metatheorems about them. These proofs are typically carried out with pen and paper, which is error-prone and can be tedious. As proof assistants are becoming easier to use, it makes sense to employ them.

In this spirit, we started an effort, called IsaFoL (Isabelle Formalization of Logic), that aims at developing libraries and methodology for formalizing modern research in the field, using the Isabelle/HOL proof assistant.[0] Our initial emphasis is on established results about propositional and first-order logic.

The main result this year has been a formalization of a large part of Bachmair and Ganzinger's chapter on resolution theorem proving in the *Handbook of Automated Reasoning*, by Anders Schlichtkrull et al. The work was conducted by Schlichtkrull largely during a visit at the MPI in Saarbrücken and was published at IJCAR 2018 [34]. The following quote of one of the reviews nicely sums up the objective of the project:

[0] https://bitbucket.org/isafol/isafol/wiki/Home

> The authors convinced me that their development is a great tool for exploring/developing calculus extensions. It will enable us to "*extend/hack without fear*."

A follow-up paper [33], also by Schlichtkrull et al., has been accepted at CPP 2019. In this work, a chain of refinement leads to a verified executable prover.

The IsaFoL repository has welcome several further additions in 2018, and there is largely finished work, which we expect will lead to at least two publications in 2019:

- After the journal publication [13] following up on an IJCAR 2016 paper and a publication at CPP 2018 [23], Fleury has improved his verified SAT solver IsaSAT further by implementing four optimizations: restarts, forgetting, blocking literals, and machine integers. IsaSAT is now by far the most efficient verified SAT solver, and it is catching up with MiniSat, a reference (but unverified) SAT solver implementation.
- Sophie Tourret and Simon Robillard have formalized a new framework, designed primarily by Uwe Waldmann, that captures abstractly the lifting from completeness of a calculus for propositional logic to a first-order prover. This will yield a simpler proof of Bachmair and Ganzinger's completeness theorem and will be reusable for reasoning about other provers (e.g., superposition provers), whether with pen and paper or in Isabelle.

Jasmin Blanchette briefly describes this ongoing research in an invited paper [20], which he will present at CPP 2019.

7.1.3. Subtropical Reasoning for Real Inequalities

Joint work with Hoon Hong (North Carolina State University, Raleigh, NC).

We consider systems of strict multivariate polynomial inequalities over the reals. All polynomial coefficients are parameters ranging over the reals, where for each coefficient we prescribe its sign. We are interested in the existence of positive real solutions of our system for all choices of coefficients subject to our sign conditions. We give a decision procedure for the existence of such solutions. In the positive case our procedure yields a parametric positive solution as a rational function in the coefficients. Our framework allows heuristic subtropical approaches to be reformulated for non-parametric systems of polynomial inequalities. Such systems have been recently used in qualitative biological network analysis and, independently, in satisfiability modulo theory solving. We apply our results to characterize the incompleteness of those methods.

The approach allows SMT solving for non-linear real arithmetic to be heuristically reduced to linear real arithmetic, to which, e.g., methods from 7.1.4 are applicable. In the special case of single inequalities one can even reduce to linear programming. [25]. This has been successfully applied to heuristic search for Hopf bifurcation fixed points in chemical and biological network analysis.

7.1.4. Reasoning in Linear Arithmetic

We have continued our work on reasoning in linear integer (LIA), linear real (LRA) and linear mixed arithmetic (LIRA). Whereas the standard branch-and-bound techniques [63] for LIA typically work well for bounded systems of inequations, they often diverge on unbounded systems. We already proposed cube techniques for this case. They comprise efficiently computable sufficient tests for the existence of a solution [58]. However, these tests are only necessary for the existence of a solution in the case of a system that is unbounded in all directions. For the case of partially unbounded systems, our combination of the Mixed-Echelon-Hermite transformation and the Double-Bounded Reduction for systems of linear mixed arithmetic preserve satisfiability, can be computed in polynomial time, and turn any LIRA system into a bounded system [22]. Existing approaches for LIRA, e.g., branch-and-bound and cuts from proofs, only explore a finite search space after the application of our two transformations. The transformations orient themselves on the structure of an input system instead of computing *a priori* (over-)approximations out of the available constants. We also developed a polynomial method for converting certificates of (un)satisfiability from the transformed to the original system.

Meanwhile our techniques have been integrated into the SMT solver veriT, but also in other SMT solvers such as Z3 [72] or MathSAT [62]. They have been substantial for our success at SMTComp2018.

7.1.5. Combination of Satisfiability Procedures

Joint work with Christophe Ringeissen (Inria Nancy – Grand Est, Pesto) and Paula Chocron (IIIA-CSIC, Bellaterra, Spain).

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. The design of a generic combination method for non-disjoint unions of theories is difficult, 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).

In 2015, we defined 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 [60]. 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 other theories [61] amenable to similar combinations: this class includes the theory of equality, the theory of absolutely free data structures, and all the theories in between.

In 2018, we have been improving the framework and unified both results. A paper is under review.

7.1.6. Quantifier Handling in SMT

Joint work with Andrew J. Reynolds (Univ. of Iowa, USA) and Cezary Kaliszyk (Univ. of Innsbruck).

SMT solvers generally rely on various instantiation techniques for handling quantifiers. We built a unifying framework encompassing quantified formulas with equality and uninterpreted functions, such that the major instantiation techniques in SMT solving can be cast in that framework. It is based on the problem of E-ground (dis)unification, a variation of the classic Rigid E-unification problem. We introduced a sound and complete calculus to solve this problem in practice: Congruence Closure with Free Variables (CCFV). Experimental evaluations of implementations of CCFV demonstrate notable improvements in the state-of-the-art solver CVC4 and make the solver veriT competitive with state-of-the-art solvers for several benchmark libraries, in particular those originating in verification problems. This was the subject of a publication in 2017 [53]. In a publication at TACAS 2018 [31], we revisit enumerative instantiation for SMT.

We are currently investigating machine learning techniques as a tool for filtering instantiations. Other ongoing work aims at lifting the above techniques to higher-order reasoning.

7.1.7. Real Quantifier Elimination, Decision, and Satisfiability and Their Applications

Effective quantifier elimination procedures for first-order theories provide a powerful tool for generically solving a wide range of problems based on logical specifications. In contrast to general first-order provers, quantifier elimination procedures are based on a fixed set of admissible logical symbols with an implicitly fixed semantics. This admits the use of sub-algorithms from symbolic computation. Specifically quantifier elimination for the reals has been successfully applied in geometry, verification, and the life sciences.

A survey paper with an invited talk at ISSAC 2018 provides a coherent view on the scientific developments of the virtual substitution method for real quantifier elimination during the past three decades [17]. Another recent survey paper had illustrated relevant applications of that method [71].

7.1.8. Non-Linear Arithmetic in SMT

Joint work with M. Ogawa and X. T. Vu (Japan Advanced Institute of Science and Technology), V. K. To (University of Engineering and Technology, VNU, Hanoi, Vietnam).

In the context of the SC^2 project (cf. sections 8.1 and 8.3), we study the theory, design techniques, and implement software to push forward the non-linear arithmetic (NLA) reasoning capabilities in SMT. Previously, we designed a framework to combine interval constraint propagation with other decision procedures for NLA, with promising results, notably in the international competition of SMT solvers. We also studied integration of these procedures into combinations of theories. These ideas were validated through an implementation within the veriT solver, together with code from the raSAT solver (from JAIST), and they were presented at the SC^2 workshop 2018 [24].

7.1.9. Proofs for SMT

We have previously developed a framework for processing formulas in automatic theorem provers, with generation of detailed proofs. The main components are a generic contextual recursion algorithm and an extensible set of inference rules. Clausification, skolemization, theory-specific simplifications, and expansion of 'let' expressions are instances of this framework. With suitable data structures, proof generation adds only a linear-time overhead, and proofs can be checked in linear time. We implemented the approach in the SMT solver veriT. This allowed us to dramatically simplify the code base while increasing the number of problems for which detailed proofs can be produced, which is important for independent checking and reconstruction in proof assistants. This was the subject of a conference publication in 2017. In 2018, we polished the approach, fully implementing proof reconstruction of veriT proofs in Isabelle. A paper has been accepted in the Journal of Automated Reasoning.

7.1.10. A More Efficient Technique for Validating Cyclic Pre-Proofs

Cyclic pre-proofs can be represented as sets of finite tree derivations with back-links. In a setting of first-order logic with inductive definitions, the nodes of the tree derivations are labelled by sequents and the back-links connect particular terminal nodes, referred to as buds, to other nodes labelled by the same sequent. However, only some back-links can constitute sound pre-proofs. Previously, it was shown that special ordering and derivability conditions, defined along the minimal cycles of the digraph representing a particular normal form of the cyclic pre-proof, are sufficient for validating the back-links. In that approach, a single constraint could be checked several times when processing different minimal cycles, hence one may require additional recording mechanisms to avoid redundant computation in order to achieve polynomial time complexity.

In [39], we presented a new approach that does not need to process minimal cycles. It is based on a normal form in which the validation conditions are defined by taking into account only the root-bud paths from the non-singleton strongly connected components of its digraph.

7.1.11. Mechanical Synthesis of Algorithms by Logical and Combinatorial Techniques

Joint work with Isabela Dramnesc (West University, Timisoara, Romania) and Tudor Jebelean (RISC, Johannes Kepler University, Linz, Austria).

In [14], we developed logical and combinatorial methods for automating the generation of sorting algorithms for binary trees, starting from input-output specifications and producing conditional rewrite rules. The main approach consists in proving (constructively) the existence of an appropriate output from every input. The proof may fail if some necessary sub-algorithms are lacking. Then, their specifications are suggested and their synthesis is performed by the same principles.

The main goal is to avoid the possibly prohibitive cost of pure resolution proofs by using a natural-style proving in which domain-specific strategies and inference steps lead to a significant increase of efficiency. We introduce novel techniques and combine them with classical techniques for natural-deduction style proving, as well as methods based on the properties of domain-specific relations and functions. In particular, we use combinatorial techniques in order to generate possible witnesses, which in certain cases lead to the discovery of new induction principles. From the proof, the algorithm is extracted by transforming inductive proof steps into recursions, and case-based proof steps into conditionals.

The approach was demonstrated using the Theorema system for developing the theory, implementing the prover, and performing the proofs of the necessary properties and synthesis conjectures. It was also validated in the Coq system, allowing us to compare the facilities of the two systems in view of our application.

7.1.12. Formal Proofs of Tarjan's Algorithm

Joint work with Ran Chen (Chinese Academy of Sciences), Cyril Cohen and Laurent Théry (Inria Sophia Antipolis Méditerranée, Marelle), and Jean-Jacques Lévy (Inria Paris, Pi.r2).

We compare formal proofs of Tarjan's algorithm for computing strongly connected components in a graph in three different proof assistants: Coq, Isabelle/HOL, and Why3. Our proofs are based on a representation of the algorithm as a functional program (rather than its more conventional imperative representation), which was verified in Why3 by Chen and Lévy [59]. The proofs in all three assistants are thus closely comparable and in particular employ the same invariants. This lets us focus on different formalizations due to idiosyncracies of the proof assistants, such as w.r.t. handling mutually recursive function definitions whose termination is not obvious according to syntactic criteria, and compare the degree of automation in the three assistants. A report is available on arXiv [45].

7.2. Formal Methods for Developing and Analyzing Algorithms and Systems

Participants: Marie Duflot-Kremer, Yann Duplouy, Margaux Duroeulx, Souad Kherroubi, Igor Konnov, Dominique Méry, Stephan Merz, Axel Palaude, Nicolas Schnepf, Christoph Weidenbach.

7.2.1. Parameterized Verification of Threshold-Guarded Fault-Tolerant Distributed Algorithms

Joint work with Nathalie Bertrand (Inria Rennes, SUMO project team) and Jure Kukovec, Marijana Lazić, Ilina Stoilkovska, Josef Widder, Florian Zuleger (TU Wien).

Many fault-tolerant distributed algorithms use threshold guards: processes broadcast messages and count the number of messages that they receive from their peers. Based on the total number n of processes and an upper bound on the number t of faulty processes, a correct process tolerates faults by receiving "sufficiently many" messages. For instance, when a correct process has received $t + 1$ messages from distinct processes, at least one of these messages must originate from a non-faulty process. The main challenge is to verify such algorithms for all combinations of parameters n and t that satisfy a resilience condition, e.g., $n > 3t$.

In earlier work, we introduced threshold automata for representing processes in such algorithms and showed that systems of threshold automata have bounded diameters that do not depend on the parameters such as n and t, provided that a single-step acceleration is allowed [66]. In the contribution [27] to CONCUR'18, we reported on various extensions of this result to less restrictive forms of automata: the guards can be non-linear, shared variables can be incremented and decremented, non-trivial loops are allowed, and more general forms of acceleration are used. In the contribution [26] to ISOLA'18, we presented a parallel extension of our tool Byzantine Model Checker (ByMC), which allows one to distribute the verification queries across the computation nodes in an MPI cluster.

Our previous results apply to asynchronous algorithms. It is well-known that distributed consensus cannot be solved in purely asynchronous systems [64]. However, when an algorithm is provided with a random coin, consensus becomes solvable [56]. In [44], we introduced an approach to parameterized verification of randomized threshold-guarded distributed algorithms, which proceed in an unbounded number of rounds and toss a coin to break symmetries. This approach integrates two levels of reasoning: (1) proving safety and liveness of a single round system with ByMC by replacing randomization with non-determinism, (2) showing almost-sure termination of an algorithm by using the verification results for the non-deterministic system. To show soundness, we proved several theorems that reduce reasoning about multiple rounds to reasoning about a single round. We verified five prominent algorithms, including Ben-Or's randomized consensus [56] and randomized one-step consensus (RS-BOSCO [70]). The verification of the latter algorithm required us to run experiments in Grid5000. A paper describing these results is under review at TACAS 2019.

Another way of making consensus solvable is to impose synchrony on the executions of a distributed system. In [48] we introduced synchronous threshold automata, which execute in lock-step and count the number of processes in given local states. In general, we showed that even reachability of a parameterized set of global states in such a distributed system is undecidable. However, we proved that systems of automata with monotonic guards have bounded diameters, which allows us to use SMT-based bounded model checking as

a complete parameterized verification technique. We introduced a procedure for computing the diameter of a counter system of synchronous threshold automata, applied it to the counter systems of 8 distributed algorithms from the literature, and found that their diameters are tiny (from 1 to 4). This makes our approach practically feasible, despite undecidability in general. A paper about this work is under review at TACAS 2019.

7.2.2. Symbolic Model Checking of TLA+ Specifications

Joint work with Jure Kukovec, Thanh Hai Tran, Josef Widder (TU Wien).

TLA$^+$ is a general language introduced by Leslie Lamport for specifying temporal behavior of computer systems [67]. The tool set for TLA$^+$ includes an explicit-state model checker TLC. As explicit state model checkers do not scale to large verification problems, we started the project APALACHE [0] on developing a symbolic model checker for TLA$^+$ in 2016.

In the contribution [28] to ABZ'18, we addressed the first principal challenge towards developing the symbolic model checker. We introduced a technique for identifying assignments in TLA$^+$ specifications and decomposing a monolithic TLA$^+$ specification into a set of symbolic transitions. At the TLA$^+$ community meeting 2018, we presented a prototype solution [46] to a second challenge. We have developed an SMT encoding of TLA$^+$ expressions for model checking purposes. We presented the first version of a symbolic model checker for TLA$^+$ specifications that works under the same assumptions as TLC: the input parameters are fixed and finite structures, and the reachable states are finite structures. The experimental results are encouraging, and we are thus preparing a conference submission. Finally, in a contribution to the DSN Workshop on Byzantine Consensus and Resilient Blockchains [47], we considered challenges for automatic verification techniques for Blockchain protocols.

7.2.3. Making Explicit Domain Knowledge in Formal System Development

Joint work with partners of the IMPEX project.

The IMPEX project (cf. section 8.1) advocates that formal modeling languages should explicitly represent the knowledge resulting from an analysis of the application domain, and that ontologies are good candidates for handling explicit domain knowledge. We strive at offering rigorous mechanisms for handling domain knowledge in design models. The main results of the project are summarized in [18] and show the importance of three operations over models, namely annotation, dependency, and refactoring [38].

7.2.4. Incremental Development of Systems and Algorithms

Joint work with Manamiary Bruno Andriamiarina, Neeraj Kumar Singh (IRIT, Toulouse), Rosemary Monahan (NUI Maynooth, Ireland), Zheng Cheng (LINA, Nantes), and Mohammed Mosbah (LaBRI, Bordeaux).

The development of distributed algorithms and, more generally, of distributed systems, is a complex, delicate, and challenging process. The approach based on refinement applies a design methodology that starts from the most abstract model and leads, in an incremental way, to a distributed solution. The use of a proof assistant gives a formal guarantee about the conformance of each refinement with the model preceding it. Our main result during 2018 is the development of patterns for different kinds of paradigms including the iterative pattern, the recursive pattern, and the distributed pattern [30].

7.2.5. Synthesis of Security Chains for Software Defined Networks

Joint work with Rémi Badonnel and Abdelkader Lahmadi of the Resist team of Inria Nancy – Grand Est.

The PhD work of Nicolas Schnepf focuses on applying formal methods techniques in the area of network communications, and in particular for the construction, analysis, and optimization of security functions in the setting of software-defined networks (SDN). In previous work, we defined an extension of the Pyretic language [65] for representing both the control and the data planes of SDN controllers and implemented a translation of that extension to the input languages of the nuXmv model checker and of SMT solvers.

[0] WWTF project APALACHE (ICT15-103): https://forsyte.at/research/apalache/

This year, our work focused on synthesizing security chains for Android applications based on their observed communications. The first step consists in inferring probabilistic finite-state automata models that represent network flows generated by Android applications. Comparing our models with automata produced by the state-of-the-art tools Invarimint and Synoptic, we obtain representations that are significantly smaller than those generated by Synoptic and as succinct as those inferred by Invarimint, but that include information about transition probability, unlike Invarimint. This work was presented at NOMS 2018 [35], [37]. In a second step, we encode security policies defined by network administrators in a rule-based program that is then used to generate a high-level representation of a security chain for the application, which is then translated to Pyretic. For example, an application that contacts different ports at the same IP address in rapid succession could be qualified as performing a port scanning attack, and these connections could then be blocked. This work was presented at AVoCS 2018 [36]. The third step consists in factorizing the chains generated for different applications in order to reduce the size of the overall chain that must be deployed in a network. A paper describing appropriate algorithms for that purpose will be presented at IM 2019.

7.2.6. Satisfiability Techniques for Reliability Assessment

Joint work with Nicolae Brînzei at Centre de Recherche en Automatique de Nancy.

The reliability of complex systems is typically assessed using probabilistic methods, based on the probabilities of failures of individual components, relying on graphical representations such as fault trees or reliability block diagrams. Mathematically, the dependency of the overall system on the working status of its components is described by its Boolean-valued *structure function*, and binary decision diagrams (BDDs) have traditionally been used to construct a succinct representation of that function. We explore the use of modern satisfiability techniques as an alternative to BDD-based algorithms. In 2018, our work focused on the encoding of dynamic fault trees whose structure function needs to take into account the order in which components fail.