RESEARCH CENTER
Nancy - Grand Est

FIELD

Activity Report 2017

Section New Results

Edition: 2018-02-19

ALGORITHMICS, PROGRAMMING, SOFTWARE AND ARCHITECTURE
1. CAMUS Team 4
2. CARAMBA Project-Team 9
3. CARTE Team 13
4. GAMBLE Project-Team 16
5. PESTO Project-Team 18
6. VERIDIS Project-Team 23
APPLIED MATHEMATICS, COMPUTATION AND SIMULATION
7. SPHINX Project-Team 29
8. TOSCA Project-Team 32
DIGITAL HEALTH, BIOLOGY AND EARTH
9. BIGS Project-Team 37
10. CAPSID Project-Team 39
11. MIMESIS Team 41
12. NEUROSYS Project-Team 47
13. TONUS Team 50
NETWORKS, SYSTEMS AND SERVICES, DISTRIBUTED COMPUTING
14. COAST Project-Team 54
15. MADYNES Team 57
PERCEPTION, COGNITION AND INTERACTION
16. ALICE Project-Team 66
17. LARSEN Project-Team 70
18. MAGRIT Project-Team 77
19. MULTISPEECH Project-Team 81
20. ORPAILLEUR Project-Team 88
21. SEMAGRAMME Project-Team 95

CAMUS Team

7. New Results

7.1. Automatic (Un-)Collapsing of Non-Rectangular Loops

Participants: Philippe Clauss, Ervin Altıntaş, Matthieu Kuhn.

Loop collapsing is a well-known loop transformation which combines some loops that are perfectly nested into one single loop. It allows to take advantage of the whole amount of parallelism exhibited by the collapsed loops, and provides a perfect load balancing of iterations among the parallel threads.

However, in the current implementations of this loop optimization, as the ones of the OpenMP language, automatic loop collapsing is limited to loops with constant loop bounds that define rectangular iteration spaces, although load imbalance is a particularly crucial issue with non-rectangular loops. The OpenMP language addresses load balance mostly through dynamic runtime scheduling of the parallel threads. Nevertheless, this runtime schedule introduces some unavoidable execution-time overhead, while preventing to exploit the entire parallelism of all the parallel loops.

We propose a technique to automatically collapse any perfectly nested loops defining non-rectangular iteration spaces, whose bounds are linear functions of the loop iterators. Such spaces may be triangular, tetrahedral, trapezoidal, rhomboidal or parallelepiped. Our solution is based on original mathematical results addressing the inversion of a multi-variate polynomial that defines a ranking of the integer points contained in a convex polyhedron.

We show on a set of non-rectangular loop nests that our technique allows to generate parallel OpenMP codes that outperform the original parallel loop nests, parallelized either by using options “static” or “dynamic” of the OpenMP-schedule clause. A conference paper presenting these results, co-authored by Philippe Clauss, Ervin Altıntaş (Master student) and Matthieu Kuhn (Inria Bordeaux Sud-Ouest, team HIEPACS), has been published at the International Parallel and Distributed Processing Symposium (IPDPS) [15].

We are currently developing a technique to also provide good load balancing when parallelizing non-rectangular loops carrying dependences. This new technique has been called *loop uncollapsing*. The idea is to split the outermost parallel loop into two nested loops, such that the new outermost loop, when parallelized, results in well-balanced parallel threads.

7.2. Code-Bones for Fast and Flexible Runtime Code Generation

Participants: Juan Manuel Martinez Caamaño, Manuel Selva, Philippe Clauss.

We have developed a new runtime code generation technique for speculative loop optimization and parallelization. The main benefit of this technique, compared to previous approaches, is to enable advanced optimizing loop transformations at runtime with an acceptable time overhead. The loop transformations that may be applied are those handled by the polyhedral model. The proposed code generation strategy is based on the generation of *code-bones* at compile-time, which are parametrized code snippets either dedicated to speculation management or to computations of the original target program. These code bones are then instantiated and assembled at runtime to constitute the speculatively-optimized code, as soon as an optimizing polyhedral transformation has been determined. Their granularity threshold is sufficient to apply any polyhedral transformation, while still enabling fast runtime code generation. This approach has been implemented in the speculative loop parallelizing framework Apollo, and has been more recently extended to also support loops exhibiting a non-linear behavior thanks to a modeling using “tubes”. The whole approach has been published in Concurrency and Computation: Practice and Experience [11].

7.3. Formal Proofs about Explicitly Parallel Programs with Clocks

Participants: Alain Ketterlin, Éric Violard, Tomofumi Yuki, Paul Feautrier.

We have continued this year our work on formalizing the *happens-before* relation in explicitly parallel programs of the X10 family. Our goal is to define, for certain classes of programs, a relation between instances of elementary instructions that guarantees that one instance necessarily executes before another. Our toy language includes static-control counted loops and conditionals, as well as the usual `finish` and `async` parallel constructs. Moreover, parallel activities can synchronize though the use of *clocks*, which are barriers with dynamic membership. Clocks partition the execution into phases, and profoundly modify the happens-before relation.

This year's work has focused on correctly accounting for the possibility to define specific activities that execute irrespective of the discipline of the clock in scope, so-called *detached* activities. The presence of such activities modifies the notion of phase number, because they let their instructions execute across a range of clock-phases. Our generic notion of phase *ranking* had to be modified. Similarly, the natural semantics we defined had to be slightly modified to correctly represent the parallel execution of both clocked and detached activities. In practice, almost every lemma of the Coq proof has changed, and new definitions were introduced. The new definition of happens-before preserves all desirable properties: it is correct and complete, and is a strict partial order. There is one unpleasant aspect of detached activities that had a strong impact on happens-before: the possibility of deadlocks. A significant part of new definitions and lemmas are devoted to explicit the conditions under which programs terminate. A useful outcome of this part of the mechanization is a static, compile-time deadlock detection criterion.

Most of this work has been described in a paper currently under submission (this paper will be on HAL as soon as anonymity constraints permit). However, the diversity of themes covered in this research (compilation of static-control programs, especially those that fit the polyhedral model, but also semantic modeling of explicitly parallel programs, and formal proofs) make us contemplate the redaction of a much longer paper, which we plan to start at the beginning of next year. At the same time, this work (especially the part about deadlocks) led us to start designing an happens-before relation for a language where multiple clocks can share (part of) their scopes. We hope to be able to advance the formalization of this new family of languages in the near future.

7.4. 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. Typically, a simulation involves a cluster of machines, each machine hosting several cores, and each core being able to execute vectorized instructions (SIMD). The challenge is to efficiently exploit these three levels of parallelism. Regarding the processing on one given multicore machine, existing algorithms either suffer from suboptimal execution time, due to sorting operations or use of atomic instructions, or suffer from suboptimal space usage. We have developed a novel parallel algorithm for PIC simulations on multicore hardware that features asymptotically-optimal memory consumption, and that does not perform unnecessary accesses to the main memory. The algorithm relies on the use of *chunk bags*, i.e., linked lists of fixed-capacity arrays, for storing particles and allowing to process them efficiently using SIMD instructions. Practical results show excellent scalability on the classical Landau damping and two-stream instability test cases. A paper was published at PPAM [12].

7.5. 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 this year 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 [29], 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 were submitted to an international conference.

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 submitted to an international conference.

7.6. Program verification and formal languages

Participant: Arthur Charguéraud.

- A. Charguéraud and François Pottier (Inria Paris) extended their formalization of the correctness and asymptotic complexity of the classic Union Find data structure, which features the bound expressed in terms of the inverse Ackermann function. The proof, conducted using CFML extended with time credits, was refined using a slightly more complex potential function, allowing to derive a simpler and richer interface for the data structure. This work appeared in the Journal of Automated Reasoning (JAR) [9].
- A. Charguéraud and F. Pottier have developed an extension of Separation Logic with temporary read-only permissions. This mechanism allows to temporarily convert any assertion (or "permission") to a read-only form. Unlike with fractional permissions, no accounting is required: the proposed read-only permissions can be freely duplicated and discarded. Where mutable data structures are temporarily accessed only for reading, the proposed read-only permissions enable more concise specifications and proofs. All the metatheory is verified in Coq. An article was presented at ESOP [14].
- 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 submitted paper to an international conference.
- A. Charguéraud has made progress towards CFML 2.0, a reimplementation of CFML entirely inside Coq. In contrast, the initial version of CFML, developed in A. Charguéraud's PhD thesis, is based on an external tool that parses OCaml source code and produces Coq axioms describing their semantics. The new version will remove the need for axioms, thereby further reducing the trusted code base. Furthermore, CFML 2.0 provides a more general memory model, designed to also accomodate formal reasoning about C-style programs, in future work. In passing, A. Charguéraud performed a complete cleanup of the TLC Coq library, which is used extensively by CFML, leading to the beta release of TLC 2.0.

- 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* [31]. 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 has been submitted, and the tool has been presented to the JavaScript standardization committee (ECMA) in November 2017.

7.7. Combining Locking and Data Management Interfaces

Participants: Jens Gustedt, Mariem Saied, Daniel Salas.

Handling data consistency in parallel and distributed settings is a challenging task, in particular if we want to allow for an easy to handle asynchronism between tasks. Our publication [1] 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 6.8 . In previous work it has demonstrated its efficiency for a large variety of platforms.

This year, wee have been able to use the knowledge of the communication structure of ORWL programs to map tasks to cores and thereby achieve interesting performance gains on multicore architectures, see [16]. We propose a topology-aware placement module that is based on the Hardware Locality framework, HWLOC, and that takes the characteristics of the application, of the runtime and of the architecture into account. The aim is double. On one hand we increase the abstraction and the portability of the framework, and on the other hand we enhance the performance of the model's runtime.

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.

7.8. Automatic Generation of Adaptive Simulation Codes

Participants: Cédric Bastoul, Maxime Schmitt.

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

This research direction started in 2015 and complements our other efforts on dynamic optimization. In 2016, we started a collaboration on this topic with Inria Nancy - Grand Est team TONUS, specialized on applied mathematics (contact: Philippe Helluy), to bring models and techniques from this field to compilers. This collaboration received the support from the excellence laboratory (LabEx) IRMIA through the funding of the thesis of Maxime Schmitt on this topic. Two papers on this new research direction has been accepted this year on this topic (IMPACT 2017 workshop, HiPC 2017 conference [20]).

7.9. Parallel Polyhedral Regions

Participants: Cédric Bastoul, Vincent Loechner, Harenome Ranaivoarivony-Razanajato.

Nowadays best performing automatic parallelizers and data locality optimizers for static control programs rely on the polyhedral model. State-of-the-art polyhedral compilers generate only one type of parallelism when targeting multicore shared memory architectures: parallel loops via the OpenMP `omp parallel for` directive.

We propose to explore how a polyhedral compiler could exploit parallel region constructs. Instead of initializing a new set of threads each time the code enters a parallel loop and synchronizing them when exiting it, the threads are initialized once for all at the entrance of the region of interest, and synchronized only when it is necessary.

Technically, the whole region containing parallel loops is embedded in an `omp parallel` construct. Inside the parallel region, the `single` construct is used when some code needs to be executed sequentially; the `for` construct is used to distribute loop iterations between threads. Thanks to the power of the polyhedral dependence analysis, we compute when it is valid to add the optional `nowait` clause, to omit the implicit barrier at the end of a worksharing construct and thus to reduce even more control overhead.

This work was published and presented at the HiPC 2017 conference [19].

7.10. Optimization of Sparse Triangular and Banded Matrix Codes

Participants: Vincent Loechner, Rachid Seghir, Toufik Baroudi.

This work is a collaboration between Vincent Loechner and Rachid Seghir from University of Batna (Algeria). Toufik Baroudi is a second year PhD student under his supervision. Rachid Seghir was visiting the CAMUS team from March 25th to April 8th, 2017.

In this work, we enabled static polyhedral optimization techniques to handle sparse matrix storage formats. When handling sparse triangular and banded matrices in their packed formats, such as in the LAPACK library band storage, loop nests bounds and array references of the resulting codes are not affine functions. We proposed to use a new 2d-packed layout and simple affine transformations to enable polyhedral optimization of sparse triangular and banded matrix operations. The effectiveness of our proposal was shown through an experimental study over a large set of linear algebra benchmarks.

These results were published in ACM TACO [8], and will be presented at the HiPEAC conference in January 2018.

CARAMBA Project-Team

7. New Results

7.1. Improved Complexity Bounds for Counting Points on Hyperelliptic Curves

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

In [16], we present 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.

7.2. Deciphering of a Code Used by a 19th Century Parisian Violin Dealer

Participant: Pierrick Gaudry.

This paper [4] is joint work with Jean-Philippe Échard, Curator at the Cité de la Musique, Paris.

The study of three ledgers from the archives of a prominent Parisian violin maker's workshop (active from 1796 to 1948) reveals that some of their content was encrypted. We present the deciphering of the code, and a discussion of its use in the context of the workshop. Charles-Adolphe Gand introduced this code around 1847 to encrypt values of antique/used violins he would buy and resell. His successors maintained the use of this code at least until 1921. Taking a few examples of instruments by Stradivari and other violin makers, we illustrate how the decoded ledgers – listing transactions for more than 2,500 instruments – are of high interest as historical sources documenting the margins, rebates, and commercial practices of these violin dealers. More generally, we contribute to better describing the evolution of the market for antique instruments of the violin family.

7.3. Discrete Logarithm Record Computation in Extension Fields

Participants: Laurent Grémy, Aurore Guillevic, Emmanuel Thomé.

Together with F. Morain from the GRACE team, we reached new record sizes for the discrete logarithm problems over non-prime finite fields of small extension degrees [19], [8]. Assessing the hardness of the discrete logarithm problem in such fields is highly relevant to the security of cryptographic pairings. Our computations are not terribly large computations compared to other record-size computations for integer factoring or discrete logarithm over prime fields, but on the other hand more novelty is present in these contexts: use of automorphisms, higher degree sieving, for example.

Further research in this direction is needed, especially regarding the effectiveness of the variants of the "tower" number field sieve variants.

Furthermore, A. Guillevic and L. Grémy have gathered in a database all published records of discrete logarithm computations in all kinds of finite fields. The database is hosted on gitlab and is open to external contributions. A web interface for browsing the database is available at http://perso.ens-lyon.fr/laurent.gremy/dldb/index.html.

7.4. Using Constraint Programming to Solve a Cryptanalytic Problem

Participant: Marine Minier.

In [7], we describe Constraint Programming (CP) models to solve a cryptanalytic problem: the related key differential attack against the standard block cipher AES. We show that CP solvers are able to solve these problems quicker than dedicated cryptanalysis tools, and we prove that the 11 rounds solution on AES-192 claimed to be optimal is wrong. Instead, we provide the best related key differential characteristic on 10 rounds of AES-192. We also improved the related-key distinguisher and the basic related-key differential attack on the full AES-256 by a factor 2^6 and the q-multicollisions by a factor 2.

7.5. Optimized Binary64 and Binary128 Arithmetic with GNU MPFR

Participant: Paul Zimmermann.

Together with Vincent Lefèvre (ARIC team, Inria Rhône-Alpes), Paul Zimmermann wrote an article "Optimized Binary64 and Binary128 Arithmetic with GNU MPFR", and presented it at the 24th IEEE Symposium on Computer Arithmetic [9]. This article describes algorithms used to optimize the GNU MPFR library when the operands fit into one or two words. On modern processors, a correctly rounded addition of two quadruple precision numbers is now performed in 22 cycles, a subtraction in 24 cycles, a multiplication in 32 cycles, a division in 64 cycles, and a square root in 69 cycles. It also introduces a new faithful rounding mode, which enables even faster computations. These optimizations will be available in version 4 of MPFR.

7.6. A New Measure for Root Optimization

Participants: Nicolas David, Paul Zimmermann.

In the General Number Field Sieve (GNFS) for integer factorization or discrete logarithm, the first stage is polynomial selection. Polynomial selection itself consists in two steps: size-optimization and root-optimization. The classical measures used to rank polynomials during the root-optimization are the so-called α and Murphy-E values. During the internship of Nicolas David, it was shown that these classical measures might be off by up to 15% between two polynomial pairs, compared to a sieving test. A new measure that better corresponds to sieving tests was designed. An article describing these new results is in preparation.

7.7. Mathematical Computation with SageMath

Participant: Paul Zimmermann.

Starting in March, Paul Zimmermann coordinated the English translation of the book "Calcul mathématique avec Sage", and the update from version 5.9 to 8.0 of Sage. He also translated several chapters and proof-read the translation of all chapters. The current state of the English translation is available under a Creative Commons license (CC BY-SA) at https://members.loria.fr/PZimmermann/sagebook/english.html. A discussion is in process with an editor to publish a paper version.

7.8. Topics in Computational Number Theory Inspired by Peter L. Montgomery

Participants: Emmanuel Thomé, Paul Zimmermann.

Emmanuel Thomé and Paul Zimmermann contributed two chapters of the book "Topics in Computational Number Theory Inspired by Peter L. Montgomery", coordinated by Arjen Lenstra and Joppe Bos, and published by Cambridge University Press. Together with Richard P. Brent and Alexander Kruppa, Paul Zimmermann wrote a chapter entitled "FFT extension for algebraic-group factorization algorithms" [12]. Emmanuel Thomé contributed a chapter entitled "The block Lanczos algorithm" [14].

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

Participant: Svyatoslav Covanov.

In [17], we describe a method improving on the exhaustive search algorithm developed in [26]. 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 prove 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.

7.10. Big Prime Field FFT on the GPU

Participant: Svyatoslav Covanov.

In colloboration with L. Chen, D. Mohajerani and M. Moreno Maza, in [11], we compare various methods for the multiplication of polynomials, using the GPU. We compare the CRT method, using k machine-word primes, to the generalized Fermat prime method, for a prime of k machine-words, inspired by the work in [28]. For some degrees and k, we prove that the arithmetic operations with the generalized Fermat primes offer attractive performance both in terms of algebraic complexity and parallelism.

7.11. CM Plane Quartics

Participant: Hugo Labrande.

As a by-product of his PhD thesis defended in late 2016, Hugo Labrande contributed to a joint work with several authors, leading to an article [21] that provides examples of smooth plane quartics over $\mathbb{Q}$ with complex multiplication over $\overline{\mathbb{Q}}$ by a maximal order with primitive CM type. Several algorithms are used, in tight connection to the computation of Theta functions which was improved in Labrande's PhD thesis: reduction of period matrices, fast computation of Dixmier-Ohno invariants, and reconstruction from these invariants.

7.12. Explicit Isogenies in Genus 2 and 3

Participant: Enea Milio.

In [22], we present a quasi-linear algorithm to compute isogenies between Jacobians of curves of genus 2 and 3 starting from the equation of the curve and a maximal isotropic subgroup of the ℓ-torsion, for ℓ an odd prime number, generalizing Vélu's formula of genus 1. This work is based on the paper "Computing functions on Jacobians and their quotients" of Jean-Marc Couveignes and Tony Ezome. We improve their genus 2 case algorithm, generalize it for genus 3 hyperelliptic curves and introduce a way to deal with the genus 3 non-hyperelliptic case, using algebraic Theta functions.

7.13. Modular Polynomials of Hilbert Surfaces

Participant: Enea Milio.

In [23], together with Damien Robert from the LFANT team, we describe an evaluation/interpolation approach to compute modular polynomials on a Hilbert surface, which parametrizes abelian surfaces with maximal real multiplication. Under some heuristics we obtain a quasi-linear algorithm. The corresponding modular polynomials are much smaller than the ones on the Siegel threefold. We explain how to compute even smaller polynomials by using pullbacks of Theta functions to the Hilbert surface, and give an application to the CRT method to construct class polynomials.

7.14. Individual Logarithm Step in Non-prime Fields

Participant: Aurore Guillevic.

In [20], the previous work [33] on speeding-up the first phase of the individual discrete logarithm computation, the initial splitting, a.k.a. smoothing phase, is extended to any non-prime finite field $\mathbb{F}_{p^n}$ where n is composite. It is also applied to the new variant Tower-NFS.

7.15. Last Year Results that Appeared in 2017

Our work [6], in collaboration with J. Fried and N. Heninger from the University of Pennsylvania, describing a kilobit discrete logarithm computation for a trapdoored prime number has been published in Eurocrypt 2017.

A paper detailing the implementation of the ECM factoring algorithm on the Kalray MPPA-256 many-core processor, written as a collaboration between Jérémie Detrey and Pierrick Gaudry from CARAMBA, and Masahiro Ishii, Atsuo Inomata, and Kazutoshi Fujikawa from NAIST (Nara, Japan), was published in IEEE Transaction on Computers [2].

In [39], the notions of Square, saturation, integrals, multisets, bit patterns and tuples cryptanalysis are revised. A new Slice & Fuse paradigm to better exploit multiset type properties of block ciphers is proposed. With this refined analysis, we improve the best bounds proposed in such contexts against the following block ciphers: Threefish, Prince, Present and Rectangle.

In [3], we improve the existing impossible-differential attacks against Rijndael-160 and Rijndael-224.

Our work [10] about the computational power of the Measurement-based Quantum Computation model, written by Luc Sanselme and Simon Perdrix (from the CARTE team at LORIA), has appeared.

CARTE Team

7. New Results

7.1. Quantum Computing

Participants: Emmanuel Jeandel, Simon Perdrix, Renaud Vilmart.

- **ZX-calculus**

  The ZX-Calculus is a powerful graphical language for quantum mechanics and quantum information processing. The completeness of the language – i.e. the ability to derive any true equation – is a crucial question. In the quest for a complete ZX-calculus, supplementarity has been recently proved to be necessary for quantum diagram reasoning [73]. Roughly speaking, supplementarity consists in merging two subdiagrams when they are parameterized by antipodal angles. In [22], we introduce a generalised supplementarity – called cyclotomic supplementarity – which consists in merging n subdiagrams at once, when the n angles divide the circle into equal parts. We show that when n is an odd prime number, the cyclotomic supplementarity cannot be derived, leading to a countable family of new axioms for diagrammatic quantum reasoning. We exhibit another new simple axiom that cannot be derived from the existing rules of the ZX-Calculus, implying in particular the incompleteness of the language for the so-called Clifford+T quantum mechanics. We end up with a new axiomatisation of an extended ZX-Calculus, including an axiom schema for the cyclotomic supplementarity. This work has been presented at MFCS 2017 [22].

  The ZX-Calculus is devoted to represent complex quantum evolutions. But the advantages of quantum computing still exist when working with rebits, and evolutions with real coefficients. Some models explicitly use rebits, but the ZX-Calculus cannot handle these evolutions as it is. Hence, in [21], we define an alternative language solely dealing with real matrices, with a new set of rules. We show that three of its non-trivial rules are not derivable from the other ones and we prove that the language is complete for the $\pi/2$-fragment. We define a generalisation of the Hadamard node, and exhibit two interpretations from and to the ZX-Calculus, showing the consistency between the two languages. This work has been presented at QPL 2017 [21].

- **Causality and Quantum Computing**

  Since the classic no-go theorems by [43] and [65], contextuality has gained great importance in the development of quantum information and computation. This key characteristic feature of quantum mechanics represents one of the most valuable resources at our disposal to break through the limits of classical computation and information processing, with various concrete application

  An important class of contextuality arguments in quantum foundations are the All-versus-Nothing (AvN) proofs, generalising a construction originally due to Mermin. In [11], we present a general formulation of All-versus-Nothing arguments, and a complete characterisation of all such arguments which arise from stabiliser states. We show that every AvN argument for an n-qubit stabiliser state can be reduced to an AvN proof for a three-qubit state which is local Clifford-equivalent to the tripartite GHZ state. This is achieved through a combinatorial characterisation of AvN arguments, the AvN triple Theorem, whose proof makes use of the theory of graph states. This result enables the development of a computational method to generate all the AvN arguments in Z2 on n-qubit stabiliser states. We also present new insights into the stabiliser formalism and its connections with logic. This work has been presented at QPL 2017 [25] and published in the Philosophical Transactions of the Royal Society A [11].

Analyzing pseudo-telepathy graph games, we propose in [15] a way to build contextuality scenarios exhibiting the quantum supremacy using graph states. We consider the combinatorial structures generating equivalent scenarios. We investigate which scenarios are more multipartite and show that there exist graphs generating scenarios with a linear multipartiteness width. This work has been presented at FCT 2017 [15].

- **Measurement-based Quantum Computing**

  Measurement-based quantum computing (MBQC) is a universal model for quantum computation [74]. The combinatorial characterisation of determinism in this model [51], [48], [69], powered by measurements, and hence, fundamentally probabilistic, is the cornerstone of most of the breakthrough results in this field. The most general known sufficient condition for a deterministic MBQC to be driven is that the underlying graph of the computation has a particular kind of flow called Pauli flow. The necessity of the Pauli flow was an open question [48]. In [23], we show that the Pauli flow is necessary for real-MBQC, and not in general providing counterexamples for (complex) MBQC. We explore the consequences of this result for real MBQC and its applications. Real MBQC and more generally real quantum computing is known to be universal for quantum computing. Real MBQC has been used for interactive proofs by McKague. The two-prover case corresponds to real-MBQC on bipartite graphs. While (complex) MBQC on bipartite graphs are universal, the universality of real MBQC on bipartite graphs was an open question. We show that real bipartite MBQC is not universal proving that all measurements of real bipartite MBQC can be parallelised leading to constant depth computations. As a consequence, McKague techniques cannot lead to two-prover interactive proofs. This work has been presented at FCT 2017 [23].

7.2. Cellular automata as a model of computation

Participants: Nazim Fatès, Irène Marcovici.

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 (India), we studied a particular form of reversibility: the possibility of returning infinitely often to the initial condition after a random number of time step,s this is the recurrence property of the system. We analysed this property for the simple rules and described the communication graph of the system [33].

We studied how to coordinate a team of agents to locate a hidden source on a two-dimensional discrete grid. The challenge is to find the position of the source with only sporadic detections. This problem arises in various situations, for instance when insects emit pheromones to attract their partners. A search mechanism named infotaxis was proposed to explain how agents may progressively approach the source by using only intermittent detections. With Q. Ladeveze, an intern, we re-examined in detail the properties of our bio-inspired algorithm that relies on the Reaction–Diffusion–Chemotaxis aggregation scheme to group agents that have limited abilities [38].

To study the robustness of asynchronous CA, we examined the coalescence phenomenon, which consists in observing the cases where two different initial conditions with the same sequence of updates quickly evolve to the same non-trivial configuration. With J. Francès de Mas, an intern, we studied the rules which always coalesce and those which exhibit a phase transition between a coalescing and non-coalescing behaviour. We proposed some formal explanations of non-trivial rapid coalescence giving lower bounds for the coalescence time of ECA 154 and ECA 62, and some first steps towards finding their upper bounds in order to prove that they have, respectively, quadratic and linear coalescence time [34].

We studied random mixtures of two deterministic Elementary Cellular Automata. There are 8088 such rules, called, diploid cellular automata. We used numerical simulations to perform some steps in the exploration of this space. As the mathematical analysis of such systems is a difficult task, we used numerical simulations to get insights into the dynamics of this class of stochastic cellular automata. We examined phase transitions and various types of symmetry breaking [17].

7.3. Extension of computable functions

Participant: Mathieu Hoyrup.

We worked on the computable aspects of an elementary problem in real analysis: extending a continuous function on a larger domain. More precisely, if a real-valued function f is defined on an interval $[0, a)$ (with $0 < a < 1$) and is computable there, under which conditions can it be extended to a computable function on $[0, 1]$? Although this question has a very simple formulation, it does not have a simple answer. We obtained many results showing how the answer depends on a and on the way f converges at a. Surprisingly, this problem provides new characterizations of already existing classes of real numbers previously defined in computability theory. This work is joint with Walid Gomaa and has been presented at LICS 2017 [19].

7.4. Genericity of weakly computable objects

Participant: Mathieu Hoyrup.

Computability theory abounds with classes of objects, defined for instance in terms of the computability content of the objects. A natural problem is then to compare these classes and separate them when possible. In order to separate two classes, one has to build an object that belongs to one class but not the other. So this object has to be computable in one sense but not the other. We show that in many cases these computability properties have a topological interpretation, and that the object to build must be at the same time computable in some weak topology (*weakly computable*) but *generic* in a stronger topology. We prove a general theorem stating the existence of such objects, thus providing a very handy tool to separate many classes. We use it in the study of the extension of computable functions (previous result) and in other situations. These results are presented in [13].

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 [9] 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 [39], 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 show that our reliable numerical approach can handle classes of examples that are not reachable by symbolic methods.

7.1.2. Computing effectively stabilizing controllers for a class of nD systems

In this paper [1], we study the internal stabilizability and internal stabilization problems for multidimensional (nD) systems. Within the fractional representation approach, a multidimensional system can be studied by means of matrices with entries in the integral domain of structurally stable rational fractions, namely the ring of rational functions which have no poles in the closed unit polydisc $\overline{\mathbb{U}}^n = \{z = (z_1, ..., z_n) \in \mathbb{C}^n \mid |z_1| \leq 1, ..., |z_n| \leq 1\}$.

It is known that the internal stabilizability of a multidimensional system can be investigated by studying a certain polynomial ideal $I = \langle p_1, ..., p_r \rangle$ that can be explicitly described in terms of the transfer matrix of the plant. More precisely the system is stabilizable if and only if $V(I) = \{z \in \mathbb{C}^n \mid p_1(z) = \cdots = p_r(z) = 0\} \cap \overline{\mathbb{U}}^n = \varnothing$. In the present article, we consider the specific class of linear nD systems (which includes the class of 2D systems) for which the ideal I is zero-dimensional, i.e., the p_i's have only a finite number of common complex zeros. We propose effective symbolic-numeric algorithms for testing if $V(I) \cap \overline{\mathbb{U}}^n = \varnothing$, as well as for computing, if it exists, a stable polynomial $p \in I$ which allows the effective computation of a stabilizing controller. We illustrate our algorithms through an example and finally provide running times of prototype implementations for 2D and 3D systems.

7.2. Non-Euclidean Computational Geometry

Participants: Vincent Despré, Iordan Iordanov, Monique Teillaud.

7.2.1. Implementing Delaunay Triangulations of the Bolza Surface

The CGAL library offers software packages to compute Delaunay triangulations of the (flat) torus of genus one in two and three dimensions. To the best of our knowledge, there is no available software for the simplest possible extension, i.e., the Bolza surface, a hyperbolic manifold homeomorphic to a torus of genus two. We present an implementation based on the theoretical results and the incremental algorithm proposed recently. We describe the representation of the triangulation, we detail the different steps of the algorithm, we study predicates, and report experimental results [5]. The implementation is publicly available in the development branch of CGAL on `github`[0] and will soon be submitted for integration in the library.

[0] https://members.loria.fr/Monique.Teillaud/DT_Bolza_SoCG17/

7.3. Probabilistic Analysis of Geometric Data Structures and Algorithms

Participants: Olivier Devillers, Charles Duménil.

7.3.1. Delaunay triangulation of a random sample of a good sample has linear size

A good sample is a point set such that any ball of radius ϵ contains a constant number of points. The Delaunay triangulation of a good sample is proved to have linear size, unfortunately this is not enough to ensure a good time complexity of the randomized incremental construction of the Delaunay triangulation. In this paper we prove that a random Bernoulli sample of a good sample has a triangulation of linear size. This result allows to prove that the randomized incremental construction needs an expected linear size and an expected $O(n \log n)$ time [8].

This work was done in collaboration with Marc Glisse (Project-team DATASHAPE).

7.3.2. Delaunay triangulation of a random sampling of a generic surface

The complexity of the Delaunay triangulation of n points distributed on a surface ranges from linear to quadratic. We prove that when the points are evenly distributed on a smooth compact generic surface the expected size of the Delaunay triangulation is $O(n)$. This result has to be compared with a bound of $O(n \log n)$ when the points are a deterministic good sample of the surface under the same hypotheses on the surface [13].

7.4. Classical Computational Geometry and Graph Drawing

Participants: Olivier Devillers, Sylvain Lazard.

7.4.1. Celestial Walk: A Terminating Oblivious Walk for Convex Subdivisions

We present a new oblivious walking strategy for convex subdivisions. Our walk is faster than the straight walk and more generally applicable than the visiblity walk. To prove termination of our walk we use a novel monotonically decreasing distance measure [10].

This work was done in collaboration with Wouter Kuijper and Victor Ermolaev (Nedap Security Management).

7.4.2. 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, the size of $\mathcal{Q}$ is $O(n^{15})$, but, under reasonable hypotheses, this complexities decreases to $O(n^5)$.

This work was done in collaboration with William J. Lenhart (Williams College, USA).

7.4.3. Explicit array-based compact data structures for triangulations

We consider the problem of designing space efficient solutions for representing triangle meshes. Our main result is a new explicit data structure for compactly representing planar triangulations: if one is allowed to permute input vertices, then a triangulation with n vertices requires at most $4n$ references ($5n$ references if vertex permutations are not allowed). Our solution combines existing techniques from mesh encoding with a novel use of maximal Schnyder woods. Our approach extends to higher genus triangulations and could be applied to other families of meshes (such as quadrangular or polygonal meshes). As far as we know, our solution provides the most parsimonious data structures for triangulations, allowing constant time navigation. Our data structures require linear construction time, and are fast decodable from a standard compressed format without using additional memory allocation. All bounds, concerning storage requirements and navigation performances, hold in the worst case. We have implemented and tested our results, and experiments confirm the practical interest of compact data structures.

This work was done in collaboration with Luca Castelli Aleardi (LIX).

PESTO Project-Team

7. New Results

7.1. Modelling

7.1.1. New protocol and adversary models

Participants: Jannik Dreier, Steve Kremer, Ludovic Robin.

Symbolic models for security protocol verification, following the seminal ideas of Dolev and Yao, come in many flavors, even though they share the same ideas. A common assumption is that the attacker has complete control over the network: he can therefore intercept any message. Depending on the precise model this may be reflected either by the fact that any protocol output is directly routed to the adversary, or communications may be among any two participants, including the attacker—the scheduling between which exact parties the communication happens is left to the attacker. These two models may seem equivalent at first glance and, depending on the verification tools, either one or the other semantics is implemented. In collaboration with Babel (IIT Bombay) we show that, unsurprisingly, they indeed coincide for reachability properties. However, when we consider equivalence properties, we prove that these two semantics are incomparable. We also introduce a new semantics, where internal communications are allowed but messages are always eavesdropped by the attacker. We show that this new semantics yields strictly stronger equivalence relations and identify two subclasses of protocols for which the three semantics coincide. These results were presented at POST'17 [16].

Isolated Execution Environments (IEEs), such as ARM TrustZone and Intel SGX, offer the possibility to execute sensitive code in isolation from other, potentially malicious programs, running on the same machine, or a potentially corrupted OS. A key feature of IEEs is the ability to produce reports binding cryptographically a message to the program that produced it, typically ensuring that this message is the result of the given program running on an IEE. In collaboration with Jacomme (ENS Cachan) and Scerri (Univ Bristol), Kremer presented a symbolic model for specifying and verifying applications that make use of such features. For this they introduced the SℓAPiC process calculus to reason about reports issued at given locations. They also provide tool support, extending the *SAPIC/TAMARIN* toolchain and demonstrate the applicability of their framework on several examples implementing secure outsourced computation (SOC), a secure licensing protocol and a one-time password protocol that all rely on such IEEs. This work has been published and presented at EuroS&P'17 [30].

Modern security protocols may involve humans in order to compare or copy short strings between different devices. Multi-factor authentication protocols, such as Google 2-factor or 3D-secure are typical examples of such protocols. However, such short strings may be subject to brute force attacks. In collaboration with Delaune (IRISA), we propose a symbolic model which includes attacker capabilities for both guessing short strings, and producing collisions when short strings result from an application of weak hash functions. We propose a new decision procedure for analysing (a bounded number of sessions of) protocols that rely on short strings. The procedure has been integrated in the *Akiss* tool and tested on protocols from the ISO/IEC 9798-6:2010 standard. This work has been published and presented at CSF'17 [26].

Most security properties are modelled as *safety* properties (*"bad things do not happen"*). Another important class of properties is that of *liveness* properties (*"eventually, good things happen"*). Reasoning about the class of *liveness* properties of cryptographic protocols, has received little attention in the literature, even though this class is vital in many security-sensitive applications, such as fair exchange protocols, or security layers in industrial control systems. In collaboration with Backes and Künnemann (Univ Saarland, Germany), Dreier and Kremer have designed a protocol and adversary model that are suitable for reasoning about liveness properties. Tool support is also provided by extending the *SAPIC/TAMARIN* tool chain and several case studies demonstrate the effectiveness of the approach. This work has been published and presented at EuroS&P'17 [17].

7.1.2. New properties

Participant: Jannik Dreier.

Industrial systems are nowadays regularly the target of cyberattacks, the most famous being Stuxnet[0]. At the same time such systems are increasingly interconnected with other systems and insecure media such as Internet. In contrast to other IT systems, industrial systems often do not only require classical properties like data confidentiality or authentication of the communication, but have special needs due to their interaction with the physical world. For example, the reordering or deletion of some commands sent to a machine can cause the system to enter an unsafe state with potentially catastrophic effects. To prevent such attacks, the integrity of the message flow is necessary.

In joint work with Lafourcade (Univ Clermont-Ferrand), Potet, and Puys (Univ Grenoble Alpes), Dreier developed a formal definition of Flow Integrity in the context of industrial systems. The framework is applied to two well-known industrial protocols: OPC-UA and MODBUS. Using *TAMARIN*, they identified several design flaws in some of the different versions of these protocols. They also discussed how to efficiently model counters and timestamps in *TAMARIN*, as they are key ingredients of the analyzed protocols. This work was presented at SECRYPT'17 [32], and won a Best Student Paper Award.

7.2. Analysis

7.2.1. 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). However, they often fail to analyse equivalence properties. Equivalence properties can express a variety of security properties, including in particular privacy properties (vote privacy, anonymity, untraceability). Several decision procedures have already been proposed but the resulting tools are often rather limited, and lack efficiency.

In the case of a passive adversary, Ringeissen, in collaboration with Marshall (Univ Mary Washington, USA) and Erbatur (LMU, Germany) present new combination techniques for the study of deducibility and static equivalence in unions of equational theories sharing constructors. This allows us to develop new modularity results for the decidability of deducibility and static equivalence. In turn, this should allow for the security analysis of protocols which previous disjoint combination methods could not address because their axiomatization corresponds to the union of non-disjoint equational theories. This work has been presented at CADE'17 [28].

In case of an active adversary, and a bounded number of sessions, we made several advances. The *Akiss* tool has been extended in two directions. Gazeau and Kremer, in collaboration with Baelde (LSV, ENS Cachan) and Delaune (IRISA) have extended the underlying theory and the *Akiss* tool with support for exclusive or. They analyse unlinkability in several RFID protocols and resistance to guessing attacks of several password-based protocols. This work has been presented at CSF'17 [18]. Gazeau and Kremer also extended the *Akiss* tool to analyse protocols with else branches. This is particularly useful when verifying equivalence properties, as one needs to model precisely the error messages sent out when tests fail. While ignoring these branches may often be safe when studying trace properties this is not the case for equivalence properties, as for instance witnessed by an attack on the European electronic passport. One appealing feature of our approach is that our extension re-uses the saturation procedure which is at the heart of the verification procedure of *Akiss* as a black box, without need to modify it. As a result we obtain the first tool that is able verify equivalence properties for protocols that may use xor and else branches. We demonstrate the tool's effectiveness on several case studies, including the AKA protocol deployed in mobile telephony. This result was presented at ESORICS'17 [29]. Cortier and Dallon, in collaboration with Delaune (IRISA) propose a novel algorithm, based on graph planning and SAT-solving, which significantly improves the efficiency of the analysis of equivalence properties. The

[0] Stuxnet. https://en.wikipedia.org/wiki/Stuxnet

resulting implementation, SAT-Equiv, can analyze several sessions where most tools have to stop after one or two sessions. The approach has been presented at CSF'17 [20] for protocols with symmetric encryption and no else branches. Finally, Cheval, Kremer, and Rakotonirina have worked on complexity results for deciding equivalence properties and provide a decision procedure in the case of a bounded number of sessions. They showed that trace equivalence and labelled bisimilarity for a large variety of cryptographic primitives—those that can be represented by a subterm convergent destructor rewrite system— are both CoNEXP complete. Moreover, the procedure has been implemented in a new tool, *DeepSec*. Extensive experiments demonstrate that it is significantly more efficient than most other similar tools (being only slightly outperformed by SAT-Equiv in some specific examples), while at the same time raises the scope of the protocols that can be analysed. These results are currently under submission.

The previous results apply for a bounded number of sessions and may still be limited for a large number of sessions. In collaboration with Maffei and Grimm, Lallemand and Cortier have devised a novel approach [24] for proving equivalence properties. Instead of *deciding* equivalence, like for the previous approaches, they design a type system, sound w.r.t. equivalence. The resulting tool TypeEquiv can consider a bounded as well as an unbounded number of sessions, or a mix of both. It induces a significant speedup compared to previous tools for a bounded number of sessions and compares similarly to ProVerif for an unbounded number of sessions, with the advantage of a tighter treatment of bounded number of sessions. It can be applied to protocols with standard primitives and else branches.

7.2.2. Analysis of stateful security protocols

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

Many real-life protocols need to maintain a global state–such as counters, tables, or more generally, memory cells–that may be read and updated by parallel threads. Modelling such mutable, global state in protocols complicates the verification problem, in particular when analyzing an unbounded number of sessions.

The *SAPIC/TAMARIN* toolchain is one of the few tools that was designed to handle such global state. Dreier, Duménil (former intern in Pesto) and Kremer, in collaboration with Sasse (ETH Zurich, Switzerland) improve the underlying theory and the *TAMARIN* tool to allow for more general user-specified equational theories: the extension supports arbitrary convergent equational theories that have the finite variant property, making *TAMARIN* the first tool to support at the same time this large set of user-defined equational theories, protocols with global mutable state, an unbounded number of sessions, and complex security properties. The effectiveness of this generalization is demonstrated by analyzing several protocols that rely on blind signatures, trapdoor commitment schemes, and ciphertext prefixes that were previously out of scope. This work has been presented at POST'17 [27].

ProVerif is a very popular tool for the analysis of security protocols, that works very well in practice. However, in the case of protocols with global states, ProVerif typically fails in its analysis, due to its internal abstraction. Instead of designing a new ad-hoc procedure, we devise a generic transformation of the security properties queried to ProVerif. We prove the soundness of our transformation and implement it into a front-end GSVerif. Our experiments show that our front-end (combined with ProVerif) outperforms the few existing tools, both in terms of efficiency and protocol coverage. We successfully apply our tool to a dozen of protocols of the literature including a deployed voting and a payment protocol. This work is under submission.

7.2.3. Analysis of e-voting protocols

Participants: Véronique Cortier, Constantin-Catalin Dragan, Mathieu Turuani.

Cortier and Dragan provide the first machine-checked proof of privacy-related properties (including ballot privacy) for an electronic voting protocol in the computational model. They target the popular Helios family of voting protocols, for which they identify appropriate levels of abstractions to allow for simplification and convenient reuse of proof steps across many variations of the voting scheme. The resulting framework enables machine-checked security proofs for several hundred variants of Helios and should serve as a stepping stone for the analysis of further variations of the scheme. In addition, they highlight some of the lessons learned regarding the gap between pen-and-paper and machine-checked proofs, and report on the experience with formalizing the security of protocols at this scale. This work has been presented at S&P'17 [21].

Turuani and Cortier, in collaboration with Galindo (Univ Birmingham), have analysed the e-voting protocol developed by Scytl and planned to be deployed in Switzerland. The formal analysis of both privacy and individual verifiability has been conducted in ProVerif. It required to develop 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 accepted at EuroSP'18 [23].

Norway used e-voting in its last political election both in September 2011 and September 2013. The underlying protocol was also developed by Scytl. Cortier, in collaboration with Wiedling, has conducted a formal analysis (by hand) of vote privacy of this protocol, considering several corruption scenarios [13].

7.2.4. Unification in Forward-Closed Theories

Participant: Christophe Ringeissen.

In collaboration with Marshall (Univ Mary Washington, USA) and Erbatur (LMU, Germany), we investigate the unification problem in equational theories involving forward-closed convergent term rewrite systems. In the class of forward-closed theories, unification is decidable and finitary since a convergent term rewrite system has a finite forward-closure if and only if it has the finite variant property. Actually, forward-closed theories are syntactic theories admitting a terminating mutation-based unification procedure. This can be shown by reusing a mutation-based unification algorithm originally developed for equational theories saturated by paramodulation, since a forward-closed theory is indeed a sufficient condition to get soundness and completeness. Building on this fact we develop a new mutation-based unification algorithm which is simpler, with regard to conflicts and number of rules, than the first algorithm. We then use this simplified algorithm as a component to develop a new method that solves the unification problem in unions of forward-closed theories with non-disjoint theories. The resulting algorithm can be viewed as a terminating instance of a procedure initiated for hierarchical combination. This work has been presented at the workshop UNIF'17 [33].

7.2.5. Analysis of Combinations of Protocols

Participant: Jannik Dreier.

When trying to prove the security of a protocol, one usually analyzes the protocol in isolation, i.e., in a network with no other protocols. But in reality, there will be many protocols operating on the same network, maybe even sharing data including keys, and an intruder may use messages of one protocol to break another. We call that a multi-protocol attack. In this work, we tried to find such attacks using the *TAMARIN* prover. We analyzed both examples that were previously analyzed by hand or using other tools, and found novel attacks. This work was presented at FPS'17 [31].

7.3. Design

7.3.1. E-voting protocols

Participants: Véronique Cortier, Alicia Filipiak.

Building upon a recently proposed voting scheme, BeleniosRF, we design a new voting scheme that ensures both verifiability and privacy against a compromised voting machine, as well as a compromised voting server. It assumes that the voter has two devices: one computer for casting a vote and another device (typically a smartphone or a tablet) to, optionally, audit the material (a voting sheet) sent to the voter. Neither the computer nor the smartphone learns how the voter voted unless they collude. The resulting protocol has been formally analysed in ProVerif w.r.t. both verifiability and privacy. Analysing verifiability in ProVerif cannot be done directly as it would require counting. Instead, we propose a set of properties that can be handled by ProVerif and that entail verifiability. This work is one of the contribution of the thesis manuscript of Alicia Filipiak and will be submitted.

7.3.2. Designing and proving an EMV-compliant payment protocol for mobile devices

Participants: Véronique Cortier, Alicia Filipiak.

In collaboration with Gharout, Traoré and Florent (Orange Labs), we devised a payment protocol that can be securely used on mobile devices, even infected by malicious applications. Our protocol only requires a light use of Secure Elements, which significantly simplifies certification procedures and protocol maintenance. It is also fully compatible with the EMV-SDA protocol and allows off-line payments for the users. We provide a formal model and full security proofs of the protocol using the *TAMARIN* prover. This work has been presented at EuroS&P'17 [22].

7.3.3. Composition and design of PKIs

Participants: Vincent Cheval, Véronique Cortier.

In protocol analysis one makes the (strong) assumption that honestly generated keys are available to all parties and that the link between identities and public keys is fixed and known to everyone. The abstraction is grounded in solid intuition but there are currently no theoretical underpinnings to justify its use. Cheval and Cortier, in collaboration with Warinschi (Univ Bristol, UK), initiate a rigorous study of how to use PKIs within other protocols, securely. They first show that the abstraction outlined above is in general unsound by exhibiting a simple protocol which is secure with idealized key distribution but fails in the presence of more realistic PKI instantiation. Their main result is a generic composition theorem that identifies under which conditions protocols that require public keys can safely use any PKI protocol (which satisfies a security notion which we identify). Interestingly, unlike most existing composition results in symbolic models they do not require full tagging of the composed protocols. Furthermore, the results confirm the recommended practice that keys used in the PKI should not be used for any other cryptographic task. This work has been presented at CSF'17 [19].

7.3.4. Privacy Protection in Social Networks

Participants: Younes Abid, Hector Dang-Nhu, Andrii Dychka, Abdessamad Imine, Michaël Rusinowitch, Valentin Salquebre.

In order to demonstrate privacy threats in social networks we show how to infer user preferences by random walks in a multiple graph representing simultaneously attributes and relationships links. For the approach to scale in a first phase we reduce the space of attribute values by partition in balanced homogeneous clusters. Following the Deepwalk approach, the random walks are considered as sentences. Hence unsupervised learning techniques from natural languages processing can be employed in a second phase to deduce semantic similarities of some attributes. We conduct initial experiments on real datasets to evaluate our approach. This work was presented at DEXA'17 [15].

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

Participants: Haftay Gebreslasie Abreha, Michaël Rusinowitch.

In a joint project with EPI Madynes and Cynapsys, we are starting to work on the design, implementation and evaluation of multi-masked techniques for building a compressed and a verifiable filtering rules in Software Defined Networks with the possibility of distributing the workload processing among several packet filtering devices operating in parallel.

VERIDIS Project-Team

7. New Results

7.1. Automated and Interactive Theorem Proving

Participants: Haniel Barbosa, Jasmin Christian Blanchette, Martin Bromberger, Simon Cruanes, Daniel El Ouraoui, Mathias Fleury, Pascal Fontaine, Stephan Merz, Martin Riener, Hans-Jörg Schurr, Martin Strecker, Thomas Sturm, Andreas Teucke, Sophie Tourret, Marco Voigt, Tung Vu Xuan, Uwe Waldmann, Daniel Wand, Christoph Weidenbach.

7.1.1. IsaFoL: Isabelle Formalization of Logic

Joint work with Andreas Halkjær From (DTU Copenhagen), Alexander Birch Jensen (DTU Copenhagen), Maximilian Kirchmeier (TU München), Peter Lammich (TU München), John Bruntse Larsen (DTU Copenhagen), Julius Michaelis (TU München), Tobias Nipkow (TU München), Nicolas Peltier (IMAG Grenoble) Anders Schlichtkrull (DTU Copenhagen), Dmitriy Traytel (ETH Zürich), and Jørgen Villadsen (DTU Copenhagen).

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. In particular, we are formalizing large parts of Weidenbach's forthcoming textbook, tentatively called *Automated Reasoning—The Art of Generic Problem Solving*.

The objective of formalization work is not to eliminate paper proofs, but to complement them with rich formal companions. Formalizations help catch mistakes, whether superficial or deep, in specifications and theorems; they make it easy to experiment with changes or variants of concepts; and they help clarify concepts left vague on paper.

The repository contains 14 completed entries and four entries that are still in development. Notably, Mathias Fleury formalized a SAT solver framework with learn, forget, restart, and incrementality. This year he extended it with key optimizations such as the two-watched-literal procedure. The corresponding paper, written together with Jasmin Blanchette and Peter Lammich, was accepted at a highly competitive conference (CPP 2018).

7.1.2. Extension of Term Orders to λ-Free Higher-Order Logic

Superposition is one of the most successful proof calculi for first-order logic today, but in contrast to resolution, tableaux, and connections, it has not yet been generalized to higher-order logic (also called simple type theory). Yet, most proof assistants and many specification languages are based on some variant of higher-order logic.

This motivates us to design a *graceful* generalization of superposition: a proof calculus that behaves like standard superposition on first-order problems and that smoothly scales up to arbitrary higher-order problems. A challenge is that superposition relies on a simplification order, which is fixed in advance of the proof attempt, to prune the search space.

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

We started our investigations by focusing on a fragment devoid of λ-abstractions, but with partial application and application of variables, two crucial higher-order features. We generalized the two main orders that are used in superposition-based provers today—the lexicographic path order (LPO) [27] and the Knuth-Bendix order (KBO) [21]. The new orders gracefully generalize their first-order counterparts and enjoy nearly all properties needed for superpositions. An exception is compatibility with contexts, which is missing for LPO and some KBO variants. Preliminary work suggests that we can define a version of the superposition calculus that works well in theory and practice (i.e., is refutationally complete and does not lead to a search-space explosion) despite the missing property.

7.1.3. A Fine-Grained Approach of Understanding First-Order Logic Complexity

By the introduction of the separated fragment [65] we have initiated a new framework for a fine-grained understanding of the complexity of fragments of first-order logic, with and without the addition of theories. We have related the classes of the polynomial hierarchy to subclasses of the separated fragment [40] and developed new decidability results [36], [41] based on the techniques of our framework for the combination of the Bernays-Schoenfinkel subfragment with linear arithmetic.

7.1.4. Theorem Proving Based on Approximation-Refinement into the Monadic Shallow Linear Fragment with Straight Dismatching Constraints

We have introduced an approximation-refinement approach for first-order theorem proving based on counterexample-guided abstraction refinement [39]. A given first-order clause set is transformed into an over-approximation contained in the fragment of monadic, shallow, linear clauses with straight dismatching constraints. We have shown the fragment to be decidable, strictly extending known results. If the abstraction obtained that way is satisfiable, so is the original clause set. However, if it is unsatisfiable, then the approximation provides a terminology for lifting the found refutation, step by step, into a proof for the original clause set. If lifting fails, the cause is analyzed to refine the original clause set such that the found refutation is ruled out for the future, and the procedure repeats. We have shown that this approach is superior to all known calculi on certain classes of first-order clauses. In particular, it is able to detect satisfiability of clause sets that have only infinite models.

7.1.5. Combination of Satisfiability Procedures

Joint work with Christophe Ringeissen from the PESTO project-team of Inria Nancy – Grand Est, and Paula Chocron at IIIA-CSIC, Bellaterra, Catalonia, 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 (to ensure the existence of an infinite model). 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 [55] a sound and complete combination procedure à la Nelson-Oppen for the theory of absolutely free data structures (including lists and trees) connected to another theory via bridging functions. This combination procedure has also been refined for standard interpretations. The resulting theory has a nice politeness property, enabling combinations with arbitrary decidable theories of elements. We also investigated other theories [56] 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 2017, we have been improving the framework and unified both results. A new paper is in preparation.

7.1.6. Quantifier Handling in SMT

Joint work with Andrew J. Reynolds, Univ. of Iowa, USA.

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 [20]. In later, unpublished work, we are revisiting enumerative instantiation for SMT. This effort takes place in the context of the Matryoshka project.

7.1.7. Non-Linear Arithmetic in SMT

In the context of the SMArT ANR-DFG (Satisfiability Modulo Arithmetic Theories), KANASA and SC^2 projects (cf. sections 9.1 and 9.3), we study the theory, design techniques, and implement software to push forward the non-linear arithmetic (NLA) reasoning capabilities in SMT. This year, 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 SMT-COMP 2017. We also studied integration of these procedures into combinations of theories. The ideas are validated within the veriT solver, together with code from the raSAT solver (from JAIST). An article is in preparation.

We also adapted the subtropical method to use in an SMT context, with valuable results. This was the subject of a publication in 2017 [33].

7.1.8. Proofs for SMT

We have 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 publication in [19]. This effort takes place in the context of the Matryoshka project.

7.1.9. Coding Modal and Description Logics in SAT solvers

The application scenario behind this research is the verification of graph transformations, which themselves are relevant for a wide range of practical problems such as pointer structures in imperative programs, graph databases or access control mechanisms.

Graph structures can typically be perceived as models of modal logics, and modal logics and variants (such as description logics that are the basis for the web ontology language OWL) are in principle suitable specification formalisms for graph transformations. It turns out, however, that pure modal logics are often not sufficiently expressive for the intended verification purpose and that extensions are needed for which traditional proof methods such as tableau calculi become complex: the termination of the calculi are often very difficult to prove, and huge efforts are required to obtain an efficient implementation.

For these reasons, we have explored methods of encoding the above-mentioned logics in SAT and SMT solvers such as CVC4 and veriT. The idea is to traverse the formula to be verified in order to span up a pre-model that possibly contains more elements (worlds in a Kripke structure) than the real model, and then to run a solver to find out which of these elements can effectively be realized. A prototype has been implemented, with encouraging results. It remains to connect this prototype to the graph verification engine and to publish this work.

7.1.10. Work on the TLA+ Proof System

We continued our work on encoding set-theoretic formulas in multi-sorted first-order logic, and in particular for SMT solvers. Specifically, we unified and streamlined a technique combining an injection of unsorted expressions into sorted languages, simplification by rewriting, and abstraction that underlies the SMT backend of the TLA$^+$ proof system TLAPS. A presentation of our technique was accepted in the journal *Science of Computer Programming*, to appear in 2018.

The proof of the join protocol in a pure-join variant of the Pastry protocol [63] implementing a distributed hash table over a peer-to-peer network is the largest case study carried out so far within TLAPS. Consisting of roughly 30k lines of proof, it was developed as part of Noran Azmy's PhD thesis, defended at the end of 2016 [51]. A presentation of the design of the protocol and its proof was accepted in the journal *Science of Computer Programming*, to appear in 2018.

7.1.11. Automated Analysis of Systems of ODE for Multistationarity

Joint work with R. Bradford and J. Davenport (Bath, UK), M. England (Coventry, UK), H. Errami, C. Hoyt, and A. Weber (Bonn, Germany), V. Gerdt (Dubna, Russia), D. Grigoriev (Lille, France), O. Radulescu (Montpellier, France)

We considered the problem of determining multiple steady states for positive real values in models of biological networks. Investigating the potential for these in models of the mitogen-activated protein kinases (MAPK) network has consumed considerable effort using special insights into the structure of corresponding models. We have applied combinations of symbolic computation methods for mixed equality/inequality systems, specifically automated deduction methods like virtual substitution, lazy real triangularization and cylindrical algebraic decomposition. We have determined multistationarity of an 11-dimensional MAPK network when numeric values are known for all but potentially one parameter. More precisely, our considered model has 11 equations in 11 variables and 19 parameters, 3 of which are of interest for symbolic treatment, and furthermore positivity conditions on all variables and parameters [28].

Subsequent work [31] demonstrates that our techniques benefit tremendously from a new graph theoretical symbolic preprocessing method. We apply our combined techniques to visualize of parameter regions for multistationarity. Comparing computation times and quality of results it turns out that our automated deduction-based approach clearly outperforms established numerical continuation methods.

While automated deduction technology is a bit under the hood here, this interdisciplinary research line addresses important questions related to contemporary research in systems biology. With researchers from that area very actively involved, the results are recognized also within their communities.

7.2. Formal Methods for Developing and Analyzing Algorithms and Systems

Participants: Marie Duflot-Kremer, Margaux Duroeulx, Souad Kherroubi, Poonam Kumari, Dominique Méry, Stephan Merz, Nicolas Schnepf, Christoph Weidenbach.

7.2.1. Making Explicit Domain Knowledge in Formal System Development

Joint work with partners of the IMPEX project.

As explained in the description of the IMPEX project in section 9.1 , we advocate 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. Our objective in doing so is to offer rigorous mechanisms for handling domain knowledge in design models.

We developed the notion of dependency for state-based models. Context-awareness is an important feature in system design. We argue that in proof systems and conceptual modelling this notion should be highlighted precisely. Since we focus on conceptual modelling, understandability and clarity are of high importance. We introduce a new definition [37] for proof context in state-based formalisms with an application to the Event-B modeling language. Furthermore, we introduce a dependency relation between two Event-B models. The contextualization of Event-B models is based on knowledge provided from domains that we classified into

constraints, hypotheses and dependencies. The dependency mechanism between two models makes it possible to structure the development of systems models, by organizing phases identified in the analyzed process. These ideas are inspired by work based on the modelling of situations in situation theory that emphasize capabilities of type theory with regard to situation modelling to represent knowledge. Our approach is illustrated on small case studies, and was validated on a development of design patterns for voting protocols.

7.2.2. 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 on the conformance of each refinement with the model preceding it.

Our main result during 2017 is the development of a proved-based pattern for integrating the local computation models and the Visidia platform [32].

7.2.3. Modeling Network Flows in View of Building Security Chains

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

We are working on the application of formal modeling and verification techniques in the area of network communications, and in particular for constructing security functions in a setting of software-defined networks (SDN). Concretely, Nicolas Schnepf defined an extension of the Pyretic language [58] taking into account 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. This work was published at NetSoft 2017 [38].

Extending this approach, we have worked on inferring probabilistic finite-state automata models that represent network flows generated by Android applications. The objective is to exploit this representation for generating security chains that detect significant deviations from the behavior represented by the automata and can initiate protective actions. Comparing our models with automata produced by the state-of-the-art tools Invarimint and Synoptic, we obtain representations that are as succinct as those inferred by Invarimint, and significantly smaller than Synoptic, but that include information about transition probability, which Invarimint does not. This work was accepted for publication at NOMS 2018.

7.2.4. 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 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 [30], we develop three different algorithms for computing minimal tie sets (i.e., component configurations that ensure that the system is functioning). Our algorithms are based on either conjunctive or disjunctive normal form representations of the structure function or on the Hasse diagram representing the configurations. These algorithms have been prototypically implemented in Python, and we are evaluating them on existing benchmarks in order to understand which algorithm works best for typical fault dependencies.

7.2.5. Statistical evaluation of the robustness of production schedules

Joint work with Alexis Aubry, Sara Himmiche, Pascale Marangé, and Jean-François Pétin at Centre de Recherche en Automatique de Nancy.

Finding a good schedule for a production system, especially when it is flexible and when several machines can perform the same operation on products, is a challenging and interesting problem. For a long time, operations research has provided state-of-the-art methods for optimizing scheduling problems. However, approaches based on Discrete Event Systems present interesting alternatives, especially when dealing with uncertainties on the demand or the production time. In this particular case, the flexibility of the automata-based modeling approach is really useful. Using probabilistic timed automata, we demonstrated [35] that statistical model checking can be used successfully for evaluating the robustness of a given schedule w.r.t. probabilistic variations of the processing time. We were thus able to compare different schedules based on their level of service (i.e., the probability that the system will complete the production process within a deadline slightly higher that the schedule time) and their sensitivity (the minimal deadline for which the level of service is greater than a given threshold) [42].

An interdisciplinary workshop on this topic was organized jointly with our colleagues of Centre de Recherche en Automatique and funded by Fédération Charles Hermite.

7.2.6. Using Cubicle for Verifying TLA+ Specifications

Cubicle [0] is a model checker for the verification of parameterized transition systems whose state is described by arrays of variables indexed by an abstract sort representing processes. During her internship, Poonam Kumari designed a translation algorithm from a restricted class of TLA$^+$ specifications into the input language of Cubicle. A prototypical implementation demonstrates the feasibility of the approach, although more work will be necessary to widen the scope of the translation. This work will be continued within the PARDI project, described in section 9.1 .

[0] http://cubicle.lri.fr

SPHINX Project-Team

7. New Results

7.1. Control and stabilization of heterogeneous systems

7.1.1. Analysis of heterogeneous systems

Participants: Jean-François Scheid, Takéo Takahashi.

In [12], we consider a single disk moving under the influence of a 2D viscous fluid and study the asymptotic as the size of the solid tends to zero. If the density of the solid is independent of the size, the energy equality is not sufficient to obtain a uniform estimate for the solid velocity. This is achieved thanks to the optimal $L^p - L^q$ decay estimates of the semigroup associated to the fluid-rigid body system and to a fixed point argument.

In [10], we propose a new model for the motion of a viscous incompressible fluid. More precisely, we consider the Navier-Stokes system with a boundary condition governed by the Coulomb friction law. With this boundary condition, the fluid can slip on the boundary if the tangential component of the stress tensor is too large. We prove the existence and uniqueness of a weak solution in the two-dimensional problem and the existence of at least one solution in the three-dimensional case. In [9], we consider this model with a rigid body. We prove that there exists a weak solution for the corresponding system.

In [13], we study a free boundary problem modeling the motion of a piston in a viscous gas. The gas-piston system fills a cylinder with fixed extremities, which possibly allow gas from the exterior to penetrate inside the cylinder. The gas is modeled by the 1D compressible Navier-Stokes system and the piston motion is described by the second Newton's law. We prove the existence and uniqueness of global in time strong solutions. The main novelty is that we include the case of non homogeneous boundary conditions.

In [31], we study the shape differentiability of the free-boundary 1-dimensional simplified model for a fluid-elasticity system. The full characterization of the associated material derivatives is given and the shape derivative of an energy functional has been obtained.

7.1.2. Control of heterogeneous systems

Participants: Thomas Chambrion, Alessandro Duca, Takéo Takahashi.

In [11], we consider the swimming into a stationary Navier-Stokes fluid. The swimmer is a rigid body $\mathcal{S} \subset \mathbb{R}^3$ immersed in an infinitely extended fluid. We are interested in self-propelled motions of $\mathcal{S}$ in the steady state regime of the rigid body-fluid system, assuming that the mechanism used by the body to reach such a motion is modeled through a distribution of velocities on the boundary. We show that this can be solved as a control problem.

In [16] we prove that the Kuramoto-Sivashinsky equation is locally controllable in 1D and in 2D with one boundary control. His method consists in combining several general results in order to reduce the null-controllability of this nonlinear parabolic equation to the exact controllability of a linear beam or plate system. This improves known results on the controllability of Kuramoto-Sivashinsky equation and gives a general strategy to handle the null-controllability of nonlinear parabolic systems.

The paper [21] is the result of a long term analysis about the restrictions to the controllability of bilinear systems induced by the regularity of the propagators for the bilinear Schrödinger equation. This paper comes along with its companion paper [20] which gives a detailed proof of the celebrated Ball-Marsden-Slemrod obstruction to exact controllability for bilinear systems with L^1 controls.

The paper [23] is concerned with the one dimensional bilinear Schrödinger equation in a bounded domain. In this article, we have given the first available upper bound estimates of the time needed to steer exactly the infinite square potential well from its first eigenstate to the second one.

In [22], we present an embedded automatic strategy for the control of a low consumption vehicle equipped with an "on/off" engine. The proposed strategy has been successfully implemented on the Vir'Volt prototype in official competition (European Shell Eco Marathon).

7.1.3. Stabilization of heterogeneous systems

Participants: David Dos Santos Ferreira, Takéo Takahashi, Julie Valein, Jean-Claude Vivalda.

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 [7], we study the feedback stabilization of a system composed by an incompressible viscous fluid and a deformable structure located at the boundary of the fluid domain. We stabilize the position and the velocity of the structure and the velocity of the fluid around a stationary state by means of a Dirichlet control, localized on the exterior boundary of the fluid domain and with values in a finite dimensional space.

In [19], we study the nonlinear Korteweg-de Vries equation with boundary time-delay feedback. Under appropriate assumption on the coefficients of the feedbacks, we first prove that this nonlinear infinite dimensional system is well-posed for small initial data. The main results of our study are two theorems stating the exponential stability of the nonlinear time delay system, using two different methods: a Lyapunov functional approach and an observability inequality approach.

In [14], we generalize a formula, due to E. Sontag *et al.*, giving explicitly a continuous stabilizing feedback for systems affine in the control; more specifically for a large class of systems which depend quadratically on the control, an explicit formula for a stabilizing feedback law is given.

7.2. Inverse problems for heterogeneous systems

7.2.1. Reconstruction of coefficients and initial conditions

Participants: Karim Ramdani, Julie Valein, Jean-Claude Vivalda.

In [79], we 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. In [29], we 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 [18], 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.

7.2.2. Geometrical inverse problems

Participants: Alexandre Munnier, Karim Ramdani, Takéo Takahashi.

In [75], we 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 [28], we 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 [15], we consider the geometrical inverse problem consisting in recovering an unknown obstacle in a viscous incompressible fluid by measurements of the Cauchy force on the exterior boundary. We deal with the case where the fluid equations are the nonstationary Stokes system and using the enclosure method, we can recover the convex hull of the obstacle and the distance from a point to the obstacle. With the same method, we can obtain the same result in the case of a linear fluid-structure system composed by a rigid body and a viscous incompressible fluid.

7.3. Numerical analysis and simulation of heterogeneous systems

Participants: Xavier Antoine, Qinglin Tang.

In [1], we propose a simple accelerated pseudo-spectral algorithm to compute the stationary states of the Gross-Piteavskii Equation (GPE) with possibly multiple components. The method is based on the adaptation of new optimization algorithms under constraints coming from mathematical imaging to the imaginary time (gradient-like) method for the GPE arising in Bose-Einstein Condensation.

In [3] we propose original efficient preconditioned conjugate gradient methods coming from molecular physics to the GPE for spectrally computing the stationary states of the GPE. The method allows a gain of a factor 100 for 3D problems with extremely large nonlinearities and fast rotations. The HPC solver is being developed.

In [17], we develop new robust and efficient algorithms for computing the dynamics of 2-components GPEs with dipolar interaction. The main particularity of the method is that high accuracy is obtained by a new FFT based evaluation of nonlocal kernels applied to the nonlinear part of the operator.

In [4], we propose an asymptotic mathematical analysis of domain decomposition techniques for solving the 1D nonlinear Schrödinger equation and GPE. The analysis uses advanced techniques related to fractional microlocal analysis for PDEs. Simulations confirm the mathematical analysis.

In [2], we extend, by some very technical mathematical analysis, approaches for the results stated in [4]. Again, numerical simulations validate the theoretical analysis.

In [5], we develop and implement in parallel simple new solvers for computing the dynamics of solutions to the Dirac equation arising in quantum physics. Numerical examples are developed to analyze the capacity of these algorithms for a parallel implementation.

In [6], we introduce the concept of Absorbing Boundary Conditions and Perfectly Matched Layers for the dynamics of nonlinear problems related to classical and relativistic quantum wave problems (Wave equation, Schrödinger equation, Dirac eqution). In particular, we show application examples and detail the methods so that they can be implemented by researchers coming for quantum physics.

TOSCA Project-Team

7. New Results

7.1. 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, Paolo Pigato, Denis Talay, Etienne Tanré, Milica Tomašević, Denis Villemonais.

7.1.1. Published works and preprints

- M. Bossy, R. Maftei and Jean-François Jabir (National Research University Higher School of Economics, Moscow) propose and analyze the convergence of a time-discretization scheme for the motion of a particle when its instantaneous velocity is drifted by the known velocity of the carrying flow, and when the motion is taking into account the collision event with a boundary wall. We propose a symetrized version of the Euler scheme and prove a convergence of order one for the weak error. The regularity analysis of the associated Kolmogorov PDE is obtained by mixed variational and stochastic flow techniques for PDE problem with specular condition [46].
- N. Champagnat and B. Henry (IECL) 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 [48].
- N. Champagnat and P.-E. Jabin (Univ. Maryland) completed the study of the functional spaces in the article [18], devoted to the study of strong existence and pathwise uniqueness for stochastic differential equations (SDE) with rough coefficients, typically in Sobolev spaces.
- N. Champagnat and D. Villemonais consider, for general absorbed Markov processes, the notion of quasi-stationary distributions (QSD), which is a stationary distribution conditionally on non-absorbtion, and the associated Q-process, defined as the original Markov process conditioned to never be absorbed. They prove that, under the conditions of [5], in addition to the uniform exponential convergence of the conditional distributions to a unique QSD and the uniform exponential ergodicity of the Q-process, one also has the uniform convergence of the law of the process contionned to survival up to time T, when $T \to +\infty$. This allows them to obtain conditional ergodic theorems [22].
- N. Champagnat and D. Villemonais obtained criteria based on Lyapunov functions allowing to check the conditions of [5] which characterize the exponential uniform convergence in total variation of conditional distributions of an absorbed Markov process to a unique quasi-stationary distribution [50]. Among the various applications they give, they prove that these conditions apply to any logistic Feller diffusions in any dimension conditionned to the non extinction of all its coordinates. This question was left partly open since the first work of Cattiaux and Méléard on this topic [63].
- N. Champagnat and D. Villemonais obtained general conditions based on Foster-Lyapunov criteria ensuring the exponential convergence in total variation of the conditional distributions of an absorbed Markov process to a quasi-stationary distribution (QSD), with a speed that can depend on the initial distribution. In particular, these results provide a non-trivial subset of the domain of attraction of the minimal QSD of an absorbed process in cases where there is not uniqueness of a QSD. Similar results were only known for the very specific branching models. They also show how these criteria can be checked for a wide range of Markov processes in discrete or continuous time and in discrete

or continuous state spaces. In all these cases, they improve significantly the best known results. A particularly remarkable result is the existence of a principal eigenfunction for the generator of elliptic diffusion processes aborbed at the boundary of an open domain without any regularity assumption on the boundary of this domain [49].

- During his internship supervised by E. Tanré and R. Veltz (MATHNEURO Inria team), Q. Cormier studied numerically and theoretically a model of spiking neurons in interactions [51]. This model generalizes classical integrate and fire models: the neurons no more spike after hitting a deterministic threshold but spikes with a rate given as a function of the membrane potential (see e.g. [64]). He showed existence and uniqueness of the corresponding limit equation, and was able to extend those results in the case of excitatory and inhibitory neurons. He is now studying the long time behavior of the model, as part of his thesis.
- M. Deaconu and S. Herrmann studied the simulation of the hitting time of some given boundaries for Bessel processes. These problems are of great interest in many application fields as finance and neurosciences. More precisely they obtained recently a new method for the simulation of hitting times for Bessel processes with a non integer dimension. The main idea is to consider the simulation of the hitting time of Bessel processes with integer dimension and provide a new algorithm by using the additivity property of the laws of squared Bessel processes [26].
- M. Deaconu and S. Herrmann studied the Initial-Boundary Value Problem for the heat equation [25]. They construct an algorithm based on a random walk on heat balls in order to approximate the solution. Even if it represents a sophisticated generalization of the Walk on Spheres (WOS) algorithm introduced to solve the Dirichlet problem for Laplace's equation, its implementation is rather easy. The definition of the random walk is based on a particular mean value formula for the heat equation and they obtained also a probabilistic formulation of this formula. They proved convergence results for this algorithm and illustrate them by numerical examples.
- M. Deaconu, S. Herrmann and S. Maire [27] introduced a new method for the simulation of the exit time and position of a δ-dimensional Brownian motion from a domain. The main interest of this method is that it avoids splitting time schemes as well as inversion of complicated series. The idea is to use the connexion between the δ-dimensional Bessel process and the δ-dimensional Brownian motion thanks to an explicit Bessel hitting time distribution associated with a particular curved boundary. This allows to build a fast and accurate numerical scheme for approximating the hitting time.
- M. Deaconu, B. Dumortier and E. Vincent (EPI Multispeech) are working with the Venathec SAS on the acoustic control of wind farms. Wind turbine noise is often annoying for humans living in close proximity to a wind farm. Reliably estimating the intensity of wind turbine noise is a necessary step towards quantifying and reducing annoyance, but it is challenging because of the overlap with background noise sources. Current approaches involve measurements with on/off turbine cycles and acoustic simulations, which are expensive and unreliable. This raises the problem of separating the noise of wind turbines from that of background noise sources and coping with the uncertainties associated with the source separation output. In their work they propose to assist a black-box source separation system with a model of wind turbine noise emission and propagation in a recursive Bayesian estimation framework. This new approach is validated on real data with simulated uncertainties using different nonlinear Kalman filters [38].
- M. Deaconu is working with L. Beznea and O. Lupaşcu (Bucharest, Romania) on the stochastic interpretation of rupture phenomena. They constructed a stochastic differential equation and a branching process for the fragmentation model. The main physical model involved in their study is the avalanche one and their model includes physical properties of the phenomenon. They introduced a new numerical algorithm issued from this study, which captures the fractal property of the avalanche [43].
- C. Fritsch, F. Campillo (Inria Sophia-Antipolis, MATHNEURO team) and O. Ovaskainen (Univ. Helsinki) proposed a numerical approach to determine mutant invasion fitness and evolutionary

singular strategies using branching processes and integro-differential models in [31]. They illustrate this method with a mass-structured individual-based chemostat model.

- P. Helson, E. Tanré and R. Veltz (MATHNEURO Inria team), have numerically and theoretically studied a model of spiking neurons in interaction with stochastic plasticity. A slow-fast analysis enabled to split the dynamic in two inhomogeneous Markov chains: one models the slow variable, the other one the fast variable. The jump rates of the slow chain is governed by the invariant distribution of the fast one. In his PhD thesis, P. Helson has proved existence and uniqueness of solution. Simple conditions for the slow variable to be recurrent and transient are given [53].
- A. Lejay, L. Lenôtre (CMAP, École Polytechnique) and G. Pichot (Inria Paris, SERENA team) have continued their work on the simulation of processes on discontinuous media [55]. A new Monte Carlo scheme, called the exponential timestepping scheme and based on closed form expression of the resolvent, is being studied.
- A. Lejay, E. Mordecki (U. de la República, Uruguay) and S. Torres (U. de Valparaíso, Chile) have continuous their work on the estimation of the parameter of the Skew Brownian motion [56].
- A. Lejay and P. Pigato have studied the estimation of the parameter of the Oscillating Brownian motion, which is a solution of a stochatic differential equation whose diffusivity takes two values [35].
- A. Lejay have given an alternative proof of the Girsanov theorem which is based on semigroups [39].
- In [60] D. Talay and M. Tomašević propose a new type of stochastic interpretation of the 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. They prove that the one-dimensional parabolic-parabolic Keller-Segel system in the whole Euclidean space and the corresponding McKean-Vlasov stochastic differential equation are well-posed for any values of the parameters of the model.
- In collaboration with Jean-François Jabir (National Research University Higher School of Economics, Moscow) D. Talay and M. Tomašević prove the well–posedness of an original singularly interacting stochastic particle system associated to the one-dimensional parabolic-parabolic Keller-Segel model. They also establish the propagation of chaos towards this model [54].
- In [44] J. Bion-Nadal (Ecole Polytechnique) and D. Talay have introduced 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. They proved that it may also be defined by means of the value function of a stochastic control problem whose Hamilton–Jacobi–Bellman equation has a smooth solution, which allows one to deduce a priori estimates or to obtain numerical evaluations. They have exhibited an optimal coupling measure and characterized it as a weak solution to an explicit stochastic differential equation, and they finally have described procedures to approximate this optimal coupling measure.

  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?
- 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 (paper in revision: [59]).
- D. Villemonais worked with his Research Project student William Oçafrain (École des Mines de Nancy) on an original mean-field particle system [36]. They proved that the mean-field particle system converges in full generality toward the distribution of a conditioned Markov process, with applications to the approximation of the quasi-stationary distribution of piecewise deterministic Markov processes.

- D. Villemonais, Camille Coron (Université Paris XI) and Sylvie Méléard (École Polytechnique) proved a criterion for the integrability of paths of one-dimensional diffusion processes in [52] from which we derive new insights on allelic fixation in several situations.
- D. Villemonais obtains a lower bound for the coarse Ricci curvature of continuous time pure jump Markov processes in [61], with an emphasis on interacting particle systems. In this preprint, several models are studied, with a detailed study of the herd behavior of a simple model of interacting agents. The lower bound is shown to be sharp for birth and death processes.

7.1.2. Other works in progress

- M. Bossy, J. Fontbona (Universidad de Chile, Chile) and H. Olivero-Quinteros are working in a model for a network of neurons interacting electrically and chemically in a mean field fashion. They have proved the synchronization of the network under suitable values for the parameters of the model and a concentration result for the mean field limit.
- N. Champagnat is working with P. Vallois (IECL and Inria BIGS team) and L. Vallat (CHRU Strasbourg) on the inference of dynamical gene networks from RNAseq and proteome data.
- N. Champagnat, C. Fritsch and S. Billiard (Univ. Lille) are working on food web modeling.
- N. Champagnat, C. Fritsch and D. Villemonais are working with A. Gégout-Petit, P. Vallois, A. Mueller-Gueudin (IECL and Inria BIGS team), A. Kurtzmann (IECL), A. Harlé, J.-L. Merlin (ICL and CRAN) and E. Pencreac'h (CHRU Strasbourg) within an ITMO Cancer project on modeling and parametric estimation of dynamical models of circulating tumor DNA (ctDNA) of tumor cells, divided into resistant and sensitive ctDNA depending on whether they hold mutations known to provide resistance to a given targeted therapy or not. The goal of the project is to predict sooner and more accurately the emergence of resistance to the targeted therapy in a patient's tumor, so that the patient's therapy can be modulated more efficiently.
- M. Deaconu and S. Herrmann are working on numerical approaches for hitting times of some general stochastic differential equations.
- M. Deaconu, O. Lupaşcu and L. Beznea (Bucharest, Romania) are working on the connexion between branching processes and partial differential equations in fluid mechanics.
- M. Deaconu, B. Dumortier and E. Vincent (EPI Multispeech) are working on handling uncertainties in the wind farms model in order to design a stochastic algorithm.
- M. Deaconu and R. Stoica (Université de Lorraine, Nancy) are working on the ABC Shadow algorithm and its possible generalizations.
- O. Faugeras, E. Soret and E. Tanré are working on Mean-Field descriptions or thermodynamics limits of large populations of neurons. They study a system of EDS which describes the evolution of membrane potential of each neuron over the time when the synaptic weights are random variables (not assumed to be independent).
- O. Faugeras, James Maclaurin (Univ. of Utah) and E. Tanré have worked on the asymptotic behavior of a model of neurons in interaction with correlated gaussian synaptic weights. They have obtained the limit equation as a singular non-linear SDE and a Large Deviation Principle for the law of the finite network.
- C. Fritsch is working with A. Gégout-Petit (Univ. Lorraine and sc Bigs team), B. Marçais (INRA, Nancy) and M. Grosdidier (INRA, Nancy) on a statistical analysis of a Chalara fraxinea model.
- P. Helson, E. Tanré and R. Veltz (MATHNEURO Inria team) are working on a mathematical framework for plasticity models. The aim is to propose a 'optimized' model of memory capacity and memory lifetime.
- A. Lejay, A. Brault (Univ. Toulouse) and L. Coutin (Univ. Toulouse) are working on a non linear generalization of the sewing lemma, which is the main technical tool in the theory of rough paths.

- V. Martin Lac, H. Olivero-Quinteros and D. Talay are working on theoretical and algorithmic questions related to the simulation of large particle systems under singular interactions and to the simulation of independent random variables with heavy tails.
- 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 are working on a mean-field game and developing a new technique to analyse it.
- E. Tanré is working with Nicolas Fournier (Univ. Pierre et Marie Curie, Paris 6) 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 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. We prove the existence and uniqueness of a heuristically derived mean-field limit of the system when $n \to \infty$.
- E. Tanré is working 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 of 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. When $H \in (\frac{1}{4}, \frac{1}{2})$, they have proved the existence in the elliptic.
- D. Villemonais works in collaboration with Éliane Albuisson (CHRU of Nancy), Athanase Benetos (CHRU of Nancy), Simon Toupance (CHRU of Nancy), Daphné Germain (École des Mines de Nancy), Anne Gégout-Petit (Inria BIGS team) and Sylvain Chabanet (École des Mines de Nancy). The aim of this collaboration is to conduct a statistical study of the time evolution of telomere's length in human cells.
- D. Villemonais started a collaboration with Cécile Mailler (University of Bath) with the aim of studying the almost sure convergence of measure valued Pólya urns models.

7.2. Financial Mathematics

Participants: Madalina Deaconu, Antoine Lejay, Paolo Pigato, Khaled Salhi, Etienne Tanré.

7.2.1. Published works and preprints

- When the underlying asset price is given by a exponential Lévy model, the market is almost incomplete. Under this hypothesis, M. Deaconu, A. Lejay and K. Salhi worked on derivatives hedging under a budget constraint on the initial capital. He considers, as criterion of optimization, the CVaR of the terminal hedging risk. First, he rewrites the problem an optimisation problem on the random fraction of the payoff that permits to respect the budget constraint. Then, he approximates the problem by relaxing the constraint and considering only a specific equivalent martingale measure. This approximate problem is solved using Neyman-Pearson's Lemma and, in the case of European options, a numerical valuation of the approximated minimal CVaR based on fast Fourier transform [28].
- A. Lejay and P. Pigato studied the estimation of the coefficients of the Geometric Oscillating Brownian motion on financial data. This stochastic process is a modification of the Black & Scholes model that takes into account leverage effect and other sudden changes in the volatility [57], [41].
- V. Reutenauer and E. Tanré have worked on extensions of the exact simulation algorithm introduced by Beskos et al. [62]. They propose an unbiased algorithm to approximate the two first derivatives with respect to the initial condition x of quantities with the form $\mathbb{E}\Psi(X_T^x)$, where X is a one-dimensional diffusion process and Ψ any test-function. They also propose an efficient modification of Beskos et al. algorithm ([58], paper in revision).

BIGS Project-Team

6. New Results

6.1. Stochastic modelling

Participants: T. Bastogne, P. Vallois, S. Wantz-Mezieres, L. Batista, A. Gégout-Petit

Because of the observation of longitudinal data for each subject in medicine, we have to care about the random effect due to the subject and to choose adapted models like mixed effect models [39], [40]. We recently improved this methodology for the analysis of data collected in vivo for growth tumor for the biopharmaceutical company Transgene. The problem was to measure the differential effect of treatments (different molecules and doses) on the dynamics of the tumor taking into account the effect of censoring [10].

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 (1), we propose different tests based on the join count statistics [6].

6.2. Estimation and control for Markov Processes

Participants: R. Azais, F. Bouguet, T. Bastogne

We have developed statistical inference techniques for estimating the jump rate of PDMPs (piecewise-deterministic Markov processes) [2] which is an essential step to build relevant application models. In [2], we state a new characterization of the jump rate when the transition kernel only charges a discrete subset of the state space and deduce from it a competitive nonparametric technique for estimating this feature of interest. Our methodologies have been illustrated on numerical examples and real data. We also investigated the probabilistic properties of the PDMPs [5] or more general Markov processes [31] that could be useful to study properties of estimators.

A bit more generally, we have made contributions to a variety of specific estimation problems. We considered the problem of estimation of integrals under Markov design, which has a large variety of applications, in particular in biology and climatology. In [24], we have developed and analyzed a technique for estimating the average value over space when sensors describe a Markovian trajectory; this method leads to rates that are better than the traditional "root n"-rate, where n is the sample size, and was applied to the evaluation of the average temperature of oceans.

Control of stochastic processes is also a way to optimise administration (dose, frequency) of therapy. In [8], we have presented the design and validation of a real time controller able to track a preset photobleaching trajectory by modulating the width of light impulses during the treatment sessions, which is useful in a Photodynamic therapy context. This innovative solution was validated by *in vivo* experiments that have shown a significantly improvement of reproducibility of the inter-individual photobleaching kinetics. This innovative controller is the first personalized solution able to adapt in realtime the dose of light to be applied in photodynamic therapy.

6.3. Algorithms and Estimation for graph data

Participants: R. Azais, F. Bouguet, T. Bastogne

Tree-structured data naturally appear in various fields, particularly in biology where plants and blood vessels may be described by trees. The paper [27] is devoted to the estimation of the relative scale of ordered trees that share the same layout. The theoretical study is achieved for the stochastic model of conditioned Galton-Watson trees. New estimators are introduced and their consistency is stated. A comparison is made with an existing approach of the literature. A simulation study shows the good behavior of our procedure on finite-sample sizes.

6.4. Regression and machine learning

Participants: A. Gégout-Petit, A. Muller-Gueudin, T. Bastogne, L. Batista, R. Azais, S. Ferrigno, K. Duarte, J.-M. Monnez

We consider the problem of sequential least square multidimensional linear regression using a stochastic approximation process. The choice of the stepsize may be crucial in this type of process. In order to avoid the risk of numerical explosion which can be encountered, we define three processes with a variable or a constant stepsize and establish their convergence. Finally these processes are compared to classic processes on 11 datasets, 6 with a continuous output and 5 with a binary output, for a fixed total number of observations used and then for a fixed processing time. It appears that the third-defined process with a very simple choice of the stepsize gives usually the best results [32].

We study many other regression models like survival analysis, spatio temporal models with covariates. Among the multiple regression models, we want to test, thanks to simulation methods, validity of their assumptions [25]. Tests of this kind are called omnibus test. An omnibus test is an overall test that examines several assumptions together, the most known omnibus test is the one for testing gaussianity (that examines both skewness and kurtosis).

In the purpose of selecting factors linked to the efficiency of a treatment in the context of high dimension (about 100.000 covariates), we have developed a new methodology to select and rank covariates associated to a variable of interest in a context of high-dimensional data under dependence but few observations. The methodology imbricates successively rough selection, clustering of variables, decorrelation of variables using Factor Latent Analysis, selection using aggregation of adapted methods and finally ranking through bootstrap replications. Simulations study shows the interest of the decorrelation inside the different clusters of covariates. The methodology was applied to select covariates among genomics, proteomics covariates linked to the success of a immunotherapy treatment for lung cancer [21], [19], [20].

We also focus on the biological context of high-throughput and high-content bioassays in which several hundreds or thousands of biological signals are measured for a posterior analysis. In this experimental context, each culture well is a biological system in which the output variable is the cell proliferation, the input variable can be an electrical or a light stimulus signal and the covariate may be the type of cells, type of medium or tested compounds. The ambition is to identify a batch of several thousands of wells in a single step with the same model structure. Mixed effects models are largely used in regression but up to now they have rarely been used in the field of dynamical system identification. Our approach aims at developing a new solution based on an ARX (Auto Regressive model with eXternal inputs) model structure using the EM (Expectation-Maximisation) algorithm for the estimation of the model parameter [13], [10].

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 [47]. We are mostly concerned with three global health problems, namely fungal and bacterial infections and hypertension. Through on-going collaborations with several Brasilian 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 [72]. 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 [47], [31]. 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 [15].

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 [57], [56]. A paper describing this work has been published in the Journal of Biomedical Semantics [20].

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 [18]. 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.

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 [21]. 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 [16].

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 [23]. A full paper has been submitted to a special issue of *BMC Bioinformatics*, and is now in review. 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 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 [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 [12] 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 [22]. 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 [17].

MIMESIS Team

7. New Results

7.1. Augmented Reality in Surgical Navigation

7.1.1. Organ Pose Estimation for Augmented Reality in Hepatic Surgery

Participants: Y. Adagolodjo, R. Trivisonne, H. Courtecuisse, S. Cotin

A contribution focusing on intra-operative organ pose estimation was publihsed at the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2017) [19]. A novel method for semi-automatic registration of 3D deformable models using 2D shape outlines (silhouettes) extracted from a monocular camera view was introduced. The proposed framework is based on the combination of a biomechanical model of the organ with a set of projective constraints influencing the deformation of the model. To enforce convergence towards a global minimum for this ill-posed problem we interactively provide a rough (rigid) estimation of the pose. We show that our approach allows for the estimation of the non-rigid 3D pose while relying only on 2D information. The method is evaluated experimentally on a soft silicone gel model of a liver, as well as on real surgical data, providing augmented reality of the liver and the kidney using a monocular laparoscopic camera. Results show that the final elastic registration can be obtained in just a few seconds, thus remaining compatible with clinical constraints. We also evaluate the sensitivity of our approach according to both the initial alignment of the model and the silhouette length and shape.

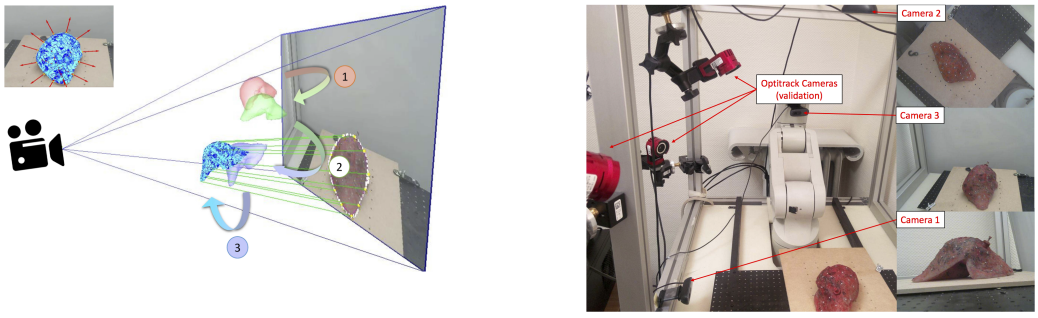

Figure 6. Left: 1) a direct simulation is applied to transform the reconstructed model obtain from the segmentation (red) in a shape close to the 3D position observed in the image (green). 2) A Rigid transformation (blue) is provided by the user to roughly align the model with the contour of the organ segmented in the image (yellow). 3) Projective constraints are applied to the biomechanical model to fit the organ contour and provide the 3D shape w.r.t. the camera position. Right: validation setup.

7.1.2. Image-driven Stochastic Estimation of Boundary Conditions

Participants: N. Haouchine, I. Peterlik, S. Cotin

A novel method was proposed in the context of image-driven stochastic simulation employed in the intra-operative navigation [25]. In the proposed approach, the boundary conditions are modeled as stochastic parameters. The method employs the reduced-order unscented Kalman filter to transform in real-time the probability distributions of the parameters, given observations extracted from intra-operative images. The method is evaluated using synthetic, phantom and real data acquired in vivo on a porcine liver. A quantitative assessment is presented and it is shown that the method significantly increases the predictive power of the biomechanical model employed by a framework implemented the augmented reality for surgical navigation.

7.2. Advanced Numerical Modeling and Simulation

7.2.1. Face-based Smoothed Finite Element Method for Real-time Simulation of Soft Tissue

Participants: A. Mendizabal, C. Paulus, R. Bessard-Duparc, I. Peterlik, S. Cotin

A method based on face-based smoothed finite element method was proposed and applied in the context of modeling of brain shift in [23]. This numerical technique has been introduced recently to overcome the overly stiff behavior of the standard FEM and to improve the solution accuracy and the convergence rate in solid mechanics problems. In this paper, a face-based smoothed finite element method (FS-FEM) using 4-node tetrahedral elements is presented. We show that in some cases, the method allows for reducing the number of degrees of freedom, while preserving the accuracy of the discretization. The method is evaluated on a simulation of a cantilever beam loaded at the free end and on a simulation of a 3D cube under traction and compression forces. Further, it is applied to the simulation of the brain shift and of the kidney's deformation. The results demonstrate that the method outperforms the standard FEM in a bending scenario and that has similar accuracy as the standard FEM in the simulations of brain shift and kidney deformation.

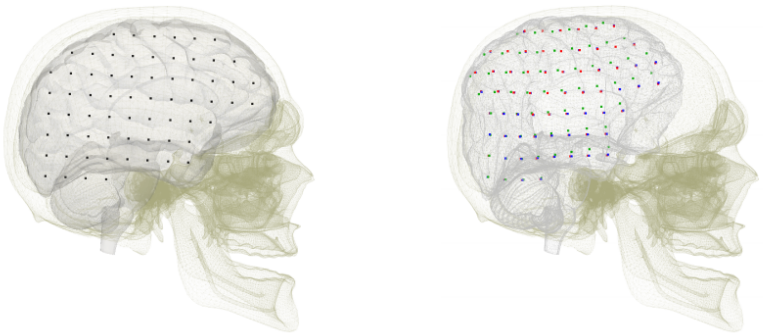

Figure 7. Grid representing tumor positions using a mesh of 7924 elements for linear FEM in blue, FS-FEM in red and non-linear FEM in green after the brain-shift. Left: rest position. Right: position after the deformation due to brain-shift.

7.2.2. Immersed Boundary Method for Real-time

Participants: C. Paulus, S. Cotin

Although the finite element method is widely used as a numerical approach in this area, it is often hindered by the need for an optimal meshing of the domain of interest. The derivation of meshes from imaging modalities such as CT or MRI can be cumbersome and time-consuming. In our contribution [24], we employed the Immersed Boundary Method (IBM) to bridge the gap between these imaging modalities and the fast simulation of soft tissue deformation on complex shapes represented by a surface mesh directly retrieved from binary images. A high resolution surface, that can be obtained from binary images using a marching cubes approach, is embedded into a hexahedral simulation grid. The details of the surface mesh are properly taken into account in the hexahedral mesh by adapting the Mirtich integration method. In addition to not requiring a dedicated meshing approach, our method results in higher accuracy for less degrees of freedom when compared to other element types. Examples on brain deformation demonstrate the potential of our method.

7.2.3. Error Control in Surgical Simulations

Participants: H. Courtecuisse, S. Cotin

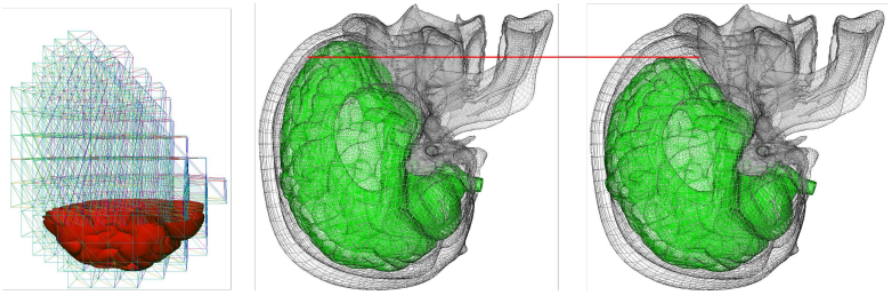

Figure 8. Simulation of brain shift using a detailed surface mesh embedded into an hexahedral grid. Boundary conditions are applied onto the exact surface, not the grid (left).

A contribution [16] presents the first real-time a posteriori error-driven adaptive finite element approach for real-time simulation and demonstrates the method on a needle insertion problem.

We use corotational elasticity and a frictional needle–tissue interaction model. The problem is solved using finite elements and the refinement strategy relies upon a hexahedron-based finite element method, combined with a posteriori error estimation driven local *h-refinement*, for simulating soft tissue deformation. We proposed to control the local and global error level in the mechanical fields (e.g. displacement or stresses) during the simulation. We show the convergence of the algorithm on academic examples, and demonstrate its practical usability on a percutaneous procedure involving needle insertion in a liver. For the latter case, we compare the force displacement curves obtained from the proposed adaptive algorithm with that obtained from a uniform refinement approach. Error control guarantees that a tolerable error level is not exceeded during the simulations. Local mesh refinement accelerates simulations. The work provides a first step to discriminate between discretization error and modeling error by providing a robust quantification of discretization error during simulations.

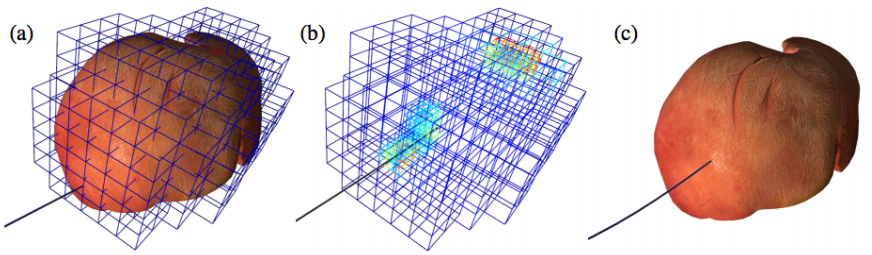

Figure 9. (a) Simulation of needle insertion in a liver; (b) Using dynamic mesh refinement scheme driven by error estimate; (c) Visual depiction. The simulation runs at 22 Hz using a PC with 4 GHz CPU.

7.3. Model-based Image Registration

7.3.1. Intraoperative Biomechanical Registration of the Liver

Participants: R. Plantefève, I. Peterlik, S. Cotin

Different aspects of model-based registration in the context of surgical navigation employing the augmented reality were analyzed in an invited contribution [17] published in the context of the attributed Prix de thèse de former Ph.D. student Rosalie Plantefève. Preoperative images such as computed tomography scans or magnetic resonance imaging contain lots of valuable information that are not easily available for surgeons during an operation. To help the clinicians better target the structures of interest during an intervention, many registration methods that align preoperative images onto the intra-operative view of the organs have been developed. For important organ deformation, biomechanical model-based registration has proven to be a method of choice. Using an existing model-based registration algorithm for laparoscopic liver surgery we investigated the influence of the heterogeneity of the liver on the registration result. It was found that the use of an heterogeneous model does not improve significantly the registration result but increases the computation time necessary to perform the registration.

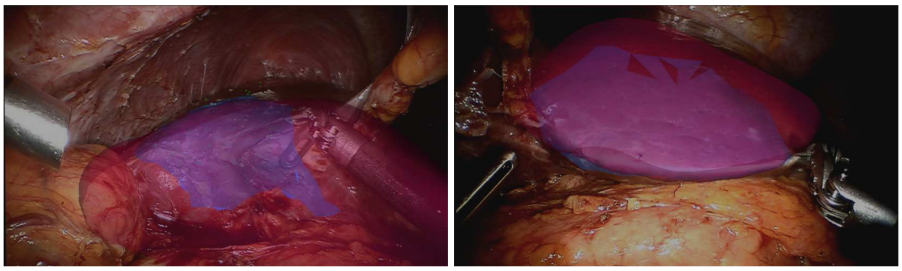

Figure 10. Registration results on in vivo data on two different views of a human liver. The registered mesh is shown in red while the partial reconstructed patch is depicted in blue.

7.3.2. Registration of Cell Nuclei in Cell Microscopy

Participants: I. Peterlik

A contribution *Registration of Cell Nuclei in 2D Live Cell Microscopy* was published in a collaboration with Centre of Biomedical Image Analysis at Masaryk University, Czech Republic [18]. The analysis of the pure motion of sub-nuclear structures without influence of the cell nucleus motion and deformation is essential in live cell imaging. We proposed a 2D contour-based image registration approach for compensation of nucleus motion and deformation in fluorescence microscopy time-lapse sequences. The proposed approach extends our previous approach which uses a static elasticity model to register cell images. Compared to that scheme, the new approach employs a dynamic elasticity model for forward simulation of nucleus motion and deformation based on the motion of its contours. The contour matching process is embedded as a constraint into the system of equations describing the elastic behavior of the nucleus. This results in better performance in terms of the registration accuracy. Our approach was successfully applied to real live cell microscopy image sequences of different types of cells including image data that was specifically designed and acquired for evaluation of cell image registration methods.

7.4. Reconstruction of Geometries from Images

7.4.1. Automatic Skeletonization of Vascular Trees in Pre-operative CT Images

Participants: R. Plantefève, I. Peterlik

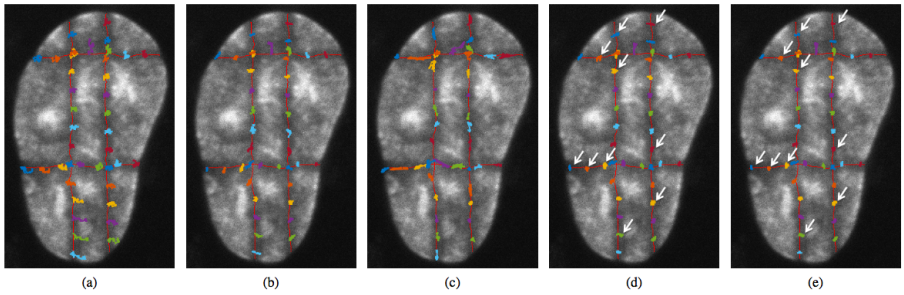

Figure 11. Tracks of line features overlaid with the first image of the sequence. The tracks represent the motion of the points of the line features sampled with 30 pixel interval for better visibility. The tracks are shown for (a) unregistered data, (b) after registration with the contour-based approach [19], (c) after registration with the intensity-based approach [9], (d) after registration with the static version of our approach, and (e) after registration with the proposed dynamic approach. White arrows indicate tracks with the most visible difference between (d) and (e).

An algorithm of an automatic skeletonization of vascularization based on Dijkstra minimum-cost spanning tree was published in [27]. The result is an extension of an existing graph-based method where the vascular topology is constructed by computation of shortest paths in a minimum-cost spanning tree obtained from binary mask of the vascularization. We suppose that the binary mask is extracted from a 3D CT image using automatic segmentation and thus suffers from important artifacts and noise. When compared to the original algorithm, the proposed method (i) employs a new weighting measure which results in smoothing of extracted topology and (ii) introduces a set of tests based on various geometric criteria which are executed in order to detect and remove spurious branches. The method is evaluated on vascular trees extracted from abdominal contrast-enhanced CT scans and MR images. The method is quantitatively compared to the original version of the algorithm showing the importance of proposed modifications. Since the branch testing depends on parameters, the para-metric study of the proposed method is presented in order to identify the optimal parametrization.

7.4.2. Template-based Recovery of Elastic Shapes from Monocular Video

Participants: N. Haouchine, S. Cotin

A method of template-based 3D recovery of elastic shapes using Lagrange multiplied was presented at a top computed-vision conference [21]. By exploiting the object's elasticity, in contrast to isometric methods that use inextensibility constraints, a large range of deformations can be handled. Our method is expressed as a saddle point problem using Lagrangian multipliers resulting in a linear system which unifies both mechanical and optical constraints and integrates Dirichlet boundary conditions, whether they are fixed or free. We experimentally show that no prior knowledge on material properties is needed, which exhibit the generic usability of our method with elastic and inelastic objects with different kinds of materials. Comparison with existing techniques are conducted on synthetic and real elastic objects with strains ranging from 25% to 130% resulting to low errors.

7.5. Simulation for Intra-operative Rehearsal

Participants: N. Haouchine, F. Roy, S. Cotin

DejaVu, a novel surgical simulation approach for intra-operative surgical gesture rehearsal was published in [22] in collaboration with UCL London. With DejaVu we aim at bridging the gap between pre-operative surgical simulation and crucial but not yet robust intra-operative surgical augmented reality. By exploiting

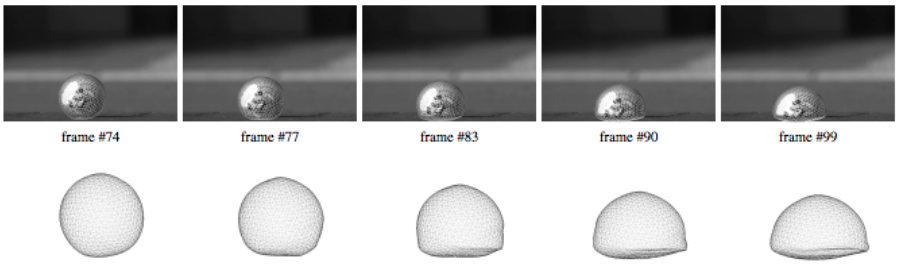

Figure 12. The proposed method illustrated on an example with a soft ball colliding the ground in slow motion. No prior knowledge of material properties in considered. The spherical volume model is composed of 512 linear P1 tetrahedral elements. The recovery and augmentation is performed in real-time at 25 FPS.

intra-operative images we produce a simulation that faithfully matches the actual procedure without visual discrepancies and with an underlying physical modeling that performs real-time deformation of organs and surrounding tissues, surgeons can interact with the targeted organs through grasping, pulling or cutting to immediately rehearse their next gesture. We present results on different in vivo surgical procedures and demonstrate the feasibility of practical use of our system.

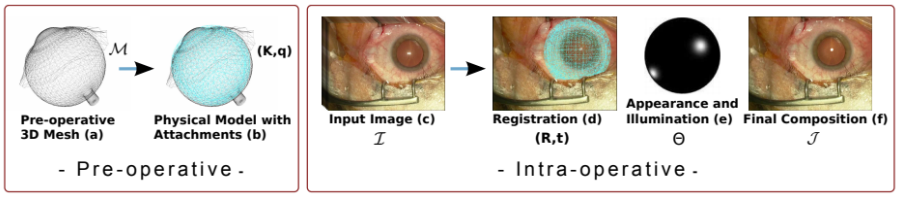

Figure 13. Schematic illustration of DejaVu Simulation. (a) preoperative model is built from tomographic images; (b) material law, tissue properties and attachments, constitute the physical model; (c) an intra-operative image is selected; (d) 3D/2D registration is performed between the physical model in (b) and the selected frame in (c); (e) appearance and illumination are estimated corresponding to specular and diffuse components and light position; (f) the final composition is build to enable surgical gesture rehearsal.

NEUROSYS Project-Team

7. New Results

7.1. From the microscopic to the mesoscopic scale

Participants: Laure Buhry, Axel Hutt, Francesco Giovannini, Mélanie Aussel, Ivan Kotiuchi.
In collaboration with Radu Ranta (university of Lorraine), Beate Knauer and Motoharu Yoshida (Ruhr university) and LieJune Shiau (university of Houston)

7.1.1. Memory and anesthesia

7.1.1.1. Modeling effects of propofol anesthesia

Neural oscillations are thought to be correlated with the execution of cognitive functions. Indeed, gamma oscillations are often recorded in functionally-coupled brain regions for cooperation during memory tasks, and this rhythmic behavior is thought to result from synaptic GABAergic interactions between interneurons. Interestingly, GABAergic synaptic and extrasynaptic receptors have been shown to be the preferred target of the most commonly used anesthetic agents. We presented a in-depth computational study [0] [1] of the action of anesthesia on neural oscillations by introducing a new mathematical model which takes into account the four main effects of the anesthetic agent propofol on GABAergic hippocampal interneurons. These are: the action on synaptic $GABA_A$ receptors, which includes an amplification and an extension of the duration of the synaptic currents, as well as an increase in current baseline, and the action on extrasynaptic $GABA_A$ receptors mediating a tonic inhibitory current. Our results indicate that propofol-mediated tonic inhibition contributes to an unexpected enhancement of synchronization in the activity of a network of hippocampal interneurons. This enhanced synchronization could provide a possible mechanism supporting the occurrence of intraoperative awareness, explicit memory formation, and even paradoxical excitation under general anesthesia, by transiently facilitating the communication between brain structures which should supposedly be not allowed to do so when anesthetized.

7.1.1.2. Stability Analysis in a model of hippocampal place cells

Ring networks, a particular form of Hopfield neural networks, can be used to model the activity of place cells, a type of cells in the hippocampus that are involved in the building and memorization of a cognitive map of one's environment. The behavior of these models is highly dependent on their recurrent synaptic connectivity matrix and on individual neurons' activation function, which must be chosen appropriately to obtain physiologically meaningful conclusions. In [4], we proposed several simpler ways to adjust this synaptic connectivity matrix compared to existing literature so as to achieve stability in a ring attractor network with a piece-wise affine activation functions, and we link these results to the possible stable states the network can converge to.

7.1.1.3. Modeling of the hippocampal formation over the sleep-wake cycle :

The hippocampus can exhibit different oscillatory rhythms within the sleep-wake cycle, each of them being involved in cognitive processes. For example, theta-nested gamma oscillations, consisting of the coupling of theta (4-12Hz) and gamma (40-100Hz) rhythms, are produced during wakefulness and are associated with spatial navigation tasks, whereas Sharp-Wave-Ripple (SWR) complexes, consisting of fast (140-200Hz) oscillatory events occurring at low ($\leq$ 0.5Hz) frequencies, are produced during slow-wave sleep and play an important role in memory consolidation. The mechanisms underlying the generation and switch between each of these rhythms is not yet fully understood, but Acetylcholine is thought to play a key role in it.

[0] F. Giovannini and L. Buhry, Tonic inhibition mediates a synchronization enhancement during propofol anesthesia in a network of hippocampal interneurons: a modeling study Journal of computational neuroscience (Submitted) 2017

In an article in preparation, we propose a computational model of the hippocampal formation based on a realistic topology and synaptic connectivity, influenced by the changing concentration of Acetylcholine between wakefulness and sleep. By using a detailed estimation of intracerebral recordings, we show 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 that appear during slow-wave sleep. The results of our simulations support the idea that the functional connectivity of the hippocampus is a key factor in controlling its rhythms.

7.2. From the Mesoscopic to the Macroscopic Scale

Participants: Laurent Bougrain, Axel Hutt, Tamara Tošić, Cecilia Lindig-León, Romain Orhand, Sébastien Rimbert, Oleksii Avilov, Rahaf Al-Chwa.
In collaboration with Stéphanie Fleck (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. Continuous and discrete

In most BCI experimental paradigms based on Motor Imagery (MI), subjects perform continuous motor imagery (CMI), i.e., a repetitive and prolonged intention of movement, for a few seconds. To improve efficiency such as detecting faster a motor imagery and thus avoid fatigue and boredom, we proposed to show the difference between discrete motor imagery (DMI), i.e., a single short MI, and CMI. The results of the experiment involving 13 healthy subjects suggest that DMI generates a robust post-MI event-related synchronization (ERS). Moreover event-related desynchronization (ERD) produced by DMI seems less variable in certain cases compared to CMI [10], [12]. We showed the difference, in term of classification, between a DMI and a CMI. The results of the experiment involving 16 healthy subjects show that a BCI based on DMI is as effective as a BCI based on CMI and could be used to allow a faster detection [6].

7.2.1.2. Profiling

The most common approach for classification consists of analyzing the signal during the course of the motor task within a frequency range including the alpha band, which attempts to detect the Event-Related Desynchronization (ERD) characteristics of the physiological phenomenon. However, to discriminate right-hand KMI and left-hand KMI, this scheme can lead to poor results on subjects for which the lateralization is not significant enough. To solve this problem, we proposed to analyze the signal at the end of the motor imagery within a higher frequency range, which contains the Event-Related Synchronization (ERS). We showed that 6 out of 15 subjects have a higher classification rate after the KMI than during the KMI, due to a higher lateralization during this period. Thus, for this population we obtained a significant improvement of 13% in classification taking into account the users lateralization profile [9].

7.2.1.3. Combined motor imageries

Combined motor imageries can be detected to deliver more commands in a Brain-Computer Interface for controlling a robotic arm. Nevertheless only a few systems use more than three motor imageries: right hand, left hand and feet. Combining them allows to get four additional commands. We presented an electrophysiological study to show that i) simple motor imageries have mainly an electrical modulation over the cortical area related the body part involved in the imagined movement and that ii) combined motor imageries reflect a superposition of the electrical activity of simple motor imageries. A shrinkage linear discriminant analysis has been used to test as a first step how a resting state and seven motor imageries can be detected. 11 healthy subjects participated in the experiment for which an intuitive assignment has been done to associate motor imageries and movements of the robotic arm with 7 degrees of freedom [2], [5].

7.2.1.4. Anesthesia

Each year, several million of general anesthesia are realized in France. A recent study shows that, between 0.1-0.2 % of patients are victims of intraoperative awareness. This kind of awakening could cause post-traumatic syndromes for the patient. Unfortunately, today, no monitoring system is able to avoid the intraoperative awareness phenomenon. Interestingly, if there is no subject movement due to curare, an electroencephalographical study of the motor cortex can help to detect an intention of movement. The dynamic study of motor cerebral activity during general anesthesia is essential if we want to create a brain-computer interface adapted to the detection of intraoperative awareness. We wrote a clinical protocol to allow EEG data recording during general anesthesia with propofol. Then, the development of temporal analysis specific methods allows us to quantify patterns of desynchronization and synchronization phases observed in delta, alpha and beta frequency bands to prevent intraoperative awareness [8].

TONUS Team

7. New Results

7.1. Palindromic methods

7.1.1. Palindromic discontinuous Galerkin method in 2D and 3D

Participants: David Coulette, Florence Drui, Emmanuel Franck, Philippe Helluy, Laurent Navoret.

In the previous year (see [7]) we have proposed a method to solve hyperbolic systems like the Euler equations with an unconditionally stable high-order method. This method is based on a kinetic representation of the hyperbolic system. The kinetic equations are solved with an upwind DG method. It requires no matrix storage. High order is obtained through palindromic composition methods. The concept has been test in 1D. During this year we extend the method to 2D and 3D and applied it to fluid mechanics. Currently we are working on improving this method on realistic cases for MHD instabilities. The objective is to compare the results with the European code JOREK.

We are also working on methods for applying boundary conditions in a stable way with the palindromic method (postdoc of Florence Drui).

7.1.2. Kinetic model for palindromic methods

Participants: David Coulette, Emmanuel Franck, Laurent Navoret.

One of the most important drawbacks of the Palindromic method is the numerical dispersion associated to the high-order time scheme. To limit this problem we propose to replace the DG method by a semi-Lagrangian method and design new kinetic representations which are more accurate. We also studied the stability of these news models. The first results were good and currently we are working on the 2D extension and the coupling with limiter technics.

7.1.3. Finite element relaxation methods for fluid models

Participants: David Coulette, Emmanuel Franck.

In parallel to our work on the Palindromic method based on a kinetic relaxation model, we studied in [17] a variant based on the Xin-Jin relaxation model. Coupled with a finite element method we obtain an implicit solver for Euler equations where we invert only Laplacians and mass matrices. The first results show that the method is more efficient in CPU costs and memory. The finite elements used are the same as in JOREK.

7.2. MHD problems

Participant: Emmanuel Franck.

7.2.1. Compatible Implicit finite element for linear MHD

In this work we consider a linear MHD problem. The aim is to design an implicit method able to preserve the energy equation and the divergence free constraints in realistic Tokamak geometry. The first idea is to use a splitting scheme between the wave and convection parts coupled with an implicit scheme for each subsystem. In order to discretize each sub-system we use compatible B-Splines FE method wich allows us to preserve the invariants and to use a reduction of the implicit problem to be inverted. The idea was improved on simple geometries. We are currently extending the method on realistic geometries.

7.2.2. Splitting and relaxation for JOREK code

The Jorek code is the main European code for the simulation of Tokamak instabilities. The inversion of the full matrix is based on Block Jacobi preconditioning which is not efficient in some cases and very greedy in memory. We are investigating a new splitting scheme similar to the one used in works on compatible Finite Elements. We have also just begun to investigate the relaxation method used in the Palindromic scheme to solve the reduced MHD model of JOREK.

7.3. Finite Volume approximations of the Euler system with variable congestion

Participants: Pierre Degond, Piotr Minakowski, Laurent Navoret, Ewelina Zatorska.

We are interested in the numerical simulations of the Euler system with variable congestion encoded by a singular pressure. This model describes for instance the macroscopic motion of a crowd with individual congestion preferences. In [3] we propose an asymptotic preserving (AP) scheme based on a conservative formulation of the system in terms of density, momentum and density fraction. A second order accuracy version of the scheme is also presented. We validate the scheme on one-dimensional test cases and compare it with a scheme developed in a previous work and extended here to higher order accuracy. We finally carry out two-dimensional numerical simulations and show that the model exhibits typical crowd dynamics.

7.4. Numerical scheme for sheath equilibria

Participants: Mehdi Badsi, 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 [19]. 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. Moreover, for preventing instabilities from developing in large time, the proposed method guaranties that the discrete Gauss equation is satisfied in time.

7.5. Recurrence phenomenon for finite element grid based Vlasov solver

Participants: Michel Mehrenberger, Laurent Navoret, Thi Nhung Pham.

When using a grid based solver (finite element/DG scheme, discontinuous Galerkin semi-Lagrangian scheme) and spatial periodic boundary conditions, the simulations of the Vlasov-Poisson system exhibit numerical reappearance of initial perturbations at some time called recurrence time. This time depends on the numerical parameters (degree and mesh size of the finite element mesh). With a given number of degrees of freedom, considering a large degree approximation makes the phenomenon appear earlier in the simulation and thus makes this choice less attractive. In our work [9], we highlight that the time and the intensity of the recurrence are related to the quadrature rules used for computing the charge density. In particular, quadratures that are exact on trigonometric polynomials weaken the recurrence effect.

7.6. PICSL

Participants: Yann Barsamian, Joackim Bernier, Sever Hirstoaga, Michel Mehrenberger.

7.6.1. Particle in Cell and Semi-Lagrangian schemes for two species plasma simulations

Thanks to a classical first order dispersion analysis, we are able to check the validity of 1D$\times$1D 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 in [14] a 2D $\times$ 2D 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 2D×2D nonlinear two species simulations with mass ratio and consider the mixing of semi-Lagrangian and Particle-in-Cell methods. This work has been initiated at CEMRACS 2016.

7.7. TARGET

Participants: Nicolas Bouzat, Guillaume Latu, Camilla Bressan, Michel Mehrenberger, Virginie Grandgirard.

7.7.1. TArgeting Realistic GEometry in Tokamak code gysela

The framework of the work in [16] is the Semi-Lagrangian setting for solving the gyrokinetic Vlasov equation and the Gysela code. A new variant for the interpolation method is proposed that can handle the mesh singularity in the poloidal plane at $r = 0$ (a polar system is used for the moment in Gysela). A non-uniform meshing of the poloidal plane is proposed, instead of a 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 shapes is used to improve the realism. Convergence studies are provided to establish the validity and robustness of our new approach. This work has been initiated at CEMRACS 2016.

7.8. Field-aligned interpolation for gyrokinetics

Participants: Yaman Güclü, Philippe Helluy, Guillaume Latu, Michel Mehrenberger, Laura Mendoza, Eric Sonnendrücker, Maurizio Ottaviani.

This work is devoted to the study of field-aligned interpolation in semi-Lagrangian codes. This work has been initiated in 2013; this year the article has been accepted [5]. In the context of numerical simulations of magnetic fusion devices, this approach is motivated by the observation that gradients of the solution along the magnetic field lines are typically much smaller than along a perpendicular direction. In toroidal geometry, field-aligned interpolation consists of a 1D interpolation along the field line, combined with 2D interpolations on the poloidal planes (at the intersections with the field line). A theoretical justification of the method is provided in the simplified context of constant advection on a 2D periodic domain: unconditional stability is proven, and error estimates are given which highlight the advantages of field-aligned interpolation. The same methodology is successfully applied to the solution of the gyrokinetic Vlasov equation, for which we present the ion temperature gradient (ITG) instability as a classical test case: first we solve this in cylindrical geometry (screw-pinch), and next in toroidal geometry (circular Tokamak). In the first case, the algorithm is implemented in Selalib (semi-Lagrangian library), and the numerical simulations provide linear growth rates that are in accordance with the linear dispersion analysis. In the second case, the algorithm is implemented in the Gysela code, and the numerical simulations are benchmarked with those employing the standard (not aligned) scheme. Numerical experiments show that field-aligned interpolation leads to considerable memory savings for the same level of accuracy; substantial savings are also expected in reactor-scale simulations.

We are also currently implementing into SCHNAPS a general transport solver for addressing non-conforming patches in complex geometries. The objective is to be able to design meshes that are able to deal with magnetic aligned geometries. The resulting scheme will be used for solving kinetic equations, of course. But it can also be the building block of a palindromic method applied on curved and non-conforming meshes.

7.9. InKS

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

7.9.1. A programming model to decouple performance from semantics in simulation codes

Existing programming models lead to a tight interleaving of semantics and computer optimization concerns in high-performance simulation codes. With the increasing complexity and heterogeneity of supercomputers 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. The report in [12] proposes InKS, a programming model that aims to improve the situation by decoupling semantics and optimizations in code so as to ease the collaboration between domain scientists and experts in high-performance optimizations. We define the InKS language that enables developers to describe the semantics of a simulation code with no concern for performance. We describe the implementation of a compiler able to automatically execute this code without making any explicit execution choice. We also describe a method to manually specify these choices to reach high-performance. Our preliminary evaluation on a 3D heat equation solver demonstrates the feasibility of the automatic approach as well as the ability to specify complex optimizations while not altering the semantic part. It shows promising performance where two distinct specifications of optimization choices in InKS offer similar performance as existing hand-tailored versions of the solver.

7.10. Performance of Particle-in-Cell methods

Participants: Yann Barsamian, Sever Hirstoaga, Eric Violard.

In a two-dimensional framework, in [6] we optimized a Particle-in-Cell (PIC) code by analyzing different data structures for the particles and for the grid fields with the aim of improving the cache reuse and by using the vectorization from the compiler. We also parallelized the code with OpenMP/MPI and satisfactory strong and weak scaling up to 8192 cores were obtained on the supercomputer CURIE.
Currently [15] we are extending and improving this work to a three-dimensional electrostatic PIC code.

7.11. Comparison of multiscale PIC methods

Participants: Nicolas Crouseilles, Sever Hirstoaga, Xiaofei Zhao.

In [2] we study different types of multiscale methods to numerically study the long-time Vlasov–Poisson equation with a strong magnetic field. The multiscale methods are an asymptotic preserving Runge–Kutta scheme, an exponential time differencing scheme, the stroboscopic averaging method and a uniformly accurate two-scale formulation. Extensive numerical experiments are conducted to investigate and compare the accuracy, efficiency, and long-time behavior of all the methods. The methods with the best performance under different parameter regimes are identified.

COAST Project-Team

5. New Results

5.1. 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. Such shared content can be edited by thousands of people. However, current consistency maintenance algorithms seem not to be adapted to massive collaborative updating involving large amount of contributors and a high velocity of changes. This year we designed new optimistic replication algorithms for maintaining consistency for complex data such as wikis. We also designed a peer-to-peer web-based real-time collaborative editor relying on our proposed algorithms as well as a mechanism that balances awareness and disturbance in this kind of systems. We also started to study collaborative editing user behavior.

Wikis are one of the most important tools of Web 2.0 allowing users to easily edit shared data. However, wikis offer limited support for merging concurrent contributions on the same pages. Users have to manually merge concurrent changes and there is no support for an automatic merging. Real-time collaborative editing reduces the number of conflicts as the time frame for concurrent work is very short. We proposed extending wiki systems with real-time collaboration and designed an automatic merging solution adapted for rich content wikis [2]. Our merging solution is based on an operational transformation approach for which we defined operations with high-level semantic capturing user intentions when editing wiki content such as move, merge and split. Our solution is the first one that deals with high level operations, existing approaches being limited to operations of insert, delete and update on textual documents.

Existing real-time collaborative editors rely on a central authority that stores user data which is a perceived privacy threat.We designed MUTE [8], a peer-to-peer web-based real-time collaborative editor that eliminates the disadvantages of central authority based systems. Users share their data with the collaborators they trust without having to store their data on a central place. MUTE features high scalability and supports offline and ad-hoc collaboration. MUTE relies on LogootSplit, a CRDT-based consistency maintenance algorithm for strings [15]. MUTE collaborative editor will be integrated in the virtual desktop of OpenPaaS::NG project [8].

When people work collaboratively on a shared document, they have two contradictory requirements on their editors that may affect the efficiency of their work. On the one hand, users would like to be aware of other users work on a particular part of the document. On the other hand, users would like to focus their attention on their own current work, with as little disturbance from the concurrent activities as possible. We designed a mechanism that lets users handle a balance between disturbance and awareness of concurrent updates [10]. Users can define focus regions and concentrate on the work in these regions without being disturbed by work of other users. Occasionally, users can preview concurrent updates and select a number of these updates to be integrated into the local copy.

We are interested in analysing user behavior during collaborative editing. This year we studied concurrency and conflicts in asynchronous collaboration [7]. We chose to study collaboration traces of distributed version control systems such as Git. We analysed Git repositories of four projects: Rails, IkiWiki, Samba and Linux Kernel. We analyzed the collaboration process of these projects at specific periods revealing how changes integration evolves during project development. We also analyzed how often users decide to rollback to previous document version when the integration process results in conflict. Finally, we studied the mechanism adopted by Git to consider changes made on two continuous lines as conflicting.

5.2. Trust-based Collaboration

Participants: Quang Vinh Dang, Claudia-Lavinia Ignat, Francois Charoy, Olivier Perrin, Mohammed Riyadh Abdmeziem, 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. In order to compute the trust score of users according to their contributions during a collaborative editing task, we need to evaluate the quality of the content of a document that has been written collaboratively. We investigated how to automatically assess the quality of Wikipedia articles in order to provide guidance for both authors and readers of Wikipedia. Most existing approaches for quality classification of Wikipedia articles rely on traditional machine learning with manual feature engineering, which requires a lot of expertise and effort and is language dependent. We proposed an approach that addresses the trade-off between accuracy, time complexity and language independence for the prediction models [5]. Our approach relying on Recurrent Neural Networks (RNN) eliminates disadvantages of feature engineering, i.e. it learns directly from raw data without human intervention and is language-neutral. Experimental results on English, French and Russian Wikipedia datasets show that our approach outperforms state-of-the-art solutions.

Rating prediction is a key task of e-commerce recommendation mechanisms. Recent studies in social recommendation enhance the performance of rating predictors by taking advantage of user relationships. However, these prediction approaches mostly rely on user personal information which is a privacy threat. We proposed dTrust [6], a simple social recommendation approach that avoids using user personal information. It relies uniquely on the topology of an anonymized trust-user-item network that combines user trust relations with user rating scores for items. This topology is fed into a deep feed-forward neural network. Experiments on real-world data sets showed that dTrust outperforms state-of-the-art in terms of Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) scores for both warm-start and cold-start problems.

One dimension of our work is dedicated to ensure consistency of the key server. We design Trusternity, which is a secure, scalable auditing mechanism using a blockchain to replace the gossiping mechanism of transparent log system. We have implemented Trusternity as a proof-of-concept, and we have led some evaluation about the detection of malicious behavior on the blockchain network.

Securing P2P collaborative system remains a critical issue for its widespread adoption. One of our goals is to ensure that communication between collaborating partner is secure from end to end. We need to encrypt exchange of operations among partners. For that we propose to rely on group keys management. One of the issue is that the composition of the partnership can change and this require to change the group key. Since we don't want a central server to manage keys, that would break the p2p nature of our approach we need to propose group key management protocols that are resilient to change in groups, even in group of large size. [3]

5.3. Cloud Provisioning for Elastic BPM

Participants: François Charoy, Samir Youcef, Guillaume Rosinosky.

Cloud computing provider do not help consumer to use optimally the available resources. For this, several approaches have been proposed [24] that take benefit from the elasticity of the Cloud, starting and stoppping 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 optimize 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 [9].

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

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

The framework is a comprehensive approach that decomposes a business process into fragments that can automatically be deployed on multiple clouds. The framework also integrates a selection algorithm that combines the security information of cloud offers and of the process with other quality of service criteria to generate an optimized configuration. It is implemented in a tool to assess cloud providers 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) [4].
- Additionally are developing a simulation tool for supporting designer decision with the ability to balance risk with cost when selecting the best cloud configuration.

5.5. 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 generaly considered as blind and stateless, which mean that in each business process execution results from past process instances are not taken into consideration.

The main objective of our current research is to exploit the data generated from previous instances in order to enhance business processes in regards with several aspects, such as improvement of process business logical correctness, optimization of business process modeling issues, or improvemment of resource allocation and scheduling procedure in order to particularly optimize costs and time (among other factors).

We focus currently on this last aspect, i.e. scheduling and resource allocation in business processes. Business Processes may contain automatic tasks and non automatic tasks, so managing resources depends on the type of those resources (human or machine) In this context, our work use machine learning techniques to analyze data generated from previous business process execution to improve business process scheduling. This step ensure the assignment of the most critical business process instance task to a qualified (and may be costly) human resource while minimizing global execution costs through assignement of “dummy” tasks to machine agents.

MADYNES Team

7. New Results

7.1. Monitoring

7.1.1. Quality of Experience Monitoring

Participants: Isabelle Chrisment [contact], Thibault Cholez, Vassili Rivron, 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 parallel, to evaluate our methods, we need a testing framework to automate testing of WiFi P2P mobile apps at scale. In [20] we proposed AndroFleet, a large-scale WiFi P2P testing framework. Androfleet can perform User Acceptance Testing for a fleet of emulators, by emulating the hardware behavior of the peer discovery, it gives the developers the ability to control P2P specific behaviors (peers joining and leaving).

7.1.2. Active Monitoring

Participants: Abdelkader Lahmadi [contact], Jérôme François, Frédéric Beck [LHS], Loic Rouch [LHS].

Following the work done in 2016, we pursued our collaboration with the regional PME TracIP (http://www.tracip.fr) on the development of attack assessment and forensics platform dedicated to industrial control systems. The platform involves multiple PLC from different manufacturers and real devices of factory automation systems (see 6.7.1).

During the year 2017, we have demonstrated that off-the-shelf hardware is sufficient to take over any Z-Wave network without knowing its topology or compromising any original devices and remaining unnoticeable for the primary controller. Our attack consists in building an adversary Z-Wave universal controller by reprogramming a mainstream USB stick controller. The technique exploits two features provided by the USB stick which allow (1) to set the network identifier (HomeID) and (2) to learn many devices identifiers even if they are not physically available. The attack has been demonstrated in Blackhat Europe 2017 by Loic Rouch (https://www.blackhat.com/eu-17/briefings/schedule/#a-universal-controller-to-take-over-a-z-wave-network-8459).

7.1.3. Service-level Monitoring of HTTPS traffic

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

We previously proposed an alternative technique to investigate HTTPS traffic which aims to be robust, privacy-preserving and practical with a service-level identification of HTTPS connections, i.e. to name the services, without relying on specific header fields that can be easily altered. We have defined dedicated features for HTTPS traffic that are used as input for a multi-level identification framework based on machine learning algorithms processing full TLS sessions. Our evaluation based on real traffic shows that we can identify encrypted web services with a high accuracy. In 2017, we finished to develop our solution to make it fully usable in real-time [1]. We now provide our prototype implementation (https://gitlab.inria.fr/swazen/HTTPSFirewall) in open-source. It operates by extending the iptables/netfilter architecture. It receives and demultiplexes the arriving HTTPS packets to a related flow. As soon as the number of packets in a given flow reaches a threshold, the identification engine extracts the features and runs the C4.5 algorithm to predict the HTTPS service of the flow.

7.1.4. Monitoring Programmable Networks

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

Software-Defined Networking brings new capabilities in operating networks including monitoring. In the state-of-the art many proposals have been made to enhance monitoring of networks using OpenFlow or other proposed programmable frameworks. In a preliminary work [11], we reviewed them in order to highlight what are the remaining challenges to be addressed in that area. The main issue is the trade-off to be made between the strong expressibility (especially stateful operations) and capability of monitoring techniques that are necessary for advanced operation purposes and the complexity it induces if we want to keep the pace with line-rate packet processing. Another important aspect is the security as adding programmable monitoring functions may lead to introduce security threats. Our current work is thus focused on adding monitoring capacity while guaranteeing line-rate operations and safety requirements even when programs are deployed on running network switches.

7.1.5. Smart Contracts Monitoring

Participants: Jérôme François [contact], Sofiane Lagraa, Radu State [University of Luxembourg], Jérémy Charlier [University of Luxembourg].

Blockchain technologies are skyrocketing and the team is interested in assessing the impact of such technologies on networking, and if necessary managing the coupling between them. Indeed, blockchain efficiency resides in an overlay network built on top of a real infrastructure which needs to properly support it. Orchestrating network ressources, *i.e.* adding some network capacity, might be helpful but supposes first an in-depth monitoring of blockchain interactions. In a first work, we thus evaluated the relation among smart contracts. We defined methods to discover smart contracts interactions and the different group properties. This approach relies on graph modelling and mining techniques as well as tensor modelling combined with stochastic processes. It underlines actual exchanges between smart contracts and targets the predictions of future interactions among the communities. Comparative study between graph analysis and tensor analysis is provided for predictions of smart contract interactions. Finally, virtual reality visualization based on Unity 3D game engine has been applied [12].

7.1.6. Sensor networks monitoring

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

Our work on IoT security monitoring has been published in IEEE Transactions on Network and Service Management [4]. This concerns more specifically our distributed monitoring architecture for detecting attacks against RPL networks. The RPL routing protocol has been standardized by IETF to enable a lightweight and robust routing in lower-power and lossy networks. After having compared existing IoT monitoring solutions, we have proposed a detection strategy for RPL version number attacks. This one relies on our monitoring architecture to preserve constrained node resources, in the context of AMI infrastructures. A versioning mechanism is incorporated into RPL in order to maintain an optimized topology. However, an attacker can exploit this mechanism to significantly damage the network and reduce its lifetime. We have exploited monitoring node collaboration to identify the attacker, the localization process being performed by the root after gathering detection information from all monitoring nodes. We have evaluated our solution through experiments and have analyzed the performance according to defined metrics. We have shown that the false positive rate of our solution can be reduced by a strategic monitoring node placement. We have also considered the scalability issue, by modeling this placement as an optimization problem and quantifying the number of required monitoring nodes to ensure acceptable false positive rates.

7.2. Security

7.2.1. Security analytics

Participants: Jérôme François [contact], Abdelkader Lahmadi, Sofiane Lagraa, Soline Blanc, Giulia de Santis, Olivier Festor, Radu State [University of Luxembourg], Christian Hammerschmidt [University of Luxembourg].

In 2017, we have continued our active cooperation with the High Security Lab (HSL) in Nancy. The latter provides the infrastructure to support two main projects in security analytics, namely the FUI HuMa project and the ATT AMICS. Thanks to darknet data of the HSL, we developped two methods based on graph-mining to extract knowledge. The first one focuses on port scanning analysis in order to profile the behaviours and patterns of attackers. By representing consecutive targeted ports in an aggregated graph format, we assess then the centrality of port number using different metrics and highlights valuable correlation among some of them. We are particularly able to identify patterns of scanning related to a specific setup (e.g. medical environment) [17]. We then extended this method to security events analysis by constructing multiple graphs to be analyzed with an outlier technique. The rationale is to represent individual behaviors and detect those which deviate from the majority. The method has been successfully applied to botnet detection in [16]. We are currently leveraging our graph analysis in order to provide to the community a new metric or distance to be applied when comparing port numbers. Indeed, numerical comparison is meaningless in that context and we could leverage either a semantic database (such as Wikipedia) or attacker database (darknet) to derive a meaningful metric, *i.e.* representing a real correlation between port numbers (TCP or UDP).

Furthermore, we continue our work on using Hidden Markov Models for analysing TCP scanning activities. We are now in a stage where individual models from different scanner tools or configurations (e.g. targeted ports) are used in order to automatically learn unique signatures then applied on non-labelled data.

7.2.2. NDN Security

Participants: Thibault Cholez [contact], Xavier Marchal, Olivier Festor, Jérôme François, Salvatore Signorello [University of Luxembourg], Radu State [University of Luxembourg], Samuel Marchal [Aalto University].

Information Centric Networking (ICN) is seen as a promising solution to re-conciliate the Internet usage with its core architecture. However, to be considered as a realistic alternative to IP, ICN must evolve from a pure academic proposition deployed in test environments to an operational solution in which security is assessed from the protocol design to its running implementation. Among ICN solutions, Named Data Networking (NDN), together with its reference implementation NDN Forwarding Daemon (NFD), acts as the most mature proposal but its vulnerability against the Content Poisoning Attack (CPA) is considered as a critical threat that can jeopardize this architecture. So far, existing works in that area have fallen into the pit of coupling a biased and partial phenomenon analysis with a proposed solution, hence lacking a comprehensive understanding of the attack's feasibility and impact in a real network. In a joint work with our colleagues from UTT and in the context of the ANR DOCTOR projet, we demonstrated through an experimental measurement campaign that CPA can easily and widely affect NDN. Our contribution is threefold: (1) we propose three realistic attack scenarios relying on both protocol design and implementation weaknesses; (2) we present their implementation and evaluation in a testbed based on the latest NFD version; and (3) we analyze their impact on the different ICN nodes (clients, access and core routers, content provider) composing a realistic topology. This work was published in IM 2017 conference [21].

Also, still in the context of the DOCTOR project, we refined our architecture to securely deploy NDN over NFV. Indeed, combining NFV fast service deployment and SDN fine grained control of data flows allows comprehensive network security monitoring. The DOCTOR architecture allows detecting, assessing and remediating attacks. NDN is an example of application made possible by SDN and NFV coexistence, since hardware implementation would be too expensive. We showed how NDN routers can be implemented and managed as VNFs. Security monitoring of the DOCTOR architecture is performed at two levels. First, host-level monitoring, provided by CyberCAPTOR, uses an attack graph approach based on network topology knowledge. It then suggests remediations to cut attack paths. We show how our monitoring tool integrates SDN and NFV specificities and how SDN and NFV make security monitoring more efficient. Then, application-level monitoring relies on the MMT probe. It monitors NDN-specific metrics from inside the VNFs and a central component can detect attack patterns corresponding to known flaws of the NDN protocol. These attacks are fed to the CyberCAPTOR module to integrate NDN attacks in attack graphs. This work was published in a book chapter "Guide to Security in SDN and NFV" from Springer's Computer Communications and Networks collection [35].

Finally, in cooperation with the University of Luxembourg, we have investigated interest flooding attacks in NDN. By nature, NDN communication assumes that requesting a content leads to emit an interest and forwarding it in the network until it reaches an appropriate content provider which then sends back data through the reverse path. Interest flooding attacks forge interests (requests) which cannot be satisfied by any data to be sent back to the emitter. As such, both the network and nodes are overloaded as the interests are flooded into the network and intermediate nodes have to store them locally in the pending interest table. We observed that most of literature mechanisms have been evaluated with very simple attack models. Actually, we had a great expertise in phishing attacks and social engineering that can be used to generate realistic phishing names for the NDN naming scheme. We thus create a new stealthy attack relying on natural language processing techniques to forge interests very similar to legitimate ones making inefficient all proposed counter-measures from the state-of-the-art [25].

7.2.3. Configuration security automation

Participants: Rémi Badonnel [contact], Abdelkader Lahmadi, Olivier Festor, Nicolas Schnepf, Maxime Compastié.

The main research challenge addressed in this work is focused on enabling configuration security automation in dynamic networks and services. In particular our objective is to support the efficient configuration and orchestration of security management operations.

The continuous growth and variety of networking significantly increases the complexity of management. It requires novel autonomic methods and techniques contributing to detection and prevention performances with respect to vulnerabilities and attacks.

We have pursued during Year 2017 the efforts on the orchestration of security functions in the context of mobile smart environments, with our joint work with Stephan Merz of the VeriDis project-team at Inria Nancy. We had already defined an automated verification technique, based on an extension of an SDN language, for checking both the control and the data planes related to security chains [24]. Complementarily, we proposed a strategy for generating SDN policies for protecting Android environments based on automata learning. Our solution collects traces of flow interactions of their applications, aggregates them in order to build finite-state models, and then infer SDN policy rules. We have designed and implemented aggregation and automata learning algorithms that allow precise and generic models of applications to be built. These models will be then used for configuring chains of security functions specified in the Pyretic language and verified with our Synaptic checker. We have developed a prototype of our solution implementing these algorithms, and evaluated its performances through a series of experiments based on the backend process miners Synoptic and Invarimint, in addition to our own algorithm. The experiments showed the benefits and limits of these methods in terms of simplicity, precision, genericity and expressivity, while varying the level of aggregation of the input flow traces.

In addition, we have worked on our software-defined security framework, for enabling the enforcement of security policies in distributed cloud environments. This framework relies on the autonomic paradigm to dynamically configure and adjust these mechanisms to distributed cloud constraints, and exploit the software-defined logic to express and propagate security policies to the considered cloud resources [13]. In particular, we have investigated during Year 2017 the exploitability of unikernels to support our framework. Unikernels permit to build highly-constrained configurations limited to the strict necessary with a time-limited validity. We take benefits of their properties to reduce the attack exposure of cloud resources. We have formalized and integrated into our software-defined security framework, on-the-fly generation mechanisms of unikernel images that cope with security policy requirements. In that context, security mechanisms are directly integrated to the unikernel images at building time. A proof of concept prototype based on MirageOS was developed and the performance of such a software-based security strategy was evaluated through extensive series of experiments. We have also compared them to other regular virtualization solutions. Our results show that the costs induced by security mechanisms integration are relatively limited, and unikernels are well suited to minimize risk exposure.

7.3. Experimentation, Emulation, Reproducible Research

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

7.3.1. Grid'5000 design and evolutions

Participants: Florent Didier, Arthur Garnier, Imed Maamria, 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.

7.3.1.1. Technical team management

Since the beginning of 2017, Lucas Nussbaum serves as the Grid'5000 *directeur technique* (CTO), managing the global technical team (9 FTE).

7.3.1.2. SILECS project

We are also heavily involved in the ongoing SILECS project, that aims at creating 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.

7.3.1.3. Promoting the testbed

In order to promote the testbed to the french devops and sysadmin community, we presented in [27] an overview of the testbed's capabilities.

7.3.1.4. Disk reservation

We contributed a new feature that will greatly help Big Data experimenters: the ability to reserve disks on nodes, in order to leave large datasets stored on nodes between nodes reservations.

7.3.1.5. Automated testing of the testbed

In order to ensure that all services remain functional, and that experimental results remain trustworthy and reproducible, we designed an infrastructure to automatically test the testbed and detect misconfigurations, regressions, uncontrolled hardware heterogeneity, etc. This work was described in [23] and later presented in [34].

7.3.1.6. Support for SDN experiments

We started the development of a tool to orchestrate SDN experiments on Grid'5000, combining KaVLAN and OpenVSwitch.

7.3.2. Emulation with Distem

Participants: Alexandre Merlin, Lucas Nussbaum [contact].

The ADT SDT project started in March. Initial work focused on improving the software developing infrastructure by adding automated regression tests on both correctness and performance. This should allow a new release in early 2018.

7.3.3. I/O access patterns analysis with eBPF

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

In the context of Abdulqawi Saif's CIFRE PhD (with Xilopix), we explored the relevance of an emerging instrumentation technology for the Linux kernel, eBPF, and used it to analyze I/O access patterns of two popular NoSQL databases. A publication on this topic is expected in early 2018.

7.3.4. Performance study of public clouds

Participants: Souha Bel Haj Hassine, Lucas Nussbaum [contact].

We worked on clouds performance in the context of an ongoing collaboration with *CloudScreener*, a French startup founded in 2012 that has developed tools for cloud price and performance benchmarks and automated cloud recommendation to optimize the decision making process in the context of cloud computing. We designed methods and tools to do performance evaluation of public clouds focusing on (1) outlining performance variability over time; (2) identifying adverse strategies that might be deployed by cloud providers in order to vary the performance level over time.

7.3.4.1. Testbeds federation and collaborations in the testbeds community

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'500 is a member.

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 gave two talks [33][34] and chaired the session on reproducibility.

7.3.4.2. Experimentation and reproducible research

In addition to the work already mentioned on testbed testing [23], [34], we worked on a survey of testbeds and their features for reproducible research [22]. We also gave several talks on reproducible research and testbeds at *École ARCHI* [5], *École RESCOM* [6], and Inria webinars on Reproducible Research [7].

7.4. Routing

7.4.1. NDN routing

Participants: Isabelle Chrisment [contact], Thomas Silverston, Elian Aubry.

As NDN relies on content names instead of host address it cannot rely on traditional Internet routing. Therefore it is essential to propose a routing scheme adapted for NDN. In [8] we have presented SRSC, our SDN-based Routing Scheme for CCN/NDN and its implementation. SRSC relies on the SDN paradigm. A controller is responsible to forward decisions and to set up rules into NDN nodes. So we have implemented SRSC into NDNx. We have deployed an NDN testbed within a virtual environment emulating a real ISP topology in order to evaluate the performances of our proposal with real-world experiments. We have demonstrated the feasibility of SRSC and its ability to forward Interest messages in a fully deployed NDN environment while keeping low overhead and computation time and high caching performances.

7.4.2. Energy-Aware and QoS Routing for Wireless Sensor Networks

Participants: Evangelia Tsiontsiou, Bernardetta Addis, Ye-Qiong Song [contact].

The main research problems in the domain of routing data packets in a multi-hop wireless sensor network are the optimisation of the energy and the routing under multi-criteria QoS constraints (e.g., energy, reliability, delay, ...). To address these problems, we proposed, in the PhD thesis of E. Tiontsiou, two contributions. The first contribution is an optimal probabilistic energy-aware routing protocol, allowing to energy usage balancing. Comparing to the existing probabilistic routing protocols, our solution is based on the computation of the optimal probabilities by solving a linear programming problem. Our second contribution is an operator calculus algebra based multi-constrained routing protocol. It is fundamentally different from the existing solutions since it can simultaneously consider several constraints, instead of their combination.

7.5. Smart*: design, multi-modeling and co-simulation and supervision of mobile CPS/IoT

Participants: Laurent Ciarletta [contact], Ye-Qiong Song, Yannick Presse, Julien Vaubourg, Emmanuel Nataf, Petro Aksonenko, Virgile Dauge, Louis Viard, Florian Greff, Virginie Galtier, Thomas Paris.

Vincent Chevrier (former Maia team, Dep 5, LORIA) is a collaborator and the correspondant for the MS4SG/MECSYCO project, as well as Christine Bourjot (former MAIA team, Dep 5, LORIA).

Sylvain Contassot-Vivier (Dep 3, Loria) is a collaborator on the Grone project and is directing Virgile Daugé with Laurent Ciarletta.

Pierre-Etienne Moreau is a collaborator on the CEOS project and is directing Louis Viard with Laurent Ciarletta.

Virginie Galtier from CentraleSupélec is now a member of the Loria laboratory and will integrate the future Simbiot team (Systems of Interactive aMBient Intelligent ObjecTs).

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

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

We proposed the AA4MM meta-model [45] that solves the core challenges of multimodeling and simulation coupling in an homogeneous perspective. In AA4MM, we chose a multi-agent point of view: a multi-model is a society of models; each model corresponds to an agent and coupling relationships correspond to interaction between agents. In the MECSYCO-NG (formerly MS4SG, Multi Simulation for Smart Grids) project which involves some members of the former MAIA team, Madynes and EDF R&D on smart-grid simulation, we developed a proof of concepts for a smart-appartment case that serves as a basis for building up use cases, and we have worked on some specific cases provided by our industrial partner. We also collaborated with researchers from the Green UL laboratory.

In 2017 we worked on the following research topics:

- Overall assessment and evaluation of complex systems.
- Cyber Physical Systems and Smart *.

  We have continued the design and implementation of the Aetournos platform at Loria which will be part of the Creativ'Lab. The collective movements of a flock of flying communicating robots / UAVs, evolving in potentially perturbed environment constitutes a good example of a Cyber Physical System. Several projects have started during the last part of 2017. One of the emerging topic in this area is the safety of Mobile IoT/CPS with regards to their environment and users.

  – The Grone (Interreg) project involves partners from the 4 countries of the Grande Région (Centrale Supélec, LIST, Univ Luxembourg, Univ Liège, Fraunhofer IZFP to name a few). The main goal is to develop UAV based solution for the surveillance of industrial and agricultural sites and the exploration of GPS denied and underground environments. A PhD has been started in March 2017 (Systèmes cyber-physiques autonomes et communicants en milieux hostiles. Application à l'exploration par robots mobiles. Virgile Daugé).

  – and the CEOS (FUI22) project involving high profile companies (Thales C&S, EDF R&D, ENEDIS and Aéroport de Lyon) as well as academic partners (AOSTE2 Inria, ESIEE) and collaborating SMEs (RT@W, ADCIS, Alerion). This project focuses on the safety of UAV based monitoring solutions for OIV (Opérateurs d'Intéret Vital) infrastructures. A PhD has been started in November 2017 (Environnement de développement et d'analyse de propriétés pour des systèmes cyber-physiques mobiles. Louis Viard).

  The work on Software Defined Real-Time Mesh networks (Florian Greff's PhD CIFRE with Thales R&T) has given many results as he plans to defend his work in march 2018 [15], [39], [14].

On more specific subject of innovative sensors for mobile and interactive IoT, a collaborative project with the KPI (Ukraine) university has been started with a projected PhD (Méthodes optimisées de calibration, d'alignement et algorithmes d'attitude avancés pour les systèmes de navigation inertiels fixés, Petro Aksonenko). Several papers have been published [9] [44] in 2017.

- (Very Serious) Gaming: Starburst Gaming. During some exploratory work, we have seen the potential of these Pervasive Computing ressources in the (Very Serious) Gaming area which led us to the Starburst Computing SATT projects in 2016 and 2017. A spin off has been founded in 2017 that is getting the licences for the resulting IP (the software is under the APP process at the time of this writing). Starburst is already involved in a AMI project with the Globlinz game studio and the lab and has officially been accepted in novembre 2017 and will be operational in 2018.
- Smart *: MS4SG / MECSYCO-NG has given us the opportunity to link simulations tools with a strong focus on FMI (Functional Mockup Interface) and network simulators (NS3/Omnet++). We have so far successfully applied our solution to the simulation of smart apartment complex and to combine the electrical and networking part of a Smart Grid. The AA4MM software is now MECSYCO and has seen constant improvements in 2017 thanks to the ressources provided by the MECSYCO-NG project in collaboration with EDF R&D (http://www.mecsyco.com), and the work of Thomas Paris and Julien Vaubourg.

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

The technical report [43] has been extended into a journal paper under revision for a publication in 2018.

7.6. Quality-of-Service

7.6.1. Self-adaptive MAC protocol for both QoS and energy efficiency in IoT

Participants: Shuguo Zhuo, François Despaux, Ye-Qiong Song [contact].

The diversity of IoT applications implies the requirement of reliable yet efficient MAC solutions for supporting transmissions for various traffic patterns. We have mainly contributed to enhance the implementation of the high efficient traffic self-adaptive MAC protocols. As part of RIOT ADT project, our main achievements are the development of two MAC protocols lw-MAC and GoMacH [26]. lw-MAC is similar to X-MAC and ContikiMAC. It allows to introduce a first duty-cycled MAC into RIOT IoT protocol stack. GoMacH is a nearly optimal protocol that provides high reliability and throughput for handling various traffic loads in IoT. GoMacH seamlessly integrates several outstanding techniques. It adopts the phase-lock scheme to achieve low-power duty-cycled communication. It also utilizes a dynamic slots allocation scheme for providing accurate and instantaneous throughput boost. Furthermore, like in TSCH, GoMacH spreads its communications onto IEEE 802.15.4's 16 channels, leading to high reliability. GoMacH has been implemented in open source on RIOT OS, and has also been seamlessly integrated into IETF's 6LoWPAN/RPL/UDP stack as well as CCN-light. Experimental results on SAMR21-xpro test-beds and IoT-LAB verify the practicality of GoMacH and its capabilities for consistently providing high throughput, high delivery ratio, and low radio duty-cycle. They are both publically available on the RIOT open source github.

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

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

The QoS must be guaranteed when dealing with real-time distributed systems interconnected by a network. Not only task schedulability in processors, but also message schedulability in networks should be analyzed for validating the system design. Fault-tolerance is another critical issue that one must take into account.

In collaboration with Thales TRT industrial partner as part of a CIFRE PhD work, we have developed a Software-Defined Real-time Network (SDRN) framework [14]. SDRN deals with the real-time flow allocation problem in mesh networks. The objective is to find a suitable path under delay constraint while allowing load balancing. For this purpose, combined online flow admission control and pathfinding algorithms have been developed on an SDN-like controller. At switch level, each output port is ruled by a credit-based weighted round robin, allowing isolation of flows. As a consequence, a freshly admitted flow will not influence existing flows, allowing incremental online admission of new flows. This approach has been applied to a RapidIO mesh network example and compared with the compositional performance analysis method. Numerical results clearly show the benefit of our proposal in terms of complexity and delay bound pessimism. In [15], Fault-tolerance issue in mesh networks has been addressed. In fact, one of the major advantages of a mesh topology is its ability to leverage the path redundancy in order to recover from link or node failures, through a flow reconfiguration process. However, one needs to ensure that hard real-time packets will keep being delivered on time during this transient reconfiguration period. Anticipating each possible fault is very complex and can result in a waste of network resource. Our contribution is the combination of an optimized content-centric source routing in nominal mode and a destination-tag flexible and scalable routing in transient recovery mode. We show the benefit of this approach in terms of flexibility and network resource utilization. Our method can ensure real-time properties enforcement even during the transient reconfiguration period. Algorithms have been developed to extend the SDRN flow allocation and routing methods in order to implement this hybrid fault-tolerant extension.

As part of Eurostars RETINA project, in the in-vehicle networking domain, we have focused on the evaluation of the worst-case response time of AVB traffic under time-aware shaper of TSN (time-sensitive networking). It is a hierarchical real-time scheduling problem, where a packet is scheduled by the credit-based shaper, priority and time-aware shaper (TDMA). We have proved that the eligible interval approach, developed for AVB, is still hold for TSN case. The worst case delay expression, as well as the feasibility condition are deduced. Our methods (analysis and simulation) are applied to an automotive use case, which is defined within the Eurostars RETINA project, and where both control data traffic and AVB traffic must be guaranteed. It has been shown that our delay bound is tight in single switch case [19].

ALICE Project-Team

7. New Results

7.1. Geometric foundations

7.1.1. Detecting the Intersection of Two Convex Shapes by Searching on the 2-sphere

Participant: Samuel Hornus.

We take a look at the problem of deciding whether two convex shapes intersect or not. We do so through the well known lens of Minkowski sums and with a bias towards applications in computer graphics and robotics. We describe a new technique that works explicitly on the unit sphere, interpreted as the sphere of directions. In extensive benchmarks against various well-known techniques, ours is found to be slightly more efficient, much more robust and comparatively easy to implement. In particular, our technique is compared favourably to the ubiquitous algorithm of Gilbert, Johnson and Keerthi (GJK), and its decision variant by Gilbert and Foo. We provide an in-depth geometrical understanding of the differences between GJK and our technique and conclude that our technique is probably a good drop-in replacement when one is not interested in the actual distance between two non-intersecting shapes.

The work was published in the journal Computer-Aided Design (special issue for the Proceedings of Solid and Physical Modeling: SPM'17) [9]. The paper has received a best paper award (2nd place) at the SPM conference.

7.1.2. Decomposition of a Hexahedron into a Set of Tetrahedra

Participant: Laurent Alonso.

This year was marked by some works on the combinatorial decomposition of a generic hexahedron in a set of nonintersecting tetrahedra up to symmetries: it is well known that there are only six decompositions of a cube into tetrahedra ; we show that there are at most 1360 potential different decompositions of a hexahedron and at least 1357 are geometrically valid. Additional work is in progress in order to show that the last 3 remaining decompositions do not correspond to valid geometrical solutions.

7.1.3. Hash-based CSG Evaluation on GPU

Participants: Cédric Zanni, Sylvain Lefebvre.

We have developed a new evaluation scheme for Constructive Solid Geometry (CSG) modeling that is well adapted to modern GPU. The approach falls into the category of screen space techniques and can handle a large range of geometric representation. 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 to subdivide the space in order to apply progressively the rendering algorithm, ensuring that required data fit in the graphic memory. This improvement over previous approach allows to handle objects of higher complexity during both modeling and slicing for additive manufacturing.

7.2. Geomety processing

7.2.1. Hexahedral Meshes: Generation, Simulation, Evaluation

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

We continued working on the generation of the so-called hexahedral (or hexahedral-dominant) meshes. It is believed that these meshes are more efficient (both in terms of required space and computational time) for certain physics and numerical simulation. However, they are much more difficult to generate, and no fully automatic method currently exists. It is a huge problem in the industry, that uses days/weeks/months of manual work to generate them, because they are preferred in certain domains (fluids, mechanics, wave propagation). In this research, we aim at answering the following questions:

- How can we generate a hexahedral (or hex-dominant) mesh in a fully automatic manner?
- How can we evaluate the quality of this mesh, suitability for numerical simulation?

In the context of Maxence Reberol's Ph.D. thesis, we developed new algorithms to answer both questions. To answer the first question, in [17], building on our previous results based on global parameterization, we proposed a method to mesh the globality of the domain, by isolating the singular zones of the parameterization and meshing them with a separate algorithm. To answer the second question, in [16], we proposed a new method to estimate the distance between two Finite Element simulations obtained from two different computational meshes / function bases. We started using our algorithm to compare the rate of convergence of the method as a function of element size h with different PDEs (Poisson, linear elasticity) using different function bases (tetrahedral: P1, P2, P3, hexahedral: Q1, Q2, Q3).

7.2.2. Surface Reconstruction

Participants: Dobrina Boltcheva, Bruno Lévy.

We developed a new algorithm [7] for surface reconstruction. Our algorithm is equivalent to Delaunay-based reconstruction, it computes the Delaunay triangulation restricted to an object computed from the input pointset. The object is a set of disks centered on the input points and perpendicular to estimated normals. The algorithm is fast and memory efficient, because the only used global data structure is a Kd-tree. Applications are demonstrated with a parallel implementation on a multicore processor, and a version for hand-held devices.

7.2.3. Geometric Algorithms for 3D modeling in Geo-sciences

Participants: Bruno Lévy.

We developed RingMesh [13], an application layer around our Geogram library, specialized for 3D modeling in Geo-sciences. The RingMesh library uses the mesh data structure and basic algorithms in Geogram to offer 3D modeling primitives well-suited to geosciences. It has datatypes to efficiently represent complicated 3D models of the underground, with the topological relationships between the interfaces (horizons and faults), as well as interfaces to 3D mesh generation softwares.

7.3. Additive Manufacturing

7.3.1. Iterative Carving for Self-supporting 3D Printed Cavities

Participants: Samuel Hornus and Sylvain Lefebvre.

This work explores the printing of shapes with as little material as possible, mostly with a view toward minimizing fabrication time for large pieces. In particular, it aims at modeling a structure of thin sheets inside a volume in such a way that the sheets and the boundary of the volume can be 3D-printed as is, without internal support.

The work is an adaptation of the technique developed earlier fo other applications in 3D printing and achieved state-of-the-art results. It is available as an Inria technical report [14].

7.3.2. Optimal Discrete Slicing

Participant: Sylvain Lefebvre.

This work is a collaboration with Marc Alexa and Kristian Hildebrand from TU Berlin. We developed a novel algorithm to compute the optimal decomposition of a 3D shape into layers of varying thickness, in a discrete setting. This answers a long standing problem in additive manufacturing. Our approach computes all optimal solutions for any number of slices by formulating the optimization as a dynamic programming problem. We developed efficient algorithms for both computing the geometric errors within each slice (based on volume difference) as well as for the optimizer. Our technique is the first to provide a provably optimal result and outperforms all existing heuristics. The work has been published in ACM TOG [5], presented at SIGGRAPH 2017 and is fully implemented within IceSL, available for public download.

7.3.3. Fabricable Tile Decors

Participants: Sylvain Lefebvre and Jonàs Martínez.

We propose a modeling technique to produce large objects whose surface is composed of user-provided decorative tiles. Such objects are very inefficient to 3D print as they occupy a large volume while in fact using little material. On low end printers they require large amounts of support structures which are difficult to remove. We propose a decomposition of the input shape into sets of planar patches that can print flat and can be later assembled into stable structures. This work is a collaboration with Hong Kong University in the context of the PrePrint3D associated team. It was published in ACM TOG [8] and presented at SIGGRAPH Asia 2017.

7.3.4. Visualizing and Fabricating Complex Internal Structures

Participant: Sylvain Lefebvre.

This work considers efficient display and manufacturing of extremely detailed internal structures described by implicit (procedural) indicator functions [15]. We describe a technique for their progressive rendering when the structures fill an envelope provided as a 3D mesh. We also describe how to efficiently extract slices for additive manufacturing, in a process that is both computationally and memory efficient. This work was presented at the Visual Analytics conference (Moscow, 2017) and is under submission to the Scientific Visualization journal.

7.3.5. Orthotropic k-nearest Foams for Additive Manufacturing

Participants: Jonàs Martínez, Haichuan Song, Jérémie Dumas, Sylvain Lefebvre.

We proposed a novel metamaterial with controllable, freely orientable, orthotropic elastic behavior – orthotropy means that elasticity is controlled independently along three orthogonal axes, which leads to materials that better adapt to uneven, directional load scenarios, and offer a more versatile material design primitive. The fine-scale structures are generated procedurally by a stochastic process, and resemble a foam. This work has been published in ACM TOG [12], and presented at SIGGRAPH 2017.

7.3.6. Color Fused Filament Fabrication

Participants: Haichuan Song, Sylvain Lefebvre.

Traditional filament printers cannot truly reproduce colored objects. The best current techniques rely on a form of dithering exploiting occlusion, that was only demonstrated for shades of two base colors and that behaves differently depending on surface slope. We explored a novel approach for 3D printing colored objects, capable of creating controlled gradients of varying sharpness. Our technique exploits off-the-shelves nozzles that are designed to mix multiple filaments in a small melting chamber, obtaining intermediate colors once the mix is stabilized. The key idea is to divide each input layer into a set of sublayers, each having a different constant color. By locally changing the thickness of the sublayers, we change the color that is perceived at a given location. By optimizing the choice of colors of each sublayer, we further improve quality and allow the use of different numbers of input filaments. We demonstrate our results by building a functional color printer using low cost, off-the-shelves components. Using our tool a user can paint a 3D model and directly produce its physical counterpart, using any material and color available for fused filament fabrication. This work has been submitted and is available at [18].

7.3.7. Anti-aliasing for Fused Filament Deposition

Participants: Haichuan Song, Sylvain Lefebvre.

Layered manufacturing inherently suffers from staircase defects along surfaces that are gently sloped with respect to the build direction. Reducing the slice thickness improves the situation but also largely increases the print time. We proposed a simple yet effective technique to improve the print accuracy for layered manufacturing by filament deposition. It better reproduces the geometry of sloped surfaces without increasing the print time. The key idea is to perform a local anti-aliasing, working at a sub-layer accuracy to produce slightly curved deposition paths and reduce approximation errors. We further split and order paths to minimize defects due to the extruder nozzle shape, avoiding any change to the existing hardware. We apply and analyze our approach on 3D printed examples, showing that our technique greatly improves surface accuracy and silhouette quality while keeping the print time nearly identical. This work has been published in the Computer Aided Design (CAD) journal [19].

LARSEN Project-Team

7. New Results

7.1. Lifelong Autonomy

7.1.1. Sensorized environment

7.1.1.1. Localisation of Robots on a Load-sensing Floor

Participants: François Charpillet, Francis Colas, Vincent Thomas.

The use of floor-sensors in ambient intelligence contexts began in the late 1990's. We designed such a sensing floor in Nancy in collaboration with the Hikob company (http://www.hikob.com) and Inria SED. This is a load-sensing floor which is composed of square tiles, each equipped with two ARM processors (Cortex M3 and A8), 4 load cells, and a wired connection to the four neighboring cells. Ninety tiles cover the floor of our experimental platform (HIS).

This year, with Aurelien Andre (master student from Univ. Lorraine), we have focused on tracking robots on several scenarios based on data originated from the sensing tiles and collected the previous years. We have proposed a new approach to build relevant clusters of tiles (based on connexity). For single robot scenarios, we have focused on basic algorithms (for instance, Kalman filter) and on Probability Data Association Filter to consider the possibility of false positive in the bayesian filter. Then, for multi-target tracking, we have investigated elaborate strategies to associate atomic measures to the tracked targets like JPDAF (Joint Probability Data Association Filter algorithm [58]) and JPDAMF (Joint Probability Data Association Merged Filter [45]) in order to consider measures resulting from several targets.

7.1.1.2. High Integrity Personal Tracking Using Fault Tolerant Multi-Sensor Data Fusion

Participants: François Charpillet, Maan Badaoui El-Najjar.

Maan Badaoui El Najjar is professor at university of Lille and he is the head of the DiCOT Team "Diagnostic, Control and Observation for fault Tolerant Systems" of the CRIStAL Laboratory.

The objective of this PhD work is to study the possibilities offered by the above mentioned load-sensing floor. The idea is to combine the information from each sensor (load sensors and accelerometers) to identify daily living activities (walking, standing, lying down, sitting, falling) and to create a positioning system for the person in the apartment. The approach is based on information theory to address the detection of outliers during the fusion process. This is based on informational filters and fault detection to identify and eliminate faulty measurements. This work was carried through the PhD Thesis of Mohamad Daher under the supervision of François Charpillet and Maan Badaoui El Najjar. This thesis was defended at university of Lille on the 13th December 2017.

Publication: [14]

7.1.1.3. Active Sensing and Multi-Camera Tracking

Participants: François Charpillet, Vincent Thomas.

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.

This problem we are focusing on consists of following the trajectories of persons with the help of several controllable cameras in the smart environment. The approach we are working on is based on probabilistic decision processes in partial observability (POMDP - Partially Observable Markov Decision Processes) and particle filters. In the past, we have proposed an original formalism rho-*POMDP* and new algorithms for representing and solving active sensing problems [43] by tracking several persons with fixed camera based on particle filters and Simultaneous Tracking and Activity Recognition approach [49].

This year, approaches based on Monte-Carlo Tree Search algorithms (MCTS) like POMCP [60] have been used to build policies for following a single person with several controllable cameras in a simulated environment.

7.1.2. Partially Observable Markovian Decision Processes (POMDP)

7.1.2.1. Solving ρ-POMDP using Lipschitz Properties

Participant: Vincent Thomas.

We are currently investigating how to solve continuous MDP and ρ-POMDP by using Lipschitz property (rather than classical Piecewise Linear and Convex property used to solve POMDP). We have proven that if the transition and reward functions are lipschtiz-continuous, the value function has the same property.

With Mathieu Fehr (Ulm ENS student), we have studied new algorithm based on HSVI (Heuristic Search Value Iteration [62]) to take advantage of the lipschitz continuity property. The properties of these algorithms are currently investigated.

7.1.3. Distributed Exploration of an Unknown Environment by a Swarm of Robots

Participants: Nassim Kalde, François Charpillet, Olivier Simonin.

Olivier Simonin is Professeur at INSA Lyon and is the scientific leader of Chroma Team.

In this PhD, we have explored the issue for a team of cooperating mobile robots to intelligently explore an unknown environment. This question has been addressed both in the framework of sequential decision making and frontier based exploration. Considered environments includes static or populated environments.

This work was carried through the PhD Thesis of Nassim Fates under the supervision of François Charpillet and Olivier Simonin. This thesis was defended on the 12th December 2017.

7.1.4. Robot Learning

7.1.4.1. Black-box Data-efficient RObot Policy Search (Black-DROPS)

Participants: Konstantinos Chatzilygeroudis, Dorian Goepp, Rituraj Kaushik, Jean-Baptiste Mouret.

The most data-efficient algorithms for reinforcement learning (RL) in robotics are based on uncertain dynamical models: after each episode, they first learn a dynamical model of the robot, then they use an optimization algorithm to find a policy that maximizes the expected return given the model and its uncertainties. It is often believed that this optimization can be tractable only if analytical, gradient-based algorithms are used; however, these algorithms require using specific families of reward functions and policies, which greatly limits the flexibility of the overall approach. We introduced a novel model-based RL algorithm [23], called Black-DROPS (Black-box Data-efficient RObot Policy Search), that: (1) does not impose any constraint on the reward function or the policy (they are treated as black-boxes), (2) is as data-efficient as the state-of-the-art algorithm for data-efficient RL in robotics, and (3) is as fast (or faster) than analytical approaches when several cores are available. The key idea is to replace the gradient-based optimization algorithm with a parallel, black-box algorithm that takes into account the model uncertainties. We demonstrate the performance of our new algorithm on two standard control benchmark problems (in simulation) and a low-cost robotic manipulator (with a real robot).

Publications: [23]

7.1.4.2. Reset-free Data-efficient Trial-and-error for Robot Damage Recovery

Participants: Konstantinos Chatzilygeroudis, Jean-Baptiste Mouret, Vassilis Vassiliades.

The state-of-the-art RL algorithms for robotics require the robot and the environment to be reset to an initial state after each episode, that is, the robot is not learning autonomously. In addition, most of the RL methods for robotics do not scale well with complex robots (e.g., walking robots) and either cannot be used at all or take too long to converge to a solution (e.g., hours of learning). We introduced a novel learning algorithm called "Reset-free Trial-and-Error" (RTE) that (1) breaks the complexity by pre-generating hundreds of possible behaviors with a dynamics simulator of the intact robot, and (2) allows complex robots to quickly recover from damage while completing their tasks and taking the environment into account [13]. We evaluated our algorithm on a simulated wheeled robot, a simulated six-legged robot, and a real six-legged walking robot that are damaged in several ways (e.g., a missing leg, a shortened leg, faulty motor, etc.) and whose objective is to reach a sequence of targets in an arena. Our experiments show that the robots can recover most of their locomotion abilities in an environment with obstacles, and without any human intervention.

Publications: [13]

7.1.5. Illumination & Quality Diversity Algorithms

7.1.5.1. Using Centroidal Voronoi Tessellations to Scale up the MAP-Elites Algorithm

Participants: Konstantinos Chatzilygeroudis, Jean-Baptiste Mouret, Vassilis Vassiliades.

The MAP-Elites algorithm [55] is a key step of our "Intelligent Trial and Error" approach [46] for data-efficient damage recovery. It works by discretizing a continuous feature space into unique regions according to the desired discretization per dimension. While simple, this algorithm has a main drawback: it cannot scale to high-dimensional feature spaces since the number of regions increase exponentially with the number of dimensions. We addressed this limitation by introducing a simple extension of MAP-Elites that has a constant, pre-defined number of regions irrespective of the dimensionality of the feature space [21]. Our main insight is that methods from computational geometry could partition a high-dimensional space into well-spread geometric regions. In particular, our algorithm uses a centroidal Voronoi tessellation (CVT) to divide the feature space into a desired number of regions; it then places every generated individual in its closest region, replacing a less fit one if the region is already occupied. We demonstrated the effectiveness of the new "CVT-MAP-Elites" algorithm in high-dimensional feature spaces through comparisons against MAP-Elites in maze navigation and hexapod locomotion tasks.

Publications: [21], [37], [38]

7.1.5.2. Aerodynamic Design Exploration through Surrogate-Assisted Illumination

Participants: Adam Gaier, Jean-Baptiste Mouret.

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 a single, near-optimal solution. Unfortunately, these algorithms currently require a large number of function evaluations, limiting their applicability. In our recent work [27], [26], we introduced a new illumination algorithm, called Surrogate-Assisted Illumination (SAIL), that creates a map of the design space according to user-defined features by leveraging surrogate modeling and intelligent sampling to minimize the number of 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 [55] or CMA-ES [52]. As shown in this article, SAIL can also produce maps of high-performing designs in a more realistic 3-dimensional aerodynamic task with an accurate flow simulation. Overall, SAIL can help designers understand what is possible, beyond what is optimal, by considering more than pure objective-based optimization.

Publications: [27], [26]

7.1.6. Applications – civil robotics

7.1.6.1. Minimally Invasive Exploration of Heritage Buildings

Participants: Jean-Baptiste Mouret, Lucien Renaud, Kapil Sawant.

In 2017, the team officially joined the ScanPyramids mission, which aims at better understanding how the pyramids of the Old Kingdom were built, but also to encourage innovations in various fields (muography, virtual reality, simulation, ...) that could be useful for the pyramids as well as for other monuments. The ScanPyramids team has discovered several previously unknown voids in the pyramid of Cheops, one of them with a size similar to the one of the Grand Gallery, called « ScanPyramids' Big Void ».

We participated to the article about the ScanPyramids' Big Void [17] and we designed several prototypes for minimally invasive exploration. We envision exploration to take place in two stages. At first, a tubular robot fitted with an omnidirectional camera would be inserted to take high-resolution pictures of the inaccessible place. In a second stage, the team would use the same hole to send an exploration robot operated remotely to travel through corridors and help mapping the interior. For this second step, we are currently designing a miniature blimp that would be folded during the insertion, then remotely inflated once in the inaccessible place. When the exploration is over, the blimp would come back to its base, be deflated, then extracted from the insertion hole.

Publications: [17]

7.1.7. Humanoid Robotics

7.1.7.1. Trial-and-error Learning of Repulsors for Humanoid QP-based Whole-Body Control

Participants: Karim Bouyarmane, Serena Ivaldi, Jean-Baptiste Mouret, Jonathan Spitz, Vassilis Vassiliades.

Whole body controllers based on quadratic programming allow humanoid robots to achieve complex motions. However, they rely on the assumption that the model perfectly captures the dynamics of the robot and its environment, whereas even the most accurate models are never perfect. We introduced a trial-and-error learning algorithm that allows whole-body controllers to operate in spite of inaccurate models, without needing to update these models [35]. The main idea is to encourage the controller to perform the task differently after each trial by introducing repulsors in the quadratic program cost function. We demonstrated our algorithm on (1) a simple 2D case and (2) a simulated iCub robot for which the model used by the controller and the one used in simulation do not match.

Publications: [35]

7.1.7.2. Safe Trajectory Optimization for Whole-body Motion of Humanoids

Participants: Serena Ivaldi, Valerio Modugno.

Multi-task prioritized controllers generate complex behaviors for humanoids that concurrently satisfy several tasks and constraints. In our previous work we automatically learned the task priorities that maximized the robot performance in whole-body reaching tasks, ensuring that the optimized priorities were leading to safe behaviors. Here, we take the opposite approach: we optimize the task trajectories for whole-body balancing tasks with switching contacts, ensuring that the optimized movements are safe and never violate any of the robot and problem constraints. We use (1+1)-CMA-ES with Constrained Covariance Adaptation as a constrained black box stochastic optimization algorithm, with an instance of (1+1)- CMA-ES for bootstrapping the search. We apply our learning framework to the prioritized whole-body torque controller of iCub, to optimize the robot's movement for standing up from a chair.

Publications: [29]

7.1.7.3. Humanoid Robot Fall Control

Participant: Karim Bouyarmane.

Falling is a major skill to be mastered by an autonomous humanoid robot, since no matter what balance controller we use, a humanoid robot will end up falling in certain circumstances. We proposed new approaches to control humanoid robots in general fall configurations and in general cluttered environment. From fall detection instant, a pre-imapct phase is triggered where a real-time configuration adaptation routine makes the robot quickly analyze the surrounding environment, choose best impact points on the environment, and adapts its configuration accordingly to meet the desired impact points (all calculations performed in the short duration of 0.7s to 1s that the fall lasts). Then right after impact a real-time motor PD gain adaptation controller

allows to set the right values for the gains in real-time to comply actively with the impact while minimizing peak torque at impact. Finally, a model-predictive approach combined with a novel formulation of admissible force polytopes accounting for both torque limits and Coulomb friction limitation ensures that the robot safely comes to a steady-state resting state at the end of the fall.

Publications: [41], [34], [33]

7.1.7.4. Stability Proof of Weighted Multi-Task Humanoid QP Controller

Participant: Karim Bouyarmane.

We proved that weighted multi-task controllers are locally exponentially stable under appropriate conditions of the task gain matrices. We also derived a number of stability properties of the underlying QP optimization problem.

Publications: [12]

7.1.7.5. Theoretical Study of Commonalities between Locomotion and Manipulation in Humanoid-like Locomotion-and-manipulation Integration System

Participant: Karim Bouyarmane.

We published our theoretical study on common ground formulations of locomotion and manipulation, and thereby their extension to integrated locomotion-and-manipulation systems, by analytically deriving their planning and control solutions in low-dimensional proof-of-concept examples based on nonlinear control and differential geometry tools.

Publications: [11]

7.1.8. Embodied Evolutionary Robotics

7.1.8.1. Online Distributed Learning for a Swarm of Robots

Participants: Iñaki Fernández Pérez, Amine Boumaza, François Charpillet.

We study how a swarm of robots adapts over time to solve a collaborative task using a distributed Embodied Evolutionary approach, where each robot runs an evolutionary algorithm and locally exchange genomes and fitness values. Particularly, we study a collaborative foraging task, where the robots are rewarded for collecting food items that are too heavy to be collected individually and need at least two robots to be collected. Furthermore, to promote collaboration, agents must agree on a signal in order to collect the items. Our experiments show that the distributed algorithm is able to evolve swarm behavior to collect items cooperatively. The experiments also reveal that effective cooperation is evolved due mostly to the ability of robots to jointly reach food items, while learning to display the right color that matches the item is done suboptimally. However, a closer analysis shows that, without a mechanism to avoid neglecting any kind of item, robots collect all of them, which means that there is some degree of learning to choose the right value for the color effector depending on the situation.

This work was carried through the PhD Thesis of Iñaki Fernández Pérez under the supervision of François Charpillet and Amine Boumaza. This thesis was defended on the 19th December 2017.

Publications: [25]

7.1.8.2. Phylogeny of Embodied Evolutionary Robotics

Participant: Amine Boumaza.

We explore the idea of analyzing Embodied Evolutionary Robotics from the perspective of genes and their dynamics using phylogenetic trees. We illustrate a general approach on a simple question regarding the dynamics of the fittest and most copied genes as an illustration using tools from spectral graph theory or computational phylogenetics, and argue that such an approach may give interesting insights on the behavior of these algorithms. This idea seems promising and further investigations are underway, especially on the links with coalescence theory.

Publications: [22]

7.2. Natural Interaction with Robotics Systems

7.2.1. Control of Interaction

7.2.1.1. Towards Human-aware Whole-Body Controllers for Physical Human-Robot Interaction

Participants: Oriane Dermy, Serena Ivaldi.

The success of robots in real-world environments is largely dependent on their ability to interact with both humans and said environment. The FP7 EU project CoDyCo focused on the latter of these two challenges by exploiting both rigid and compliant contacts dynamics in the robot control problem. Regarding the former, to properly manage interaction dynamics on the robot control side, an estimation of the human behaviours and intentions is necessary. We contributed to the building blocks of such a human-in-the-loop controller, and validate them in both simulation and on the iCub humanoid robot for the final demo of the CoDyCo project where a human assists the robot in standing up from being seated on a bench.

The controller is the basis for our current researches in the AnDy project.

Publications: [20]

7.2.1.2. Generating Motions for a Humanoid Robot that Assists a Human in a Co-manipulation Task

Participants: Karim Bouyarmane, Kazuya Otani, Serena Ivaldi.

We proposed a method to make a humanoid robot adapt its motion to help a human collaborator in simulation realize a collaborative manipulation task with the robot while the robot figures out its configuration in real-time through symmetric retargeting.

Publications: [40]

7.2.1.3. Human-to-humanoid Motion Retargeting

Participants: Karim Bouyarmane, Kazuya Otani.

We continue the development of our human-to-humanoid motion retargeting method by extending it to whole-body manipulation motions based on our previously-proposed multi-robot QP paradigm. The motion retargeting system is now able to autonomously adapt the motion of the robot to dynamics parameters of the manipulated object that substantially differ from those used to provide the human demonstration.

Publications: [31]

7.2.2. Non-verbal Interaction

7.2.2.1. Multimodal Prediction of Intention via Probabilistic Movement Primitives (ProMP)

Participants: François Charpillet, Oriane Dermy, Serena Ivaldi.

We designed a method for predicting the intention of a user interacting (physically or not) with the humanoid robot iCub, and implemented an associated open-source software (cf. ProMP_iCub in the Software section). Our goal is to allow the robot to infer the intention of the human partner during collaboration, by predicting the future intended trajectory: this capability is critical to design anticipatory behaviors that are crucial in human–robot collaborative scenarios, such as in co-manipulation, cooperative assembly, or transportation. We propose an approach to endow the iCub with basic capabilities of intention recognition, based on Probabilistic Movement Primitives (ProMPs), a versatile method for representing, generalizing, and reproducing complex motor skills. The robot learns a set of motion primitives from several demonstrations, provided by the human via physical interaction. During training, we model the collaborative scenario using human demonstrations. During the reproduction of the collaborative task, we use the acquired knowledge to recognize the intention of the human partner. Using a few early observations of the state of the robot, we can not only infer the intention of the partner but also complete the movement, even if the user breaks the physical interaction with the robot. We evaluated our approach both in simulation and with the real iCub robot. We also proposed a method to exploit referential gaze and combine it with physical interaction, to improve the prediction of primitives. The software implementing our approach is open source and available on the GitHub platform. In addition, we provide tutorials and videos.

Publications: [15]

7.2.2.2. PsyPhINe: Cogito Ergo Es

Participant: Amine Boumaza.

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

This year an experimental campaigns was organized in which around 40 test subjects where asked to solve a task in front of the moving robotic device. These interactions were recorded on video along with eye tracking data. To analyze the data, a web application was created that crowd-sources video annotation to internet users. A preliminary analysis of the data was presented at the third edition of the PsyPhINe workshop organized by the group, gathering top researchers from philosophy, anthropology, psychology and computer science to discuss and exchange on our methodology (see 10.1.1.1).

7.2.2.3. Active Audio Source Localization

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

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

We considered, here, the task of audio source localization using a microphone array on a mobile robot. Active localization algorithms have been proposed in the literature that can estimate the 3D position of a source by fusing the measurements taken for different poses of the robot. However, the robot movements are typically fixed or they obey heuristic strategies, such as turning the head and moving towards the source, which may be suboptimal. This work proposes an approach to control the robot movements so as to locate the source as quickly as possible using the Monte-Carlo Tree Search algorithm [30]. We represent the belief about the source using our mixture Kalman filter that explicitly includes the discrete activity of the source in the estimated state vector, alongside the continuous states such as the position of the robot or the sound source.

This work was carried through the PhD Thesis of Van Quan Nguyen under the supervision of Emmanuel Vincent and Francis Colas. This thesis was defended on the 3rd November 2017.

Publication: [30]

MAGRIT Project-Team

7. New Results

7.1. Matching and localization

Participants: Marie-Odile Berger, Vincent Gaudilliere, Antoine Fond, Pierre Rolin, Gilles Simon, Frédéric Sur.

Pose initialization

Estimating the pose of a camera from a model of the scene is a challenging problem when the camera is in a position not covered by the views used to build the model, because feature matching is difficult in such a situation. Several viewpoint simulation techniques have been recently proposed in this context. They generally come with a high computational cost, are limited to specific scenes such as urban environments or object-centered scenes, or need an initial guess for the pose. In his PhD thesis [12], P. Rolin has proposed a viewpoint simulation method well suited to most scenes and query views. Two major problems have been addressed: the positioning of the virtual viewpoints with respect to the scene, and the synthesis of geometrically consistent patches. Experimental results showed that patch synthesis dramatically improves the accuracy of the pose in case of difficult registration, with a limited additional computational cost.

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. we are interested in VP detection from uncalibrated monocular images. As any two parallel lines intersect in a VP, grouping line segments is a difficult problem that often yields a large number of spurious VPs. However, many tasks in computer vision, including the examples mentioned above, only require that the vertical (so-called *zenith*) VP and two or more horizontal VPs are detected. In that case, a lot of spurious VPs can be avoided by first detecting the zenith and the *horizon line* (HL), and then constraining the horizontal VPs on the HL. The zenith is generally easy to detect, as many lines converge towards that point in man-made environments. However, until recently, the HL was detected as an alignment of VPs, which led to a "chicken-and-egg" problem.

Last year, we showed that, assuming that the HL is inside the image boundaries, this line can usually be detected as an alignment of oriented line segments. This comes from the fact that any horizontal line segment at the height of the camera's optical center projects to the HL regardless of its 3-D direction. In practice, doors, windows, floor separation lines but also man-made objects such as cars, road signs, street furniture, and so on, are often placed at eye level, so that alignments of oriented line segments around the HLs are indeed observed in most images from urban or indoor scenes. This allowed us to propose a new method for VP detection, that was fast in execution and easy to implement. However, it was only middle rank in terms of accuracy. This year, we effectively put the HL detection into an *a contrario* framework. This transposal along with other improvements allows us to obtain top-ranked results in terms of both rapidity of computation and accuracy of the HL, along with more relevant VPs than with the previous top-ranked methods. This work has been submitted to CVPR 2018 (IEEE Conference on Computer Vision and Pattern Recognition).

Facade detection and localization

Planar building facades are semantically meaningful city-scale landmarks. Such landmarks are essential for localization and guidance tasks in GPS-denied areas which are common in urban environments. Detection of facades is also key in augmented reality systems that allow for the annotation of prominent features in the user's view. We proposed in [19] a novel object proposals method specific to building facades. We define new image cues that measure typical facade characteristics such as semantics, symmetry and repetitions. They are combined to generate a few facade candidates in urban environments fast. We show that our method outperforms state-of-the-art object proposals techniques for this task on the 1000 images of the Zurich Building Database. We demonstrated the interest of this procedure for augmented reality through facade recognition and camera pose initialization. In a very time-efficient pipeline we classify the candidates and match them to a facade references database using CNN-based descriptors. We proved that this approach is more robust to severe changes of viewpoint and occlusions than standard object recognition methods.

We are currently investigating ways to perform registration from this set of facade proposals. As point-based approaches may be inefficient to perform image/model matching due to changes in the illumination conditions, we propose to rely on semantic segmentation to improve the accuracy of this initial registration. Registration is here improved through an Expectation-Maximization framework. We especially introduce a Bayesian model that uses prior semantic segmentation as well as geometric structure of the facade reference modeled by L p Gaussian mixtures. This work has been submitted to CVPR'2018.

AR in industrial environments
As industrial environments are normally inundated with textureless objects and specular surfaces, it is difficult to capture enough features and build accurate 3D models for camera pose estimation using traditional 2D/3D matching-based approaches. Moreover, as people usually check industrial objects with free motions, recent CNN-based approaches could easily fail if the training data is not properly collected (e.g. does not cover enough views around the objects) and augmented (e.g. over-zoomed and over-augmented). For these challenges, we presented a novel protocol for six degrees of freedom (6-DOF) camera pose learning and estimation without any 3D reconstruction and matching processes. In particular, we proposed a visually controllable method to collect sufficient training images and their 6-DOF camera poses from different views and camera-object distances. Building upon this, we proposed a transfer learning scheme to train convolutional neural networks to detect objects and estimate the coarse camera pose from a single RGB image in an end-to-end manner. Experiments show that the trained convolutional network estimates each camera pose in about 5 ms and obtains approximately 13.3mm and 4.8 deg accuracy, which is compatible for training or maintenance tasks in industrial environments.

This work has been submitted to WACV 2018 (IEEE Winter Conf. on Applications of Computer Vision), and an extended version to TVCG (IEEE Transactions on Visualization and Computer Graphics).

7.2. Handling non-rigid deformations

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

Elastic multi-modal registration
Image-guided hepatic surgery is progressively becoming a standard for certain interventions. However, requirements on limited radiation dosis result in lower quality images, making it difficult to localize tumors and other structures of interest. Within J. Guevara's PhD thesis, we have proposed an automatic registration method exploiting the matching of the vascular trees, visible in both pre- and intra-operative images. The graphs are automatically matched using an algorithm combining Gaussian Process Regression and biomechanical model [20]. Indeed, Gaussian Process regression allows for a rigorous and fast error propagation but is extremely versatile. On the contrary, using biomecanical transformations is slower but provides physically correct hypotheses. Integrating the two approaches allows us to dramatically improve the quality of the matching for moderate or large organ deformations while reducing significantly the computational cost.

Individual-specific heart valve modeling
In this work, we focused on the segmentation of the valve cords. As dataset, we used 8 CT images of porcine hearts. Those data were acquired during various times with a microCT scan machine.

Within D. Panicheva's Master thesis, we worked on modeling the mitral valve chordae by applying a RANSAC-based method designed to extract cylinders with elliptical basis from a set of 3D contour points. To limit the search area, the results of segmentation obtained with classical methods for tubular structures extraction were used as initial assumptions of cords location.

The proposed method allows us to significantly improve cords segmentation results compared with classical methods, in particular, the section size and the endpoints of the cords are accurately defined which is important for future mechanical modeling of the mitral valve.

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

The motivation for the project is the serious medical condition, called ventilator induced diaphragmatic dysfunction (VIDD). During mechanical ventilation, air is pushed into the lungs resulting in a passive displacement of the diaphragm. This unnatural forcing results in loss of function in the muscle tissue. Our goal is to develop a simulator that allows for an in-silico exploration of the respiratory function with and without mechanical ventilation in combination with intervention measures that can reduce or prevent the risk for VIDD in the patients.

In the first year of the project, we worked on extracting a mesh from the segmented medical data that includes the boundary conditions. This work relies on analyzing the physiological constrains (rib motions, thoracic, abdominal and lung pressures) that can be measured.

We also worked on a method to solve differential equations on a complex geometric domain using the Radial Basis Function Partition of Unity Collocation Method (RBF-PUM). To use RBF-PUM for solving differential equations a covering of the domain has to be formed. The test implemented Poisson's diffusion equation on a domain defined by the diaphragm geometry. The diaphragm is not an easy case due to its thickness and shape.

This work is funded by the Swedish Research Council and realized within a collaboration with Uppsala University.

7.3. Interventional neuroradiology

Participants: Marie-Odile Berger, Charlotte Delmas, Erwan Kerrien, Raffaella Trivisonne.

Tools reconstruction for interventional neuro-radiology

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

In collaboration with GE Healthcare, our project aims at devising ways to reconstruct the micro-tools in 3D from fluoroscopy images. Charlotte Delmas has been working as a PhD CIFRE student on this subject since April 2013. She presented her research and results about fluoroscopic image segmentation, live stereo reconstruction of the guidewire, and fast 3D coil reconstruction, together with in-depth validation, in her PhD manuscript [11].

Image driven simulation

We consider image-driven simulation, applied to interventional neuroradiology as a coupled system of interactive computer-based simulation (interventional devices in blood vessels) and on-line medical image acquisitions (X-ray fluoroscopy). The main idea is to use the live X-ray images as references to continuously refine the parameters used to simulate the blood vessels and the interventional devices (micro-guide, micro-catheter, coil).

Raffaella Trivisonne started her PhD thesis in November 2015 (co-supervised by Stéphane Cotin, from MIMESIS team in Strasbourg) to address this research topic. We investigated various image and mechanical constraints, and proposed an efficient constrained shape from template approach where a set of radio-opaque markers on the catheter are tracked in the fluoroscopic images, and the surface of the vessel defines a set of unilateral constraints to prevent the catheter from crossing the vessel wall [22]. In particular, a constraint on the insertion point of the catheter at the groin was necessary to retrieve an accurate 3D shape of the micro-device.

7.4. Assessing metrological performance in experimental mechanics

Participant: Frédéric Sur.

In experimental mechanics, displacement and strain fields are estimated through the analysis of the deformation of patterns deposited on the surface of the tested specimen. Regular patterns such as grids are processed with spectral methods (the so-called "grid method"), and random speckle patterns are processed with digital image correlation (DIC). The scientific contribution obtained in 2017 concerns the comparison of the grid method and DIC. Since neither guidelines nor precise standard are available to perform a fair comparison between them, a methodology must first be defined. In [13], it is proposed to rely on three metrological parameters, namely the displacement resolution, the bias and the spatial resolution, which are not independent but linked.For the value of the bias fixed in the study of [13], the grid method features a better compromise than subset-based local DIC between displacement resolution and spatial resolution, in spite of its additional cost due to grid depositing. Work in progress concerns several aspects of DIC-based methods. In particular, we worked on synthetic speckle image rendering: ground truth databases are indeed crucial to assess the performance of the algorithms estimating displacement fields. It is, however, not straightforward to control any aspect of the rendering algorithm to ensure that performance assessment is not biased by the rendering algorithm. In addition, a popularization paper has been published in ERCIM News [15]. This work is part of a collaboration between Magrit project-team and Institut Pascal (Clermont-Ferrand).

MULTISPEECH Project-Team

7. New Results

7.1. Explicit Modeling of Speech Production and Perception

Participants: Anne Bonneau, Vincent Colotte, Yves Laprie, Slim Ouni, Agnès Piquard-Kipffer, Benjamin Elie, Theo Biasutto-Lervat, Sara Dahmani, Ioannis Douros, Valérian Girard, Yang Liu, Anastasiia Tsukanova.

7.1.1. Articulatory modeling

7.1.1.1. Articulatory models and synthesis

The geometry of the vocal tract is essential to guarantee the success of articulatory synthesis. This year we worked on the construction of an articulatory model of the epiglottis from MRI images and X-ray films. The new model takes into account the influences of the mandible, tongue and larynx via a multi-linear regression applied to the contours of the epiglottis [44]. Once these influences are removed from the contours, principle component analysis is applied to the control points of the B-spline representing the centerline of the epiglottis. The main advantage of using the centerline is to reduce the effect of delineation errors. Following the same idea, we also developed an articulatory model of the velum.

Geometry of the vocal tract is an input of articulatory synthesis and an algorithm for controlling the positions of speech articulators (jaw, tongue, lips, velum, larynx and epiglottis) is required to produce given speech sounds, syllables and phrases. This control has to take into account coarticulation and be flexible enough to be able to vary strategies for speech production [65]. The data for the algorithm are 97 static MRI images capturing the articulation of French vowels and blocked consonant-vowel syllables. The results of this synthesis are evaluated visually, acoustically and perceptually, and the problems encountered are broken down by their origin: the dataset, its modeling, the algorithm for managing the vocal tract shapes, their translation to the area functions, and the acoustic simulation.

7.1.1.2. Acoustic simulations

The acquisition of EPGG data (ElectroPhotoGlottoGraphy) data in collaboration with LPP in Paris has allowed the exploration of the production of voiced and unvoiced fricatives with realistic glottis opening profiles. These data show that the glottal opening is gradual and starts well before the fricative itself. Production of fricatives were studied by using acoustic simulations based on classic lumped circuit element methods to compute the propagation of the acoustic wave along the vocal tract. The glottis model incorporating a glottal chink developed last year is connected to the wave solver to simulate a partial abduction of the vocal folds during their self-oscillating cycles. Area functions of fricatives at the three places of articulation of French (palato-alveolar, alveolar, and labiodental) have been extracted from static MRI acquisitions. Simulations highlight the existence of three distinct regimes, named A, B, and C, depending on the degree of abduction of the glottis. They are characterized by the frication noise level: A exhibits a voiced signal with a low frication noise level, B is a mixed noise/voiced signal, and C contains only frication noise [33], [12].

Following the same approach of coupling articulatory data and acoustic simulation we investigated the acoustic simulation of alveolar trills, and the articulatory and phonatory configurations that are required to produce them. Using a realistic geometry of the vocal tract, derived from cineMRI data of a real speaker, the mechanical behavior of a lumped two-mass model of the tongue tip was studied [13]. The incomplete occlusion of the vocal tract during linguopalatal contacts was modeled by adding a lateral acoustic waveguide. Finally, the simulation framework is used to study the impact of a set of parameters on the characteristic features of the produced alveolar trills. It shows that the production of trills is favored when the tongue tip position is slightly away of the alveolar zone, and when the glottis is fully adducted.

7.1.1.3. Acquisition of articulatory data

The effort of acquiring new articulatory data was quite strong this year: **(i)** acquisition of MRI films (136 x 136 pixel images at a sampling rate of 55Hz) of continuous speech in Max Planck Institute Göttingen with Prof. Jens Frahm. We collected 2 hours of speech for 2 male speakers covering sentences and spontaneous speech. The sentences were designed so as to contain all the consonants and consonant clusters (excepted the very rare ones) in four vocalic contexts (the three cardinal vowels and /y/) and some intermediate vowels to check how they can be derived from those extreme vowels. The acoustic speech signal was recorded and denoised. Orthographic annotations of speech are available and the phonetic alignments were computed from the denoised speech signal. **(ii)** acquisition of EPGG (ElectroPhotoGlottoGraphy) data in LPP (Laboratoire de Phonologie et de Phonétique in Paris). The principle is to measure the flow of light (infrared light) which crosses the glottis. The emitting source is placed above the glottis and a light sensor below. The flow of light crossing the obstacle is roughly proportional to the surface of the glottis. Data acquired cover VCVs for fricatives and stops and some consonant clusters. These data were used to study the coordination between glottis opening and the realization of constrictions in the vocal tract. **(iii)** acquisition of fibroscopy data in HEGP (Georges Pompidou European Hospital). The principle is to introduce a smooth endoscope through the nostrils up to the top of the pharynx so as to image the glottis opening. This technique only allows a frequency close to 50 Hz which is not sufficient to observe the smooth glottis opening profiles accompanying the production of fricatives. Data have been collected for one female speaker and two male speakers.

7.1.2. Expressive acoustic and visual synthesis

We have improved our audiovisual acquisition techniques by acquiring a very advanced 8-camera motion capture system that allows capturing 3D data with higher temporal resolution and accuracy. We have acquired a small corpus for testing and evaluation purpose.

Within the framework expressive audiovisual speech synthesis, a perceptive case study on the quality of the expressiveness of a set of emotions acted by a semi-professional actor has been conducted. We have analyzed the production of this actor pronouncing a set of sentences with acted emotions, during a human emotion-recognition task. We have observed different modalities: audio, real video, 3D-extracted data, as unimodal presentations and bimodal presentations (with audio). The results of this study show the necessity of such perceptive evaluation prior to further exploitation of the data for the synthesis system. The comparison of the modalities shows clearly what the emotions are, that need to be improved during production and how audio and visual components have a strong mutual influence on emotional perception [57].

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

7.1.3.1. Categorization of sounds for native speech

Concerning the mother tongue, we conducted empirical research. We followed 170 young people, aged from 6 to 20 years old, with langage deficiencies - dyslexia and Specific Language Impairment (SLI) - including categorization of sounds. We examined the links between those difficulties and their schooling experience and observed how they constituted a point of major obstacle at the time of learning to read and to write, which the pupils do not overcome. All of them were in a handicap situation [18].

We conducted two descriptive studies which aims were to give an overview of educational systems for students with special educational needs, including pupils with learning and sound categorization disabilities (LD). Around the world, schooling is different from one country to another, according to the languages, even every country follows the international movement of school for all. For these students, the question of the best mode of inclusion remains topical [16]. In France, different types of scooling are observed. We focused our study on a particular system of teaching - a local unit for inclusive education - for children aged from 6 to 12 with specific language disorders - dyslexia and SLI - and learning disabilities, in a specialised school. We described a few exemples of pedagogical multimodal accomodations [15].

7.1.3.2. Digital books for language impaired children

In the framework of Handicom ADT project [7], we used one of the digital books prototypes set up with the use of a 3D avatar as narrator and multimodal speech, combining oral, written language and visual clues

(i.e. LPC, french cued speech), specially targeting children between 3 and 6. After the study conducted with digital album users, speech-therapists or re-eductors with hearing impaired impaired children, SLI and children with autism [81], we conducted another study, following children at school to investigate how technological innovations could help kindergarten children's (with and without langage difficulties) to improve their speech and language habilities.

7.1.3.3. Analysis of non-native pronunciations

Deviations in L2 intonation affect a number of prosodic characteristics including pitch range, declination line, or the rises of non-final intonation phrases, and might lead to misunderstandings or contribute to the perception of foreign-accent. This study investigates the characteristics of non-native speech at the boundary between prosodic constituents [67]. We analyzed a French declarative sentence, extracted from the IFCASL corpus (http://www.ifcasl.org), made up of four constituents and pronounced with a neutral intonation. Each constituent has three syllables and the sentence is realized typically by French speakers with four accentual –prosodic- groups, corresponding to the four constituents. Fourty German learners of French (beginners, and advanced speakers) and fifty four French speakers read the sentence once. We used the software ProsodyPro from Yi Xu for the prosodic analysis. We determined the presence of pauses and evaluated for each prosodic group: the (normalized) F0 maximum on the last syllable; the F0 excursion (max-min) of the final contour, and its maximum of velocity. In order to analyze the temporal course of F0 on the final contour, we also compared the values of the F0 excursion on the vowel and before it. On the basis of acoustic cues, non-native speakers, especially beginners, appear to realize more important prosodic boundaries (in particular higher F0 maxima, especially at the very end of the prosodic group, and more pauses) than French speakers, whereas native speakers appear to show more anticipation.

7.2. Statistical Modeling of Speech

Participants: Vincent Colotte, Dominique Fohr, Irène Illina, Denis Jouvet, Antoine Liutkus, Odile Mella, Romain Serizel, Emmanuel Vincent, Md Sahidullah, Guillaume Carbajal, Ken Deguernel, Mathieu Fontaine, Amal Houidhek, Aditya Nugraha, Laureline Perotin, Imran Sheikh, Sunit Sivasankaran, Ziteng Wang, Ismaël Bada.

7.2.1. Source separation

We wrote an extensive overview article about multichannel source separation and speech enhancement [14] and two book chapters about single-channel [72] and multichannel separation based on nonnegative matrix factorization [74].

7.2.1.1. Deep neural models for source separation and echo suppression

We pursued our research on the use of deep learning for multichannel source separation. In our previous work, which we summarized in a book chapter [73], we estimated the short-time spectra of the sound sources by a deep neural network and their spatial covariance matrices by a classical expectation-maximization (EM) algorithm and we derived the source signals by a multichannel Wiener filter. We also explored several variants of the multichannel Wiener filter, which turned out to result in better speech recognition performance on the CHiME-3 dataset [23]. We developed a new "end-to-end" approach which estimates both the short-time spectra and the spatial covariance matrices by a dedicated deep neural network architecture and which outperforms previously proposed approaches on CHiME-3. Arie Aditya Nugraha described the latter approach in his thesis, which he successfully defended. We started exploring the usage of deep neural networks for reducing the residual nonlinear echo after linear acoustic echo cancellation [80] and for separating multiple speakers from each other.

We also continued our work on music source separation, with the organization of the successful Signal Separation Evaluation Challenge (SiSEC 2016 [46]), as well as with national and international collaborations on this topic [34], [47], [58], [59], [60]. This research activity features several important research directions, described below.

7.2.1.2. Alpha-stable modeling of audio signals

Under the KAMoulox funding, we investigated the use of alpha-stable probabilistic models for source separation. As opposed to their more classical counterparts, these models feature very heavy tails, which allows to better account for the large dynamics found in audio signals. In close collaboration with national and international partners, we published several papers in international conferences on these topics. We demonstrated that alpha-stable processes allow to understand long-standing practices in speech enhancement [36]. More specifically, we showed that parameterized Wiener filters, dating back to the early 80s, can be understood as the optimal filtering strategy when sources are distributed with respect to alpha-stable distributions of different characteristic exponents. Interestingly, this gives a rationale for setting filtering parameters that were always manually tuned. Stable distributions also allow generalizing Wiener filtering for nonnegative sources [48], [49], and are interesting for robust multichannel separation [45], in the sense that they permit to compensate for model mismatch efficiently.

7.2.1.3. Scalable source localization

In the context of KAMoulox, we studied how probabilistic modeling of multichannel audio with alpha-stable distributions leads to models for microphone arrays that allow for scalable inference for the source positions [37], [38]. The core points of these methods are twofold. First, heaviness of the tails of alpha-stable distributions allows to efficiently model the marginal distribution of sources spectra. This is in sharp contrast with Gaussian distributions, that can only correctly represent audio signals adequately if each time-frequency point has its own distribution. On the contrary, while alpha-stable distributions give a high probability mass to small magnitudes, they also allow for the important deviations to be expected when the source is active. The advantage of such a model for marginal distributions over the whole time-frequency plane is to dramatically reduce the number of parameters and thus lead to much robust estimation methods. The second innovation brought in by the proposed localization method is to compute a summarized representation of the data, and to proceed to inference on this representation instead of using the -massive- original data.

7.2.1.4. Interference reduction

Under the DYCI2 schedule, we significantly extended our previous research on interference reduction for musical recordings. This task consists in reducing inter-microphone leakage in live recordings and has many applications in the audio engineering industry. This lead us to propose two important contributions on this respect. First, we amended previous methods to correctly exploit the proposed probabilistic model: previous research indeed featured some ad-hoc and suboptimal steps. This was corrected and the corresponding extension proved to behave much better [30]. Second, we investigated whether the proposed methods can be generalized to process full-length recordings. This is indeed an important and challenging question, because full-length multitrack recordings are extremely large and cannot reasonably be processed with current methods. This line of research lead us to propose inferring some parameters on compressed representations, which is promising ongoing research.

7.2.2. Acoustic modeling

7.2.2.1. Noise-robust acoustic modeling

In many real-world conditions, the target speech signal is reverberated and noisy. We conducted an extensive evaluation of several approaches for speech recognition in varied reverberation conditions, including both established and newly proposed approaches [21].

Speech enhancement and automatic speech recognition (ASR) are most often evaluated in matched (or multicondition) settings where the acoustic conditions of the training data match (or cover) those of the test data. We conducted a systematic assessment of the impact of acoustic mismatches (noise environment, microphone response, data simulation) between training and test data on the performance of recent DNN-based speech enhancement and ASR techniques [22]. The results show that multi-condition training outperforms matched training on average, but training on a subset of noise environments only is preferable in a few specific cases [25]. This raises the question: what are the optimal training conditions given the task to be solved, the deep neural network architecture, and the test conditions? We provided a preliminary answer to this question

by means of a discriminative importance weighting algorithm which aims to select the most useful training data in a rigorous optimization framework [64].

In order to motivate further work by the community, we created the series of CHiME Speech Separation and Recognition Challenges in 2011. Following the organization of the CHiME-3 Challenge in 2015, we edited a special issue [9] of *Computer Speech and Language*, which includes a detailed description of its outcomes [10]. We also published a book chapter that summarizes the outcomes of the whole series of challenges [70].

7.2.2.2. Environmental sounds

Following the recruitment of Romain Serizel in Fall 2016, our team has become more involved in the community on environmental sound recognition. In collaboration with Carnegie Mellon University (USA), we co-organized the first ever large-scale environmental sound recognition evaluation. This evaluation relied on the Audioset corpus released by Google and was part of the DCASE 2017 Challenge [24]. It focused on the problem of learning from weak labels for an application to smart cars.

We continued our work on acoustic scene classification. In particular, we focused on exploiting matrix factorization techniques for features learning. We extended previous work that used these learned features as an input to a linear classifier [11] to the deep learning framework [27], [28] and we proposed to jointly learn the deep-learning based classifier and the dictionary matrix [27]. A system based on this approach was submitted to DCASE challenge and was among the top 25% systems [28].

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. The first results, achieved in the MGB Challenge framework, show an improvement of the ASR word error rate compared to a Gaussian Mixture Model (GMM) based speech/non-speech segmenter.

7.2.2.4. Data selection

Training a speech recognition system needs audio data and their corresponding exact transcriptions. However, manual transcribing is expensive, labor intensive and error-prone. Some sources, such as TV broadcast, have subtitles. Subtitles are closed to the exact transcription, but not exactly the same. Some sentences might be paraphrased, deleted, changed in word order, etc. Building automatic speech recognition from inexact subtitles may result in a poor model and low performance system. Therefore, selecting data is crucial to obtain highly efficient models. We study data selection methods based on phone matched error rate and average word duration [26]

7.2.2.5. Transcription systems

We designed a new automatic transcription system based on deep learning with an acoustic modeling done by TDNN-HMM and a language model rescoring using RNN. In the framework of the AMIS project, we developped automatic systems for the transcription of TV shows in English, in French and in Arabic [52] [51].

7.2.2.6. Speaker identification

We proposed supervised feature learning approaches for speaker identification that rely on nonnegative matrix factorization [61]. The approach integrates a recent method that relies on group nonnegative matrix factorization into a task-driven supervised framework for speaker identification [11]. The goal is to capture both the speaker variability and the session variability while exploiting the discriminative learning aspect of the task-driven approach.

7.2.3. Language modeling

7.2.3.1. Out-of-vocabulary proper name retrieval

The diachronic nature of broadcast news causes frequent variations in the linguistic content and vocabulary, leading to the problem of Out-Of-Vocabulary (OOV) words in automatic speech recognition. Most of the OOV words are found to be proper names whereas proper names are important for automatic indexing of audio-video content as well as for obtaining reliable automatic transcriptions. New proper names missed by the speech

recognition system can be recovered by a dynamic vocabulary multi-pass recognition approach in which new proper names are added to the speech recognition vocabulary based on the context of the spoken content. We proposed a Neural Bag-of-Weighted Words (NBOW2) model which learns to assign higher weights to words that are important for retrieval of an OOV PN. [20]. We explored topic segmentation in ASR transcripts using bidirectional RNNs for change detection [62].

7.2.3.2. Adding words in a language model

We proposes new approaches to OOV proper noun probability estimation using Recurrent Neural Network Language Model (RNNLM). The proposed approaches are based on the notion of closest in-vocabulary words (list of brothers) to a given OOV proper noun. The probabilities of these words are used to estimate the probabilities of OOV proper nouns thanks to RNNLM [40].

7.2.3.3. Updating speech recognition vocabularies

In the framework of 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 [42]. Results shows that a significant reduction of the amount of out-of-vocabulary words is observed for the three languages, and that, for a given vocabulary size, the percentage of out-of-vocabulary words is higher for Arabic than for the other languages.

7.2.3.4. Segmentation and classification of opinions

Automatic opinion/sentiment analysis is essential for analysing large amounts of text as well as audio/video data communicated by users. This analysis provides highly valuable information to companies, government and other entities, who want to understand the likes, dislikes and feedback of the users and people in general. We proposed a recurrent neural network model with bi-directional LSTM-RNN, to perform joint segmentation and classification of opinions [63].

7.2.3.5. 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 [79]. In the context of ANR DYCI2, we described a general framework for automatic music improvisation that encompasses three existing paradigms [56] and that relies on our previous work about combining a multi-dimensional probabilistic model encoding the musical experience of the system and a factor oracle encoding the local context of the improvisation. Inspired in particular by the regularity of the temporal structure of popular music pieces [19], we proposed a new polyphonic music improvisation approach that takes the structure of the musical piece at multiple time scales into account [32].

7.2.4. Speech generation

Work on Arabic speech synthesis was carried out within a CMCU PHC project with ENIT (École Nationale d'Ingénieurs de Tunis, Tunisia, cf. 9.4.2.1), using HMM and NN based approaches applied to Modern Standard Arabic language.

HMM-based speech synthesis system relies 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, the vowel quantity and the consonant gemination. This year, we studied thoroughly the modeling of these phenomena. Results of objective and subjective evaluations showed that the results are similar between the different approaches that have been studied [39]. Other similar experiments are on-going using neural-network-based synthesis.

A particular weakness point of HMM-based synthesis quality may be due to the prediction of prosodic features which is based on a decision tree approach. Neural network are known for their ability to model complex relationships. This year, we studied the modeling of phoneme duration with NN approaches. Predicted phoneme durations will then be included in the Modern Standard Arabic synthesis system.

In parallel, the neural network based approach has also been tested on the French language.

7.3. Uncertainty Estimation and Exploitation in Speech Processing

Participants: Vincent Colotte, Dominique Fohr, Denis Jouvet, Yves Laprie, Odile Mella, Emmanuel Vincent, Yassine Boudi, Mathieu Hu, Karan Nathwani.

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. We conducted an extensive experimental investigation of existing uncertainty estimation and propagation techniques using deep neural network acoustic models on two different datasets (CHiME-2 and CHiME-3) [53]. We also proposed a deep neural network-based uncertainty estimator and a consistent way of accounting for uncertainty in both the training and decoding stage [54]. Overall, we were the first to report a significant improvement using uncertainty estimation and propagation compared to a competitive deep neural network acoustic modeling baseline based on feature-domain maximum likelihood linear regression (fMLLR) features.

7.3.1.2. Uncertainty in other applications

Besides the above applications, we pursued our exploration of uncertainty modeling for robot audition and wind turbine control. In the first context, uncertainty arises about the location of acoustic sources and the robot is controlled to locate the sources as quickly as possible [55]. In his successfully defended thesis, Quan Van Nguyen also described a way of locating multiple sources. In the second context, uncertainty arises about the noise intensity of each wind turbine and the turbines are controlled to maximize electrical production under a maximum noise threshold [31].

7.3.2. Uncertainty and phonetic segmentation

In the framework of the LCHN CPER project (cf. 9.1.1), for studying prosodic correlates of discourse particles in French, phonetic boundaries of discourse particles and adjacent words have been checked and manually corrected; this shows that there is still a need for performance improvement of the automatic speech-text alignment process.

We also worked on speech-to-speech alignement, with the goal of obtaining a precise alignement between two speakers pronouncing the same sentence. This task is difficult due to the fact that the speakers may pronounce certain sounds in a different way, or they may insert or remove silences between words. We introduced explicit phoneme duration and insertion/deletion models for alignment and evaluated them on real data.

7.3.3. Uncertainty and prosody

The fundamental frequency is one of the prosodic features. Numerous approaches exist for the computation of F0. Most of them lead to good performance on good quality speech. The performance degradation with respect to noise level has been studied on reference databases, for several (about ten) F0 detection approaches. It was observed that for each algorithm, a large part of the errors are due to incorrect voiced/unvoiced decision [43]. A first set of experiments have been conducted for computing a confidence measure on the estimated F0 values through the use of neural network approaches [29].

Study of discourse particles in French has continued thanks to the support of the CPER LCHN project. So far a few French words frequently used as discourse particles have been studied. Several thousands occurrences have been extracted from the ESTER and the ORFEO speech corpora, and annotated as discourse particle or not. The pragmatic function of the discourse particles has also been annotated. Prosodic correlates of these words have been analyzed with respect to their function (discourse particle or not, as well as pragmatic function) [66], and some automatic classification processes have been investigated [41].

ORPAILLEUR Project-Team

7. New Results

7.1. Mining of Complex Data

Participants: Quentin Brabant, Miguel Couceiro, Adrien Coulet, Esther Catherine Galbrun, Nyoman Juniarta, Florence Le Ber, Joël Legrand, Pierre Monnin, Tatiana Makhalova, Amedeo Napoli, Justine Reynaud, Chedy Raïssi, Mohsen Sayed, Yannick Toussaint.

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

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 [77]) 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 [90] 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 also worked on a generic framework based on FCA in which we can define the pattern mining process at a formal level [3]. We consider several types of patterns and we are making precise the mining of complex patterns represented as sequences, trees and graphs.

We also worked on a significant extension of previous work on the discovery of skyline patterns (or "skypatterns") based on the theoretical relationships with condensed representations of patterns. We have shown how these relationships facilitate the computation of skypatterns. Thus we proposed a flexible and efficient approach to mine skypatterns using a dynamic constraint satisfaction problems (CSP) framework [30].

7.1.2. Text Mining

In the context of the PractikPharma ANR Project, we study cross-corpus training with Tree-LSTM for the extraction of biomedical relationships from texts, especially, how large annotated corpora developed for alternative tasks may improve the performance on biomedicine related tasks, for which few annotated resources are available [55]. We experiment two deep learning-based models to extract relationships from biomedical texts with high performance. The first one combines locally extracted features using a Convolutional Neural Network (CNN) model, while the second exploits the syntactic structure of sentences using a Recursive Neural Network (RNN) architecture. Our experiments show that the latter benefits from a cross-corpus learning strategy to improve the performance of relationship extraction tasks. Indeed our approach leads to state-of-the-art performances for four biomedical tasks for which few annotated resources are available (less than 400 manually annotated sentences). This may have a particular impact in specialized domains in which training resources are scarce, because they would benefit from the training data of other domains for which large annotated corpora do exist.

In the framework of the Hybride ANR project (terminated at the end of 2016), Mohsen Sayed Hassan proposed an original machine learning approach for identifying in texts about diseases phenotypes that are not yet represented within existing ontologies [9]. The result of the extraction is used to enrich existing ontologies of the considered domain. We studied three research directions: (1) extracting relationships from texts, i.e., extracting Disease-Phenotype (D-P) relationships, (2) identifying new complex entities standing as phenotypes of a rare disease, and (3) enriching an existing rare disease ontology on the basis of the relationships previously extracted.

A collection of abstracts of scientific articles is represented as a collection of dependency graphs used for discovering relevant pieces of biomedical knowledge. We focused on the completion of rare disease descriptions, by extracting Disease-Phenotype relationships. We developed an automatic approach named SPARE⋆, for extracting Disease-Phenotype relationships from PubMed abstracts, where phenotypes and rare diseases are previously annotated by a Named Entity Recognizer. SPARE⋆ is the resulting hybrid approach that combines a graph-pattern based method, called SPARE, and a machine learning method (SVM). It benefits both from the good precision of SPARE and from the good recall of SVM. Finally, we applied pattern structures for classifying rare diseases and enriching an existing ontology about such diseases.

7.1.3. Mining Sequences and Trajectories

Nowadays datasets are available in very complex and heterogeneous ways. Mining of such data collections is essential to support many real-world applications ranging from healthcare to marketing. This year we finished a work on the analysis of "complex" sequential data and its usage in video games for the analysis of strategy "balance" in those games [14].

7.1.4. 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 [89]. 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.

In a preceding work [83], we focused on the problem of pattern selection, developing a method for filtering a set of redescription to identify a non-redundant, interesting subset to present to the analyst. Also, we showcased the usability of redescription mining on an application in the political domain [76]. More specifically, we applied redescription mining to the exploratory analysis of the profiles and opinions of candidates to the parliamentary elections in Finland in 2011 and 2015.

We presented an introductory tutorial on redescription mining at SDM in April 2017 to help foster the research on these techniques and widen their use (http://siren.mpi-inf.mpg.de/tutorial_sdm2017/main/).

7.1.5. 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.

In [1], we argue that the greedy nature of most of these approaches produce pattern sets that lack of diversity. Consequently, 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.

7.1.6. Data Privacy: Online link disclosure strategies for social networks

Online social networks are transforming our culture and world. While online social networks have become an important channel for social interactions, they also raise ethical and privacy issues. A well known fact is that social networks leak information, that may be sensitive, about users. However, performing accurate real world online privacy attacks in a reasonable time frame remains a challenging task. We continued our work on this aspect and we address the problem of rapidly disclosing many friendship links using only legitimate queries (i.e., queries and tools provided by the targeted social network). The results of this joint work with the Pesto Inria Team are published in [31].

7.1.7. Aggregation

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 [79], [82]. 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 our recent work, 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 [38], [59]. Problems related to feature selection and model elicitation where tackled in [15].

7.2. Knowledge Discovery in Healthcare and Life Sciences

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

7.2.1. Ontology-based Clustering of Biological Linked Open Data

Increasing amounts of biomedical data provided as Linked Open Data (LOD) offer novel opportunities for knowledge discovery in bio-medicine. We proposed an approach for selecting, integrating, and mining LOD with the goal of discovering genes responsible for a disease [87]. We are currently working on the integration of LOD about known phenotypes and genes responsible for diseases along with relevant bio-ontologies. We are also defining a corpus-based semantic distance. One possible application of this work is to build and compare possible "diseaseomes", i.e. global graphs representing all diseases connected according to their pairwise similarity values.

7.2.2. Biological Data Aggregation for Knowledge Discovery

This specific research takes place within two multi-disciplinary projects initiated in 2016, in collaboration with the Capsid Team, with a group of clinicians from the Regional University Hospital (CHU Nancy) and bio-statisticians from the Maths Lab (IECL). The first project is entitled ITM2P [0] and depends on the so-called CPER 2015–2020 framework. We are involved in the design of the SMEC platform as a support for "Simulation, Modeling and Knowledge Extraction from Bio-Medical Data".

The second project is a RHU [0] project entitled *Fight Heart Failure* (FHF), where we are in charge of a workpackage about entitled "Network-based analysis and integration". Accordingly, we are working on the definition of multidimensional similarity measure for comparing and clustering sets of patients. Each cluster should correspond to a bioprofile, i.e. a subgroup of patients sharing the same form of the disease and thus the same diagnosis and care strategy. The first results were presented at the "International Symposium on Aggregation and Structures (ISAS 2016)" [74] where we proposed an approach for complex graph aggregation resulting in a similarity graph between a subset of nodes. In a recent work we explored an alternative to define and efficiently compute pairwise patient similarity thanks to "Unsupervised Extremely Randomized Trees" [62].

The next challenge is to build a prediction model for each bioprofile/subgroup, once validated by clinicians, to be integrated in a decision support system. Currently, we are investigating "Statistical Relational Learning" and analogy-based methods for achieving this goal.

7.2.3. Validation of Pharmacogenomics Knowledge

A standard task in pharmacogenomics research is identifying genes that may be involved in drug response variability. Those genes are called "pharmacogenes". As genomic experiments in this domain tend to generate many false positives, computational approaches based on background knowledge may generate more valuable results. Until now, the latter have only used molecular network databases or biomedical literature. We developed a new method that takes advantage of various linked data sources to evaluate the validity of uncertain drug-gene relationships, i.e. pharmacogenes [5]. One advantage relies on the standard implementation of linked data that facilitates the joint use of various sources and makes easier to consider features of various origins. The second advantage is related to graph mining approaches that we are using, which consider linked data in their original form, i.e. as graphs. We selected, formatted, interconnected and published an initial set of linked data sources relevant to pharmacogenomics, named PGxLOD (for "PharmacoGenomic Linked Open Data"). We applied and compared distinct numerical classification methods on these data and identified candidate pharmacogenes.

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

This work is a first attempt for validating state-of-the-art knowledge in pharmacogenomics, which is one objective of the ANR project "PractiKPharma" initiated in 2016 (http://practikpharma.loria.fr/). This year, we improved and enriched PGxLOD in various ways. Firstly, we wanted PGxLOD to be able to encompass pharmacogenomic knowledge of various origin, such as scientific literature, specialized databases, or Electronic Health Records (EHRs). To represent the fact that a given knowledge unit may have distinct provenances, we developed a simple ontology named PGxO ("Pharmacogenomic Ontology") which relies on the Standard Ontology PROV-O to represent provenance. This makes possible to compare similar knowledge units that may have distinct origins [45].

7.2.4. Analysis of biomedical data annotated with ontologies

In the context of the Snowflake Inria Associate Team (at present Snowball), 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. In this case, 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 the first research work 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 and analyzed and may be used as a basis of a recommendation system [29].

7.3. Knowledge Engineering and Web of Data

Participants: Emmanuelle Gaillard, 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

The Taaable project was originally created as a challenger of the Computer Cooking Contest (ICCBR Conference) [72]. Beyond its participation to the CCC challenges, the Taaable project aims at federating various research themes including case-based reasoning (CBR), knowledge discovery, knowledge engineering and belief change theory [6]. CBR performs adaptation of recipes w.r.t. user constraints. The reasoning process is based on a cooking domain ontology (especially hierarchies of classes) and adaptation rules. The knowledge base is encoded within a semantic wiki containing the recipes, the domain ontology and adaptation rules.

Adaptation rules have been used to manage ingredient adaptation with a restrictive set of available ingredients [43]. Three types of rule have been identified. The first type is about the substitution of ingredients belonging to a same category (e.g. dairy) by the sole available ingredient of this category (e.g. yogurt). The second type of rule is in concern with substitution, according to the role the ingredients play in the recipe, e.g. egg can be replaced by salmon in salad recipes because they are both playing the role of a protein. The last type of rules consists in removing ingredients of original recipes when they are not concerned by a rule of the first nor second type.

FCA allows the classification of objects according to the properties they share into a concept lattice. A lattice has been built from a large set a cocktail recipes according to the ingredients they use, producing a hierarchy of ingredient combinations. For example, when a cocktail recipe R has to be adapted, this lattice can be used to search the best ingredient combinations in the concepts that are the closest to the concept representing R [43].

Two main research works were carried out about the application of CBR in medicine. Imaging, in particular in nuclear medicine, is getting more and more complex over the years. Each year, new radiotracers and machines are developed and tested. Despite this rapid evolution, few studies address the issue of image interpretation and imaging report. In [35], we show how nuclear image interpretation is improved by Tetra, a new case-based decision support system.

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 [48], a system is presented to assist operators in the interpretation of best medical coding practices.

Finally, an approach to adaptation based on the principles of analogical transfer applied to the formalism RDFS has been developed. It is based on the problem-solution dependency represented as an RDFS graph: this dependency within the source case is modified so that it fits the context of the target problem [2]. This is implemented within the so-called SQTRL system (for "SPARQL Query Transformation Rule Language" http://tuuurbine.loria.fr/sqtrl/) [2]. The development of SQTRL is based on a collaboration between Orpailleur team and the *Archives Henri Poincaré* (http://poincare.univ-lorraine.fr/).

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 [47]. 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.

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 [46]. The result of this comparison shows which "new axioms" can be proposed to ontology developers for guiding their design work.

7.4. Advances in Graph Theory, Clone Theory and Multiple-Valued Logic

Participants: Quentin Brabant, Miguel Couceiro, Amedeo Napoli, François Pirot, Chedy Raïssi, Jean-Sébastien Sereni.

Keywords: graph theory, graph colouring, extremal graph theory, chromatic number, multiple-valued logic, clone theory

7.4.1. Graph Theory

Proper colouring of triangle-free planar graphs is an active research topic with interesting algorithmic ramifications. It has been known for more than fifty years that such graphs can be properly 3-coloured, and Thomassen conjectured in 2007 that they actually admit an exponential number of such colourings. This statement is still wide open, and to bring forward further insight we established [75] it to be equivalent to the following:
there exists a positive real α such that whenever G is a planar graph and A is a subset of its edges whose deletion makes G triangle-free, there exists a subset A' of A of size at least $\alpha|A|$ such that $G-(A \setminus A')$ is 3-colourable. This equivalence allows us to study restricted situations, where we can prove the statement to be true.

Still on graph colouring, we demonstrated [93] a conjecture by Zhang and Whu made in 2011, that for every positive integer Δ, every K_4-minor-free graph with maximum degree Δ admits an equitable colouring with k colours whenever $k \geq \frac{\Delta+3}{2}$. A key ingredient was to *not* use the discharging method and rather exploit decomposition trees of K_4-minor-free graphs.

We also considered [88] distance colouring in graphs of maximum degree at most d and how excluding one fixed cycle of length ℓ affects the number of colours required as $d \to \infty$. For vertex-colouring and $t \geq 1$, if any two distinct vertices connected by a path of at most t edges are required to be coloured differently, then a reduction by a logarithmic (in d) factor against the trivial bound $O(d^t)$ can be obtained by excluding an odd cycle length $\ell \geq +3t$ if t is odd or by excluding an even cycle length $\ell \geq 2t + 2$. For edge-colouring and $t \geq 2$, if any two distinct edges connected by a path of fewer than t edges are required to be coloured differently, then excluding an even cycle length $\ell \geq 2t$ is sufficient for a logarithmic factor reduction. For $t \geq 2$, neither of the above statements are possible for other parity combinations of ℓ and t. These results can be considered extensions of results due to Johansson (1996) and Mahdian (2000), and are related to open problems of Alon and Mohar (2002) and Kaiser and Kang (2014).

7.4.2. Multiple-Valued Logic and (Partial) Clone Theory

Clone theory was primarily motivated by the study of Boolean logic, and it currently constitutes a major subject in universal algebra, multiple-valued logic, and theoretical computer science. A clone on a set A is a class of functions $f : A^n \to A$, $n \geq 1$, that contains all projections and that is closed under compositions. Clones on a set A constitute a closure system, in fact, an algebraic lattice where meet is given by set-intersection. Clones on a 2-element set were completely classified by Emil Post. Since Post's classification several studies on clone theory have appeared and many variants and generalizations have been proposed.

As a closure system, clones can be specified within a Galois framework, namely, through the well known Pol-Inv Galois connection by the polarity between functions and the relations they preserve. This Galois connection became the main tool in several studies, in particular, in the classification of the complexity classes of CSPs ("Constraint Satisfaction Problems") [92]. Another, rather surprisingly, application of this Galois framework led to the description of of analogy-preserving Boolean classifiers [4].

Similarly, clones of partial functions (i.e., functions $f : D \to A$ for $D \subseteq A^n$) can be described by the relations its members preserve. Unlike the lattice of clones, the lattice of partial clones is of continuum cardinality even in the case of 2-element underlying sets. This shows that a complete description of this lattice is hard to attain. However, many efforts have been made towards local descriptions of this lattice, for instance, concerning the classification of its intervals that has entailed a long lasting open problem. This was settled [20] in the form of a dichotomy theorem showing that such intervals are either finite or of continuum cardinality, and we presented precise descriptions of the structure some challenging intervals in [21]. Further developments and related problems were also tackled in [11], [24], [40], [39].

SEMAGRAMME Project-Team

6. New Results

6.1. Syntax-Semantics Interface

Participants: Philippe de Groote, Sylvain Pogodalla.

6.1.1. Lexical Semantics

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

The modeling of these concepts, and how they relate to each other, gave rise to Frames Semantics as a representation format of conceptual and lexical knowledge [40], [34], [28], [52]. Frames consist of directed graphs where nodes correspond to entities (individuals, events, ...) and edges correspond to (functional or non-functional) relations between these entities. Providing a fine-grained representation of the internal concept structure allows both for a *decomposition* of the lexical meaning and for a precise description of the sub-structural interactions in the semantic composition process [51].

Following up on our previous work [46], [47] based on Hybrid Logic (HL) [30], [27] on linking Frames and truth-logical semantics, we used the flexibility of the approach to model semantic coercion as induced by verbs such as *read* that can syntactically have an entity as argument (*John began a book*) while it semantically relates to an event (e.g., *reading*, *writing*, etc.) [7].

6.2. Discourse Dynamics

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

6.2.1. Dynamic Generalized Quantifiers

We have started a classification of the (French) determiners according to the dynamic properties of the generalized quantifiers they denote [12], [17].

Following Groenendijk and Stokhof [43], we say that a generalized quantifier is *internally dynamic* in case the dynamic binders occurring in its restriction have the capacity of binding material that occurs in their scopes. We also say that a generalized quantifier is *externally dynamic* in case the dynamic binders occurring in both its arguments have the capacity of binding material that occur in the continuation of the discourse. In addition to these notions of internal and external dynamicity, we consider a third notion that we call *intrinsic dynamicity*. We say that a generalized quantifier is intrinsically dynamic in case it introduces new referent markers and makes them available to the continuation of the discourse.

Using these three notions, we have defined three classes of dynamic generalized quantifiers, which fairly correspond to the notions of specific (e.g., *the*, *this*, *his*), general (e.g., *a*, *some*, *another*), and quantificational determiners (e.g., *every*, *no*). We then have shown how the dynamic generalized quantifiers belonging to these three classes may be formalized using the continuation-based approach introduced in [5].

6.2.2. Dialogue Modeling

Studying dialogical interactions is a major subject in natural language processing, since dialogues represent the basis of human communication. Addressing this problem requires relating approaches from fields such as semantics, pragmatics, and, more generally, logic, and cognition. We have presented a compositional dynamic model of questions and answers mechanisms in a dialogical setting. We address dialogical and lexical issues starting from the formal definitions of frame semantics given in [7]. We achieve compositionality and dynamicity in our model by constructing it on top of concepts inherited from Type Theoretical Dynamic Logic [5]. We introduce control in the common (accessible to all participants of a dialogue) context of a conversation by formulating the concept of dialogical context and elaborating corresponding storage operations. We apply our model to real non-controlled examples of dialogical interactions provided by the Schizophrenia and Language, Analysis and Modeling corpus [29]. The linguistic analysis of dialogues between patients with schizophrenia and psychologists has revealed specific language-driven manifestations of cognitive dysfunction. This approach to dialogue modeling in a dynamic framework allowed us to develop tools to handle specifics of dialogical interactions on top of already existing methods for general discourse.

6.2.3. Discourse Structure

A text as a whole must exhibit some coherence that makes it more than just a bag of sentences. This coherence hinges on the discourse relations (DRs). The latter express the articulations between the different pieces of information of the text. There is still debate about the number and the nature of these DRs. Yet, typical DRs include `Contrast`, `Consequence`, or `Explanation`. Using a discourse connective (*because*, *instead*, *although*) is usually the most direct and reliable way to express a DR. These lexical items have specific syntactic, semantic, and pragmatic properties. In particular, one can often observe a mismatch between the arguments of a DR and the (syntactic) ones of the connective lexicalizing it. It happens in configurations in which the argument of the DR does not directly correspond to syntactic argument of the discourse marker. In (1), for instance, the second argument of the `Explanation` relation is not the whole conditional, its antecedent, nor its consequent. But it is the possibility of the conditional, paraphrasable by *she might miss her train*. The discourse argument is here presupposed by the conditional (i.e., the syntactic argument).

1. Mary is worried because if there is too much delay, she will miss her train.
2. John did not come to the party although Mary said he was already back in town.

Another common case occurs when an attitude verb (*think*, *believe*) or a verb of report (*say*, *tell*) is used evidentially as in (2). In such cases, the contrast expressed by the writer holds between *John did not come to the party* and *he was already back in town*. The main function of the evidential (*Mary said ...*) is to introduce the argument of a DR without being itself part of the discourse structure.

Whereas DRs have two arguments, some discourse markers, such as adverbial connectives (*so*, *otherwise*), have only one syntactic argument. It then seems natural to use an anaphoric mechanisms to describe how the other argument of the DR they lexicalise is determined from the context. We extended this idea to all connectives and showed how this view can explain most usual cases of mismatch. Additionally, considering that discourse arguments are implicit semantic objects akin to the events introduced in the Davidsonian theory, it is possible to implement this proposal in Type Theoretic Dynamic Logic, without the need of a syntactic parse above the sentence level, and in a strictly compositional way, using continuations.

6.3. Common Basic Resources

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

6.3.1. Crowdsourcing Complex Language Resources

Using a Wikipedia corpus, we showed that participants in a game with a purpose can produce quality dependency syntax annotations [44]. In [15], we have been considering a more complex corpus of scientific language. We ran an experiment aiming at evaluating the production of the participants of the game, and compared it to a gold corpus, annotated and adjudicated by experts of the domain.

We also ran two surveys on ZombiLingo's players, in order to better understand who they are and what their motivations in playing the game are, and improve the participation in the game [14].

6.3.2. Universal Dependencies

We participated to the development of new versions of the French part of the Universal Dependencies project (UD, http://universaldependencies.org/). Version 2.0 [58] was released in March 2017. In this version, a new French corpus *UD_French-Sequoia* was added. We built this corpus with an automatic conversion (using the Grew software) from the data built in the Sequoia project.

Version 2.1 [24] was released in November 2017. The conversion process, using Grew, was applied to the FrenchTreebank corpus, and led to a new corpus in Universal Dependencies: *UD_French-FTB*. In version 2.1, we worked on the harmonization of the subset of French treebanks. The Grew software was used to explore, to check consistency, and to systematically correct the data.

The "enhanced dependencies" sketched in the UD 2.0 guidelines is a promising attempt in the direction of deep syntax, an abstraction of the surface syntax towards semantics. In [13] (collaboration with Marie Candito and Djamé Seddah), we proposed to go further and enrich the enhanced dependency scheme along two axes: extending the cases of recovered arguments of non-finite verbs, and neutralizing syntactic alternations. Doing so leads to both richer and more uniform structures, while remaining at the syntactic level, and thus rather neutral with respect to the type of semantic representation that can be further obtained. We implemented this proposal in two UD treebanks of French, using deterministic graph rewriting rules. Evaluation on a 200-sentence gold standard showed that deep syntactic graphs can be obtained from surface syntax annotations with a high accuracy. Among all semantic arguments of verbs in the gold standard, 13.91% are impacted by syntactic alternation normalization, and 18.93% are additional edges corresponding to deep syntactic relations.

In [16], we present a reflection on the annotation of written French corpora in syntax and semantics. This reflection is the result of work carried out on the SEQUOIA and the UD-FRENCH corpora.

6.3.3. FR-Fracas

There are two major levels of processing that are significant in the use of a computational semantic frameworks: 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 [37].

There currently exists a multilingual version of the resource for Farsi, German, Greek, and Mandarin. We started the translation into French. 10% of the resource has been translated so far as a testbed, in order to setup guidelines for the translations. We plan to complete the translation following these guidelines and use it as an experimental tool.