RESEARCH CENTER
Nancy - Grand Est

FIELD

Activity Report 2019

Section New Results

Edition: 2020-03-21

ALICE Team

7. New Results

7.1. Curved slicing

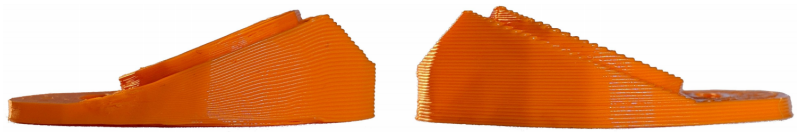


Figure 3. Sides views of curved slicing print (left) and adaptive slicing (right). Curved slicing eliminates all staircasing while closely following the input.

When printing 3D objects with Fused Filament Fabrication technology, the plastic is deposited by following a 2D path for producing the first layer. Each following layer is printed with the same method on the top of the previous layers. For technical reasons, it is convenient to use horizontal layers with constant height, but this generates aliasing errors that are especially visible (Figure 3 , right) when the object's surface is close to horizontal. The objective of this project is to reduce these artefacts by printing curved layers (Figure 3 , left). Printing curved layers is a challenging task because all technical aspects of printing have to be adapted to the curved case. The key idea of our approach is to (virtually) deform the object in such a way that the surface that is close to horizontal becomes exactly horizontal, then define all the printing instructions (tool path, slicing, pressure, etc.) in this deformed space with standard algorithms. The final printing instructions are obtained by coming back to the original space. In collaboration with MFX team, we have worked on the problem of finding the deformation by a global optimization method that tries to make horizontal large portions of the object's surface under constraints of layer thickness, tools collisions, object self-intersections, etc. The results were published at SIGGRAPH this year [7].

7.2. Coarse polycube meshes

This work is done as part of an informal (soon to be formalized) collaboration between our team and CEA. Many simulation codes require block-structured meshes. This requires decomposing the geometric domain into a set of hexahedral blocks, each one being discretized by a regular grid. Our approach to generate such structures is to generate global parameterizations. Those methods give promising results in many cases, but still face many robustness issues. To tackle those issues, we are currently working on a subset of those methods, called Polycube deformation. The idea is to deform our original domain Ω to align its boundary with a regular grid. We start by determining a set of constraints on the boundary of Ω. We then compute a map M that deforms the interior according to those constraints into a polycube. The inverse deformation M^{-1} applied to the polycube produces a structured hexmesh of the domain Ω, refer to Figure 4 . While relying on valid boundary constraints, this method is more robust than global parameterizations methods and gives good results on many models. We focus on obtaining coarse block structures, a very challenging problem with many robustness issues. Now we are able to generate as-coarse-as-possible hexahedral meshes (Figure 4 , right). We are preparing a publication of these results.

7.3. Roof fitting

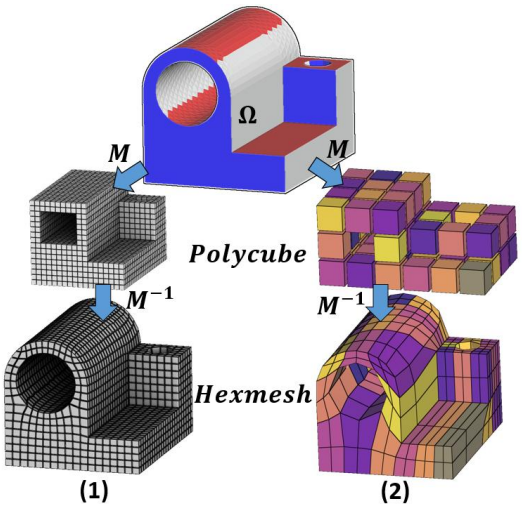


*Figure 4. The state of the art allows us to create fine polycube meshes (**left**), whereas we are trying to create meshes as coarse as possible (**right**).*

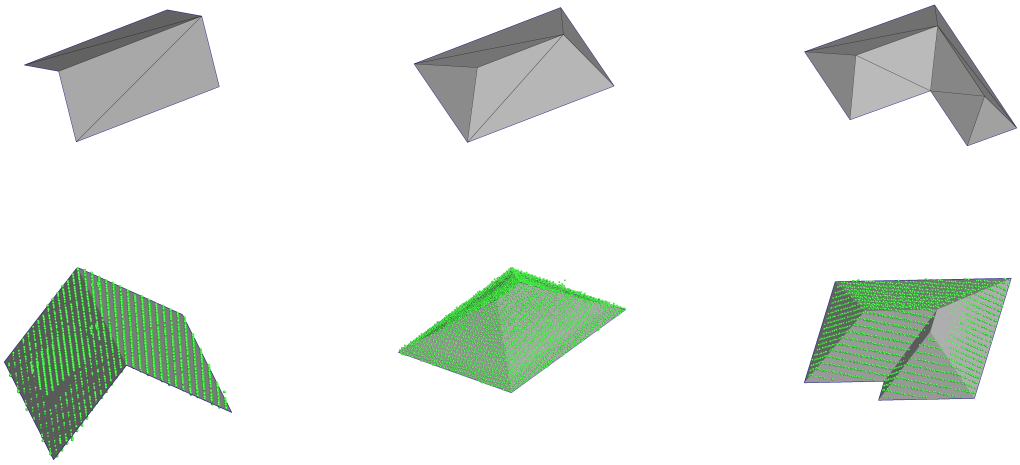


Figure 5. Top row: examples of different roof patterns. Bottom row: fitting of the patterns on LIDAR scans.

This work is done as part of an informal (soon to be formalized) collaboration between our team and RhinoTerrain. We have roof models in the form of surface meshes (Figure 5). Our data are LIDAR point clouds. Based on this data and a roof model chosen by the user, we seek to optimize the position of the model so that it "best" matches the data. This optimization must comply with two constraints:

- It is important to ignore possible outliers in the point cloud, such as parts that do not belong to the roof (trees, electrical wires, *etc.*) or should not be taken into account by the model (chimney, skylight, parabolic antenna, *etc.*);
- The roof geometry is subject to certain constraints, such as the planarity of certain rectangular faces or the alignment of certain axes.

This work is an extension of the VSDM algorithm (*Voronoï Squared Distance Minimization*) developed by the team [31]. The idea is to optimize a well-chosen energy function, the overall minimum of which corresponds to the desired position for the mesh size. The preliminary results are very promising, and we are preparing a publication.

BIGS Project-Team

7. New Results

7.1. Stochastic modelling

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

In the framework of the esca-illness of vines, we developed different spatial models and spatio-temporal models for different purposes: (1) study the distribution and the dynamics of esca vines in order to tackle the aggregation and the potential spread of the illness (2) propose a spatio-temporal model in order to capture the dynamics of cases and measure the effects of environmental covariates. For purpose (2), we developed an autologistic model (centered in a new way), have proposed estimators of the parameters and showed their properties and proposed a way to choose between several neighborhood models. These results were published in Spatial Statistics [4].

In a collaboration with physicists from Nancy CHRU, we have worked about the interest to use the whole distribution of telomeres lengths until the mean that is usually used to characterise ageing of a cell. We have shown that the shape of the distribution can be seen as a individuals's signature. It is the object of the paper published in Scientific Reports [8].

After preliminary suggestions on the building of models for low-grade gliomas [3], we focused our attention on the diffuse character of such tumors. We characterized the infiltrating phenotype (infiltration rate, direction of infiltration, evolution of morphology over time) as a new variable to consider in a context of multifactorial modelling (submitted article). A monocentric retrospective study has been conducted on the local database, estimating survival paramaters and comparing the effects of treatments (writing article). A brain cartography obtained by sensorial simulations during awake surgery with the aid of clustering analysis has been published in "Brain - A Journal of Neurology" [6].

7.2. Optimal Control of Markov Processes

Participants: B. Scherrer

Finite-horizon lookahead policies are abundantly used in Reinforcement Learning and demonstrate impressive empirical success. Usually, the lookahead policies are implemented with specific planning methods such as Monte Carlo Tree Search (e.g. in AlphaZero). Referring to the planning problem as tree search, a reasonable practice in these implementations is to back up the value only at the leaves while the information obtained at the root is not leveraged other than for updating the policy. Here, we question the potency of this approach. Namely, the latter procedure is non-contractive in general, and its convergence is not guaranteed. Our proposed enhancement, in [9], published in AAAI'2019, is straightforward and simple: use the return from the optimal tree path to back up the values at the descendants of the root. This leads to a γ^h-contracting procedure, where γ is the discount factor and h is the tree depth. To establish our results, we first introduce a notion called *multiple-step greedy consistency*. We then provide convergence rates for two algorithmic instantiations of the above enhancement in the presence of noise injected to both the tree search stage and value estimation stage.

Value iteration is a method to generate optimal control inputs for generic nonlinear systems and cost functions. Its implementation typically leads to approximation errors, which may have a major impact on the closed-loop system performance. We talk in this case of approximate value iteration (AVI). In [24], published in CDC'2019, we investigate the stability of systems for which the inputs are obtained by AVI. We consider deterministic discrete-time nonlinear plants and a class of general, possibly discounted, costs. We model the closed-loop system as a family of systems parameterized by tunable parameters, which are used for the approximation of the value function at different iterations, the discount factor and the iteration step at which we stop running the algorithm. It is shown, under natural stabilizability and detectability properties as well as mild conditions on the approximation errors, that the family of closed-loop systems exhibit local practical

stability properties. The analysis is based on the construction of a Lyapunov function given by the sum of the approximate value function and the Lyapunov-like function that characterizes the detectability of the system. By strengthening our conditions, asymptotic and exponential stability properties are guaranteed.

Many recent successful (deep) reinforcement learning algorithms make use of regularization, generally based on entropy or Kullback-Leibler divergence. In [10], published in ICML'2019, we propose a general theory of regularized Markov Decision Processes that generalizes these approaches in two directions: we consider a larger class of regularizers, and we consider the general modified policy iteration approach, encompassing both policy iteration and value iteration. The core building blocks of this theory are a notion of regularized Bellman operator and the Legendre-Fenchel transform, a classical tool of convex optimization. This approach allows for error propagation analyses of general algorithmic schemes of which (possibly variants of) classical algorithms such as Trust Region Policy Optimization, Soft Q-learning, Stochastic Actor Critic or Dynamic Policy Programming are special cases. This also draws connections to proximal convex optimization, especially to Mirror Descent.

7.3. Algorithms and Estimation for graph data

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

We consider the problem of graph estimation in a zero-inflated Gaussian model. In this model, zero-inflation is obtained by double truncation (right and left) of a Gaussian vector. The goal is to recover the latent graph structure of the Gaussian vector with observations of the zero-inflated truncated vector. We propose a two step estimation procedure. The first step consists in estimating each term of the covariance matrix by maximising the corresponding bivariate marginal log-likelihood of the truncated vector. The second one uses the graphical lasso procedure to estimate the sparsity of the precision matrix, which encodes the graph structure. We then state some theoretical results about the convergence rate of the covariance matrix and precision matrix estimators. These results allow us to establish consistency of our procedure with respect to graph structure recovery. We also present some simulation studies to corroborate the efficiency of our procedure. It is the object of the submitted paper [29], a part of the PhD thesis [1] and the communications [16] [15].

7.4. Regression and machine learning

Participants: E. Albuisson, T. Bastogne, S. Ferrigno, A. Gégout-Petit, F. Greciet, P. Guyot, C. Karmann, J.-M. Monnez, N. Sahki, S. Mézières, B. Lalloué

Through a collaboration with the pharmaceutical company Transgene (Strasbourg, France), we have developed a method for selecting covariates. The problem posed by Transgene was to establish patient profiles on the basis of their response to a treatment developed by Transgene. We have then proposed a new methodology for selecting and ranking covariates associated with a variable of interest in a context of high-dimensional data under dependence but few observations. The methodology successively intertwines the clustering of covariates, decorrelation of covariates using Factor Latent Analysis, selection using aggregation of adapted methods and finally ranking. A simulation study shows the interest of the decorrelation inside the different clusters of covariates. We have applied our method to the data of Transgene. For instance, transcriptomic data of 37 patients with advanced non-small-cell lung cancer who have received chemotherapy, to select the transcriptomic covariates that explain the survival outcome of the treatment. Our method has also been applied in another collaboration with biologists (CRAN laboratory, Nancy, France). In that case, our method has been applied to transcriptomic data of 79 breast tumor samples, to define patient profiles for a new metastatic biomarker and associated gene network. Our developed method is a contribution to the development of personalized medicine. We have published the method, as well as the two applications in [27].

In order to detect change of health state for lung-transplanted patient, we have begun to work on breakdowns in multivariate physiological signals. We consider the score-based CUSUM statistic and propose to evaluate the detection performance of some thresholds on simulation data. Two thresholds come from the literature: Wald's constant and Margavio's instantaneous threshold, and three contribution thresholds built by a simulation-based procedure: the first one is constant, the second instantaneous and the third is a dynamical version of

the previous one. The threshold's performance is evaluated for several scenarii, according to the detection objective and the real change in the data. The simulation results show that Margavio's threshold is the best at minimizing the detection delay while maintaining the given false alarm rate. But on real data, we suggest to use the dynamic instantaneous threshold because it is the easiest to build for practical implementation. It is the purpose of the communication [11] and the submitted paper [35].

We consider the problem of variable selection in regression models. In particular, we are interested in selecting explanatory covariates linked with the response variable and we want to determine which covariates are relevant, that is which covariates are involved in the model. In this framework, we deal with L1-penalised regression models. To handle the choice of the penalty parameter to perform variable selection, we develop a new method based on knockoffs. This revisited knockoffs method is general, suitable for a wide range of regressions with various types of response variables. Besides, it also works when the number of observations is smaller than the number of covariates and gives an order of importance of the covariates. Finally, we provide many experimental results to corroborate our method and compare it with other variable selection methods. It is the object of communication [17], the submitted paper [30] and a chapter of the PhD thesis [1].

In order to model crack propagation rate, continuous physical phenomenon that presents several regimes, we proposed a piecewise polynomial regression model under continuity and/or derivability assumptions as well as a statistical inference method to estimate the transition times and the parameters of each regime. We proposed several algorithms and studied their efficiency. The most efficient algorithm relies on dynamic programming. It is the object of the communication [14] and the PhD thesis of Florine Greciet.

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

In epidemiology, we are working with INSERM clinicians and biostatisticians to study fetal development in the last two trimesters of pregnancy. Reference or standard curves are required in this kind of biomedical problems. Values which lie outside the limits of these reference curves may indicated the presence of disorder. Data are from the French EDEN mother-child cohort (INSERM). It is a mother-child cohort study investigating the prenatal and early postnatal determinants of child health and development. 2002 pregnant women were recruited before 24 weeks of amenorrhoea in two maternity clinics from middle-sized French cities (Nancy and Poitiers). From May 2003 to September 2006, 1899 newborns were then included. The main outcomes of interest are fetal (via ultra-sound) and postnatal growth, adiposity development, respiratory health, atopy, behaviour and bone, cognitive and motor development. We are studying fetal weight that depends on the gestional age in the second and the third trimesters of mother's pregnancy. Some classical empirical and parametric methods as polynomials are first used to construct these curves. Polynomial regression is one of the most common parametric approach for modelling growth data espacially during the prenatal period. However, some of them require strong assumptions. So, we propose to work with semi-parametric LMS method, by modifying the response variable (fetal weight) with a Box-cox transformation. Nonparametric methods as Nadaraya-Watson kernel estimation or local polynomial estimation are also proposed to construct these curves. It is the object of the communication [28] and a paper is in progress. In addition, we want to develop a test, based on Z-scores, to detect any slope breaks in the fetal development curves (work in progress).

Many articles were devoted to the problem of recursively estimating eigenvectors corresponding to eigenvalues in decreasing order of the expectation of a random matrix using an i.i.d. sample of it. The present study makes the following contributions: the convergence of processes to normed eigenvectors is proved under two sets of more general assumptions, the observed random matrices are no more supposed i.i.d.; moreover, the scope of these processes is widened. The application to online principal component analysis of a data stream is treated, assuming that data are realizations of a random vector Z whose expectation is unknown and is estimated online, as well as possibly the metric used when it depends on unknown characteristics of Z; two types of processes are studied: we are no more bound to use a data mini-batch at each step, but we can use all previously observed data up to the current step without storing them, thus taking into account all the information contained in previous data. The conducted experiments have shown that processes of the second type are faster than those of the first type. It is the object of the submitted paper [32] and the communication [21].

The study addresses the problem of constrained binary logistic regression, particularly in the case of a data stream, using a stochastic approximation process. To avoid a numerical explosion which can be encountered, we propose to use a process with online standardized data instead of raw data. This type of process can also be used when we have to standardize the explanatory variables, for example in the case of a shrinkage method such as LASSO. Herein, we define and study the almost sure convergence of an averaged constrained stochastic gradient process with online standardized data. Moreover we propose to use a piecewise constant step-size in order that the step-size does not decrease too quickly and reduce the speed of convergence. Processes of this type are compared to classical processes on real and simulated datasets. The results of conducted experiments confirm the validity of the choices made. This will be used in an ongoing application to online updating of a score in heart failure patients. It is the object of the submitted paper [31] and the communications [20],[19].

CAMUS Project-Team

7. New Results

7.1. The Polyhedral Model Beyond Loops

Participants: Salwa Kobeissi, Philippe Clauss.

There may be a huge gap between the statements outlined by programmers in a program source code and instructions that are actually performed by a given processor architecture when running the executable code. This gap is due to the way the input code has been interpreted, translated and transformed by the compiler and the final processor hardware. Thus, there is an opportunity for efficient optimization strategies, that are dedicated to specific control structures and memory access patterns, to be applied as soon as the actual runtime behavior has been discovered, even if they could not have been applied on the original source code.

We develop this idea by identifying code excerpts that behave as polyhedral-compliant loops at runtime, while not having been outlined at all as loops in the original source code. In particular, we are interested in recursive functions whose runtime behavior can be modeled as polyhedral loops. Therefore, the scope of this study exclusively includes recursive functions whose control flow and memory accesses exhibit an affine behavior, which means that there exists a semantically equivalent affine loop nest, candidate for polyhedral optimizations. Accordingly, our approach is based on analyzing early executions of a recursive program using a Nested Loop Recognition (NLR) algorithm [3], performing the affine loop modeling of the original program runtime behavior, which is then used to generate an equivalent iterative program, finally optimized using the polyhedral compiler Polly. We present some preliminary results showing that this approach brings recursion optimization techniques into a higher level in addition to widening the scope of the polyhedral model to include originally non-loop programs.

This work is the topic of Salwa Kobeissi's PhD. A first paper has been published at the 9th International Workshop on Polyhedral Compilation Techniques [22].

7.2. New release of Apollo

Participants: Muthena Abdul-Wahab, Philippe Clauss.

Apollo has been updated to use LLVM/Clang version 6.0.1. The unmodified sources are now included, as tar-files, in the APOLLO distribution.

Regarding the build system:

- All components of APOLLO are now installed into the installation directory. Once installed, APOLLO does not need the build directory to be kept.
- The RPATH on APOLLO libraries has been set to the installation directory. This allows APOLLO to be run without having to set up library paths.
- APOLLO_BUILD_JOBS has been introduced to specify the maximum number of build jobs to use. The replaces NB_JOBS which is still supported but deprecated.
- The sources for external dependencies are now included in the APOLLO distribution. They are no longer downloaded during a build.
- A new build target 'check' has been added to run the testsuite. This is supported by Makefiles ('make check') and Ninja ('ninja check').
- The build type (Debug/Release) for LLVM/Clang is now the same as the rest of APOLLO. New build variable APOLLO_LLVM_BUILD_TYPE can be used to specify a separate build type for LLVM/Clang.

Regarding bug fixes:

- Valid code using floating point types (float or double) could make APOLLO stop with an message about unsupported scalars. This has been fixed by removing the Loop Invariant Code Motion (LICM) pass in such cases, preventing floating-point scalars to be generated.
- Code containing try-catch blocks could make APOLLO crash. This has been fixed.
- Dynamic loop bounds were no more instrumented and interpolated. This has been fixed.

7.3. Uniform Random Sampling in Polyhedra

Participant: Philippe Clauss.

We propose a method for generating uniform samples among a domain of integer points defined by a polyhedron in a multi-dimensional space. The method extends to domains defined by parametric polyhedra, in which a subset of the variables are symbolic. We motivate this work by a list of applications for the method in computer science. The proposed method relies on polyhedral ranking functions, as well as a recent inversion method for them, named *trahrhe* expressions. This work has been accomplished in collaboration with Benoît Meister from Reservoir Labs, New York, USA, and has been published at the 10th International Workshop on Polyhedral Compilation Techniques, January 2020.

7.4. Runtime Multi-Versioning and Specialization

Participant: Philippe Clauss.

We have developped an extension of APOLLO that implements code multi-versioning and specialization to optimize and parallelize loop kernels that are invoked many times with varying parameters. These parameters may influence the code structure, the touched memory locations, the workload, and the runtime performance. They may also impact the validity of the parallelizing and optimizing polyhedral transformations that are applied on-the-fly.

For a target loop kernel and its associated parameters, a different optimizing and parallelizing transformation is evaluated at each invocation, among a finite set of transformations (multi-versioning and specialization). The best performing transformed code version is stored and indexed using its associated parameters. When every optimizing transformation has been evaluated, the best performing code version regarding the current parameters, which has been stored, is relaunched at next invocations (memoization).

This work has been accomplished in collaboration with Raquel Lazcano and Eduardo Juarez of the Universidad Politécnica de Madrid, Spain, and has been published at the ACM SIGPLAN 2020 International Conference on Compiler Construction (CC 2020).

7.5. AutoParallel: Automatic parallelization and distributed execution of affine loop nests in Python

Participant: Philippe Clauss.

The last improvements in programming languages and models have focused on simplicity and abstraction; leading Python to the top of the list of the programming languages. However, there is still room for improvement when preventing users from dealing directly with distributed and parallel computing issues. We propose AutoParallel, a Python module to automatically find an appropriate task-based parallelisation of affine loop nests to execute them in parallel in a distributed computing infrastructure. This parallelization can also include the building of data blocks to increase tasks' granularity in order to achieve a good execution performance. Moreover, AutoParallel is based on sequential programming and only contains a small annotation in the form of a Python decorator so that anyone with intermediate-level programming skills can scale up an application to hundreds of cores.

This work has been accomplished in collaboration with Cristian Ramon-Cortes, Ramon Amela, Jorge Ejarque and Rosa M. Badia of the Barcelona Supercomputing Center (BSC), Spain. A journal paper is in preparation.

7.6. Combining Locking and Data Management Interfaces

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

An implementation (ORWL) of our ideas of combining control and data management in C has been undertaken, see Section 6.3 . In previous work it has demonstrated its efficiency for a large variety of platforms.

In the framework of the ASNAP project we have used ordered read-write locks (ORWL) as a model to dynamically schedule a pipeline of parallel tasks that realize a parallel control flow of two nested loops; an outer *iteration* loop and an inner *data traversal* loop. Other than dataflow programming, for each individual data object we conserve the same modification order as the sequential algorithm. As a consequence the visible side effects on any object can be guaranteed to be identical to a sequential execution. Thus the set of optimizations that are performed are compatible with C's abstract state machine and compilers could perform them, in principle, automatically and unobserved. See [16] for first results.

In the context of the Prim'Eau project (see 9.1.2) we use ORWL to integrate parallelism into an already existing `Fortran` application that computes floods in the region that is subject to the study. A first step of such a parallelization has been started by using ORWL on a process level. Our final goal will be to extend it to the thread level and to use the application structure for automatic placement on compute nodes. A first step to this goal has been a specific decomposition of geological data, see [21].

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. This year, Daniel has successfully defended his thesis, see [7].

7.7. Granularity Control for Parallel Programs

Participant: Arthur Charguéraud.

Arthur Charguéraud studied 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.

The proposed approach combines the use of asymptotic complexity functions provided by the programmer, with runtime measurements to estimate the constant factors that apply. Exploiting these two sources of information makes it possible 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 [52], extending them with a carefully-designed algorithm for ensuring convergence of the estimation of the constant factors deduced from the measures, even in the face of noise and cache effects, which are taken into account in the analysis. The approach is demonstrated on a range of benchmarks taken from the state-of-the-art PBBS benchmark suite. These results have been accepted for publication at PPoPP'19 [14].

7.8. Program Verification and Formal Languages

Participant: Arthur Charguéraud.

- Armaël Guéneau, a PhD student advised by A. Charguéraud and F. Pottier (Cambium), has developed a formal proof of the functional correctness and the asymptotic complexity of a state-of-the-art incremental cycle detection algorithm due to Bender, Fineman, Gilbert, and Tarjan. This work moreover proposes a simple change that allows the algorithm to be regarded as genuinely online. The verification proof is carried out by exploiting Separation Logic with Time Credits, in the CFML tool, to simultaneously verify the correctness and the worst-case amortized asymptotic complexity of the modified algorithm. This work was published at ITP'19 [17]. It leverages previous work on the extension of the CFML verification tool to allow the specification of the asymptotic complexity of higher-order, imperative programs [55], and shows that this framework scales up to larger, more complex programs.
- Arthur Charguéraud, together with Jean-Christophe Filliâtre and Cláudio Lourenço (CNRS, Inria and Université Paris Saclay), and Mário Pereira (NOVA LINCS & DI, Universidade Nova de Lisboa), developed a behavioral specification language for OCaml, called GOSPEL. It is designed to enable modular verification of data structures and algorithms. Compared with writing specifications directly in Separation Logic, it provides a high-level syntax that greatly improves conciseness and makes it accessible to programmers with no familiarity with Separation Logic. GOSPEL is applied to the development of a formally verified library of general-purpose OCaml data structures. This work was published at the World Congress on Formal Methods (FM) 2019 [15].

7.9. Improvement of Schnaps on multi-GPU nodes using the LAHeteroprio Scheduler

Participant: Bérenger Bramas.

The TONUS team has developed Schnaps, a discontinuous finite element solver with OpenCL and StarPU. The team members have been facing challenges in the scalability of their application when using more than one GPU. This has been the starting point of a collaboration in which Bérenger Bramas has participated in the development of Schnaps and plugged its StarPU scheduler called LAHeteroprio [9]. The improvements obtained were significant and included in a paper [50] (currently under revision).

The potential of LAHeteroprio is now demonstrated. However, setting up this scheduler remains a complicated task. Therefore, we plan to work on its automatic configuration, which will require us to perform on the fly analysis of the graph of tasks.

7.10. Improving Parallel Executions by Increasing Task Granularity in Task-based Runtime Systems using Acyclic DAG Clustering

Participants: Bérenger Bramas, Alain Ketterlin.

Bérenger Bramas and Alain Ketterlin collaborate with the TONUS team in the development of a parallel solver for the resolution of conservative hyperbolic upwind kinetic of unstructured tokamaks [49]. In their methods, they must solve the transport equation on an unstructured mesh, which can be seen as having a wave propagating from neighbor-to-neighbor. The resulting computation can be represented using a direct acyclic graph (DAG) of operations, where each operation is a tiny task. Therefore, Bérenger Bramas and Alain Ketterlin contributed mainly on two aspects. First, they have proposed a highly optimized lock-free parallel implementation of the solution based on atomic instructions. Second, they have improved an existing algorithm from the literature to cluster a DAG of tasks with the aim of increasing the granularity of the tasks and to reduce the overhead of the parallelization consequently. This new approach has been accepted in a dedicated paper (accepted but not yet published).

7.11. FMM Kernel for the Integral Equation Formulation of the N-body Dielectric Spheres Problem

Participant: Bérenger Bramas.

Bérenger Bramas worked with Benjamin Stamm and Muhammad Hassan (RWTH) to create a kernel for the fast multipole method (FMM). The kernel relies on the previously developed kernel with spherical harmonics and accelerated by rotations. It has been extended to accept spherical harmonics (with orders different from the ones used in the kernel) instead of points as input. The kernel allowed us to accelerate the computation and was used for a complexity analysis that has been submitted [54].

7.12. Automatic Task-Based Parallelization using Source to Source Transformations

Participants: Bérenger Bramas, Garip Kusolgu.

Bérenger Bramas and Garip Kusolgu worked on a new approach to parallelize automatically any application written in an object-oriented language. The main idea is to parallelize a code as an HPC expert would do it using the task-based method. With this aim, they created a new source-to-source compiler on top of CLang-LLVM called APAC. APAC is able to insert tasks in a source-code by evaluating data accesses and thus generating the correct dependencies. An important and challenging part of the work consists in managing the granularity, which requires to work both statically on the code but also by delegating decisions at runtime.

7.13. Large Scale Particle Fusion Algorithm for Tracing Systems in Fluid Mechanics Applications

Participant: Bérenger Bramas.

Bérenger Bramas worked with Michael Wilczek and Cristian Lalescu (Max Planck Institute for Dynamics and Self-Organization) in designing a new method to merge particles in a large scale application (*i.e.,* designed to run on thousands of computing nodes). In this context, the particles are originally used in a tracing system to extract information from a vector field in fluid mechanics. However, the physicists are now interested having the particles interacting and even fusioning. Due to the constraints of large scale computing, the system tries to reduce the number and amount of communications. This development has been done in the TurTLE application (not publicly available) and is currently under evaluation.

7.14. Pipelined Multithreaded Code Generation

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

State-of-the-art automatic polyhedral parallelizers extract and express parallelism as isolated parallel loops. For example, the Pluto high-level compiler generates and annotates loops with `#pragma omp parallel for` directives. In this work, we took advantage of pipelined multithreading, a parallelization strategy that can address a wider class of codes, currently not handled by automatic parallelizers. Pipelined multithreading requires interlacing iterations of some loops in a controlled way that enables the parallel execution of these iterations.

This work has been accepted for presentation at the International Workshop on Polyhedral Compilation Techniques (IMPACT 2020), in conjunction with HiPEAC '20 (Jan. 2020, Bologna, Italy).

7.15. Raster Image Processing (RIP) Optimization

Participants: Cédric Bastoul, Paul Godard, Vincent Loechner.

In the context of our collaboration with the Caldera company, we are interested in original challenges for the computer systems in charge of driving very wide printer farms and very fast digital presses.

We explored new approaches inspired by the high performance computing field to speedup the graphics processing (RIP) necessary to digital printing. To achieve this goal, we developed a distributed system which provides the adequate flexibility and performance by exploiting and optimizing both processing and synchronization techniques. Our architecture meets the specific constraints on generating streams for printing purpose. We performed an evaluation of our solution and provided experimental evidence of its great performance and viability. This work has been presented at the 2019 IEEE International Parallel and Distributed Processing Symposium Workshop (IPDPSW): PDSEC '19, in May 2019, Rio de Janeiro.

The second topic we worked on during this collaboration is an out-of-core and out-of-place rectangular matrix transposition and rotation algorithm. An originality of our processing algorithm is to rely on an optimized use of the page cache mechanism. It is parallel, optimized by several levels of tiling and independent of any disk block size. We evaluated our approach on four common storage configurations: HDD, hybrid HDD-SSD, SSD and software RAID 0 of several SSDs. We showed that it brings significant performance improvement over a hand-tuned optimized reference implementation developed by the Caldera company and we confront it against the roofline speed of a straight file copy. This work is under submission in the IEEE Transaction on Computers.

Paul Godard has defended his PhD thesis on Dec. 16th, 2019.

7.16. Static Versus Dynamic Memory Allocation

Participant: Vincent Loechner.

Vincent Loechner and Toufik Baroudi (PhD student, Univ. Batna, Algeria) compared the performance of linear algebra kernels using different array allocation modes: as static declared arrays or as dynamically allocated arrays of pointers. They studied the possible reasons of the difference in performance of parallelized or sequential linear algebra kernels on two different architectures: an AMD (Magny-Cours) and an Intel Xeon (Haswell-EP). Static or dynamic memory allocation has an impact on performance in many cases. Both the processor architecture and the compiler can provoke significant and sometimes surprising variations in the number of cache misses and vectorization opportunities taken by the compiler.

This work has been accepted for presentation at the International Workshop on Polyhedral Compilation Techniques (IMPACT 2020), in conjunction with HiPEAC '20 (Jan. 2020, Bologna, Italy).

7.17. Automatic Adaptive Approximation for Stencil Computations

Participants: Maxime Schmitt, Cédric Bastoul.

This work has been done in collaboration with Philippe Helluy (TONUS).

Approximate computing is necessary to meet deadlines in some compute-intensive applications like simulation. Building them requires a high level of expertise from the application designers as well as a significant development effort. Some application programming interfaces greatly facilitate their conception but they still heavily rely on the developer's domain-specific knowledge and require many modifications to successfully generate an approximate version of the program. In this work we designed new techniques to semi-automatically discover relevant approximate computing parameters. We believe that superior compiler-user interaction is the key to improved productivity. After pinpointing the region of interest to optimize, the developer is guided by the compiler in making the best implementation choices. Static analysis and runtime monitoring are used to infer approximation parameter values for the application. We evaluated these techniques on multiple application kernels that support approximation and show that with the help of our method, we achieve similar performance as non-assisted, hand-tuned version while requiring minimal intervention from the user.

These techiques and the underlying compiler infrastructure are a significant output of collaboration with the Inria Nancy - Grand Est team TONUS, specialized on applied mathematics (contact: Philippe Helluy), to bring models and techniques from this field to compilers. A paper presenting these extensions has been accepted to the CC international conference [18].

Maxime Schmitt has defended his PhD thesis on Sep. 30th, 2019 [8].

CAPSID Project-Team

7. New Results

7.1. Axis 1 : New Approaches for Knowledge Discovery in Structural Databases

7.1.1. Biomedical Knowledge Discovery

Our collaboration with clinicians at the CHRU Nancy in the framework of the RHU FIGHT-HF program and of the Contrat d'Interface has lead to two publications demonstrating the added value of database and knowledge graph exploitation when analyzing observational or prospective cohorts. In a retrospective observational study, we have identified and characterized patient subgroups presenting stable or unstable positivity to anti-phospholipid antobodies assays [15]. In the European FibroTarget cohort study, we have contributed to the characterization of at-risk phenotypic groups using proteomic biomarkers [16].

Another application is carried out in collaboration with the Orpailleur Team and concerns the PraktikPharma ANR project. We aim at building explanations for severe drug side effects (such as drug-induced liver injury or severe cutaneous adverse reaction) from pharmacogenomics RDF graph (PGXlod). We obtained a podium abstract at the MedInfo 2019 conference for providing molecular characterization for unexplained adverse drug reactions using pharmacogenomics RDF graph (PGXlod) [30].

7.1.2. Stochastic Decision Trees for Similarity Computation

In the frame of Kévin Dalleau's PhD thesis, we have designed a method to compute similarities on unlabeled data using stochastic decision trees [31], [27]. The main idea of Unsupervised Extremely Randomized Trees (UET) is to randomly and iteratively split the data until a stopping criterion is met. Pairwise similarity values are computed based on the co-occurrence of samples in the leaves of each generated tree. We evaluate our method on synthetic and real-world datasets by comparing the mean similarities between samples with the same label and the mean similarities between samples with distinct labels. Empirical studies show that the method effectively gives distinct similarity values between samples belonging to distinct clusters, and gives indiscernible values when there is no cluster structure. We also assessed some interesting properties such as invariance under monotone transformations of variables and robustness to correlated variables and noise. Our experiments show that the algorithm outperforms existing methods in some cases, and can reduce the amount of preprocessing needed with many real-world datasets. We extended the approach to the computation of pairwise similarity for graph nodes. The experimental results are competitive with state of the art methods. We are currently working on merging the two similarity methods (on attribute-value objects and on graph nodes) to attributed graphs where the nodes are described by attributes.

We plan to study the application of this pairwise similarity computation to quantify protein structural similarities. Two interesting problems will concern the representation of the protein structure and how to tackle extra constraints such as invariance under rotational and translational transformations.

7.1.3. Protein Annotation and Machine Learning

We have been involved in the 3rd international CAFA Challenge ("Critical Assessment of Functional Annotation") through our work on (i) domain functional annotation (Zia Alborzi's PhD thesis) and (ii) label propagation in graphs (Bishnu Sarker's PhD thesis). We were therefore contributors of the general report published this year [23].
As part of his PhD work, Bishnu Sarker developed and tested on UniProt/SwissProt a new method for functional annotation of proteins using domain embedding-based sequence classification [25].

Multiple Instance Learning (MIL) is a machine learning strategy that can be applied to sets of sequences describing organisms displaying a given property. The purpose here is to be able to classify a new organism with respect to this property based on its sequences and their similarity to the sequences of classified organisms. New MIL algorithms have been described and tested in the framework of a collaboration [26], [24]. Another collaborative work has lead to the development of a distributed algorithm for large-scale graph clustering [34].

7.2. Axis 2 : Integrative Multi-Component Assembly and Modeling

7.2.1. EROS-DOCK algorithm and its extensions

We have adapted our EROS-DOCK protein-protein docking software [35], [19] to account for experimental knowledge on the protein-protein interface to be modeled. Indeed, structural biology experiments can identify pairs of amino-acids from each protein in a protein-protein interface that are likely to be in close contact. This additional restraint is used to pre-prune the 3D rotational space of one protein toward another, by eliminating cones of rotations that cannot fulfill the distance between the two points at the protein surfaces. Using a single restraint permits to decrease the average execution time by at least 90 percent.

We also developed a new version of EROS-DOCK for multi-body docking (modeling assemblies of more than 2 proteins), using a combinatorial approach. We assembled trimers by docking in a first stage all possible combinations of pairs of proteins involved in the multi-body complex. Possible trimer solutions are assembled by fixing one protein, the "root-protein" (protein A, say) at the origin and by placing the other two around it using the transformations,T[AB] and T[AC], from the corresponding pairwise solution lists returned by EROS-DOCK. If the three transformations together form a near-native (biologically relevant) trimer structure, then it is natural to suppose that T[BC] should be found in the list of B-C pairwise solutions.

Both extensions of the EROS-DOCK algorithm reported last year and published early this year [19] have been presented by Maria-Elisa Ruiz Echartea at the 2019 CAPRI meeting in april 2019 (http://www.capri-docking.org/events/) and at the MASIM meeting in november 2019 [28]. These results are part of her PhD Thesis that was defended on december 18, 2019 (the thesis will soon be available on HAL). A paper describing EROS-DOCK adaptation to multi-body docking in under revision in *Proteins*.

7.2.2. Protein docking

The regular participation of the Capsid team to the CAPRI challenge is acknowledged through its contribution to the review published this year on CAPRI round 46 [17].

We also contributed to an evaluation of docking software performance in protein-glycosaminoglycan systems [22].

7.2.3. 3D modeling and virtual screening

We have built a 3D model by homology of a new class of relaxase involved in the horizontal transfer of DNA in a group of bacteria called Firmicutes [21].

We also built a 3D model of a chemosensory GPCR as a potential target to control a parasite in plants [13].

Virtual screening was applied on various targets in a re-purposing strategy and led to the discovery of small molecules active against invasive fungal disease [14], [18].

CARAMBA Project-Team

7. New Results

7.1. Algebraic Curves for Cryptology

7.1.1. Cocks-Pinch Curves of Embedding Degrees Five to Eight and Optimal Ate Pairing Computation

Participants: Aurore Guillevic, Simon Masson, Emmanuel Thomé.

In [21] we explored a modification of the Cocks-Pinch method to generate pairing-friendly curves resistant to the Special-Tower-NFS algorithm (STNFS). We carefully estimated the cost of the STNFS attack for existing families of curves, and chose curves of embedding degree five to eight. For prime embedding degrees 5 and 7, our curves are naturally immune to the STNFS attack, but their performance level is not high. For composite embedding degrees 6 and 8 for which the TNFS attack applies, we chose the parameters from a family that is general enough to thwart the "special" variant STNFS; we also optimized these parameter choices so that these curves can have a reasonably efficient pairing computation, close with the very best possible curve choices.

7.1.2. A Short-List of Pairing-Friendly Curves Resistant to Special TNFS at the 128-bit Security Level

Participant: Aurore Guillevic.

The preprint [20] applies the refinements of the paper [22] to estimate the cost of the Special Tower NFS algorithm for particular pairing-friendly curves, whose target group is $\mathbb{F}_{p^n}$, and where the characteristic is special, parameterized by a low degree polynomial. We show that with a new variant of the polynomial selection, the estimated cost is reduced, but stays above the theoretical bound of the Special NFS $L_{p^n}(1/3, (32/9)^{1/3})$. This variant does not apply to the Cocks-Pinch curves of [21]. We list nine interesting pairing-friendly curves of embedding degrees between 10 and 16 at the 128-bit security level.

7.1.3. A Practical Attack on ECDSA Implementations Using wNAF Representation

Participants: Gabrielle de Micheli, Cécile Pierrot, Rémi Piau.

ECDSA is a widely deployed public key signature protocol that uses elliptic curves. One way of attacking ECDSA with wNAF implementation for the scalar multiplication is to perform a side-channel analysis to collect information, then use a lattice based method to recover the secret key. In [18], we re-investigate the construction of the lattice used in one of these methods, the Extended Hidden Number Problem (EHNP). We find the secret key with only 3 signatures, thus reaching the theoretical bound never achieved before. Our attack is more efficient than previous attacks, has better probability of success, and is still able to find the secret key with a small amount of erroneous traces, up to 2% of false digits.

7.1.4. Algorithmic Aspects of Elliptic Bases in Finite Field Discrete Logarithm Algorithms

Participant: Cécile Pierrot.

Elliptic bases give an elegant way of representing finite field extensions and were used as a starting point for small characteristic finite field discrete logarithm algorithms. This idea has been proposed by two groups, in order to achieve provable quasi-polynomial time algorithms for computing discrete logarithms in small characteristic finite fields. In [23], together with Antoine Joux, we do not try to achieve a provable algorithm, but instead we investigate the practicality of heuristic algorithms based on elliptic bases.

7.1.5. A Fast Randomized Geometric Algorithm for Computing Riemann-Roch Spaces

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

In [7], we proposed a probabilistic variant of Brill-Noether's algorithm for computing a basis of the Riemann-Roch space $L(D)$ associated to a divisor D on a projective plane curve $\mathcal{C}$ over a sufficiently large perfect field k. Most of the results of this work have been obtained in 2018. In 2019, we have strengthened these results and revised the associated paper. This new version of the paper has been accepted for publication in the journal Mathematics of Computation.

7.1.6. Counting Points on Hyperelliptic Curves

Participants: Pierrick Gaudry, Pierre-Jean Spaenlehauer.

Two works with Simon Abelard [1], [2] following his PhD thesis about improved complexities for counting point algorithms of hyperelliptic curves with or without real multiplication are now formally published as journal articles.

7.1.7. Verifiable Delay Functions from Supersingular Isogenies and Pairings

Participant: Simon Masson.

Together with Luca De Feo, Christophe Petit and Antonio Sanso, we introduce in [11] two verifiable delay functions based on isogenies of supersingular elliptic curves and pairing. We discuss both the advantages and drawbacks of our constructions, we study their security and we demonstrate their practicality with a proof-of-concept implementation. This work appears in the proceedings of ASIACRYPT'2019.

7.1.8. Isogeny Graphs With Maximal Real Multiplication

Participant: Emmanuel Thomé.

Emmanuel Thomé and Sorina Ionica (post-doctoral fellow in the former CARAMEL team in 2012) worked on a new algorithm for computing isogeny graphs for Jacobians of curves having the special property that the intersection of their endomorphism ring with its real subfield is maximal. The resulting algorithm is the first depth-first algorithm for this task. The work [6] was finally published.

7.2. The Number Field Sieve – High-Level Results

7.2.1. A New Ranking Function for Polynomial Selection in the Number Field Sieve

Participant: Paul Zimmermann.

With Nicolas David (ÉNS Paris-Saclay, France), we designed a new ranking function for polynomial selection in the Number Field Sieve. The previous ranking function was only considering the *mean* of the so-called α-value, which measures how small primes divide the norm of the polynomial. The new function also takes into account the *variance* of the corresponding distribution. This partially explains why the previous function did sometimes fail to correctly identify the best polynomials. The new ranking function is implemented in Cado-NFS (branch `dist-alpha`) and is detailed in [3].

7.2.2. On the Alpha Value of Polynomials in the Tower Number Field Sieve Algorithm

Participant: Aurore Guillevic.

With Shashank Singh from IISER Bhopal (former post-doc at CARAMBA in 2017), we generalized the ranking function α for the Tower setting of the Number Field Sieve in [22]. In the relation collection of the NFS algorithm, one tests the smoothness of algebraic norms (computed with resultants). The α function measures the bias of the average valuation at small primes of algebraic norms, compared to the average valuation at random integers of the same size. A negative α means more small divisors than average. We then estimate the total number of relations with a Monte-Carlo simulation, as a generalized Murphy's E function, and finally give a rough estimate of the total cost of TNFS for finite fields $\mathbb{F}_{p^k}$ of popular pairing-friendly curves.

7.2.3. Faster Individual Discrete Logarithms in Finite Fields of Composite Extension Degree

Participant: Aurore Guillevic.

We improved the previous work [30] on speeding-up the first phase of the individual discrete logarithm computation, the initial splitting, a.k.a. the smoothing phase. We extended the algorithm to any non-prime finite field $\mathbb{F}_{p^n}$ where n is composite. We also applied it to the new variant Tower-NFS. The paper was finally published in 2019 [4].

7.3. The Number Field Sieve – Implementation Results

7.3.1. Parallel Structured Gaussian Elimination for the Number Field Sieve

Participant: Paul Zimmermann.

Together with Charles Bouillaguet (University of Lille, France), we completely re-designed the structured Gaussian elimination step of Cado-NFS (called `merge`). The new algorithm is fully parallel, and scales quite well. With 32 cores on modern hardware, the `merge`-step of RSA-512 (factored in 1999) now takes only 20 seconds, and for the hidden SNFS DLP-1024 record (done in 2017) it takes only 140 seconds [16].

7.4. Computer Arithmetic

7.4.1. Breaking Randomized Mixed-Radix Scalar Multiplication Algorithms

Participant: Jérémie Detrey.

Together with Laurent Imbert (LIRMM, France), we designed in [13] an attack against a recently published randomized elliptic-curve scalar multiplication scheme based on covering systems of congruences. We also proposed a more robust algorithm based on a mixed-radix representation of the scalar. However, under strong security hypotheses, this algorithm may still allow a virtual powerful attacker to recover much more information than what was first expected. This led us to the conclusion that randomized algorithms based on the mixed-radix number system should be avoided.

7.5. Symmetric Cryptology

7.5.1. Vectorial Boolean Functions with Very Low Differential-Linear Uniformity Using Maiorana-McFarland Type Construction

Participant: Bimal Mandal.

With Deng Tang and Subhamoy Maitra, we constructed in [14] a new class of balanced vectorial Boolean functions with very low differential-linear uniformity, whose coordinate functions are derived by modifying the Maiorana–McFarland bent functions. Further, we provided a combinatorial count of hardware gates required to implement such circuits.

7.5.2. Analysis of Boolean Functions in a Restricted (Biased) Domain

Participant: Bimal Mandal.

This work with Subhamoy Maitra, Thor Martinsen, Dibyendu Roy and Pantelimon Stanica [8] is a substantially revised and extended version of the paper "Tools in analyzing linear approximation for Boolean functions related to FLIP" that appeared in the proceedings of Indocrypt 2018 [32]. We proposed a technique to study the cryptographic properties of Boolean functions, whose inputs do not follow uniform distribution, and obtain a lower bound for the bias of the nonlinear filter function of FLIP by using biased Walsh–Hadamard transform. Our results provided more accurate calculation of the biases of Boolean function over restricted domain, which help to determine the security parameter of FLIP type ciphers.

7.5.3. Forkcipher: a New Primitive for Authenticated Encryption of Very Short Messages

Participant: Virginie Lallemand.

Together with Elena Andreeva, Antoon Purnal, Reza Reyhanitabar, Arnab Roy and Damian Vizár, we proposed a candidate to the NIST Lightweight competition that we also published at Asiacrypt 2019 [10]. Our proposal is based on the so-called forkcipher construction that was previously presented and investigated by a subset of the authors and which provides authenticated encryption optimized for short messages. Our NIST candidate is called ForkAE, and as required by NIST it is based on well investigated primitives, out of which the Skinny tweakable cipher. ForkAE is one of the 32 candidates that were selected to continue to Round 2 out of 56 valid submissions.

7.5.4. Computing AES Related-Key Differential Characteristics With Constraint Programming

Participant: Marine Minier.

In [5], with David Gérault, Pascal Lafourcade, and Christine Solnon, we improve existing Constraint Programming (CP) approaches for computing optimal related-key differential characteristics: we add new constraints that detect inconsistencies sooner, and we introduce a new decomposition of the problem in two steps. These improvements allow us to compute all optimal related-key differential characteristics for AES-128, AES-192 and AES-256 in a few hours.

7.5.5. Participation in the NIST Lightweight Cryptography Standardization Process

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

The team is actively taking part in the lightweight cryptography standardization process of the NIST. The two major actions that have been taken are the following:

- Proposition of two candidates, namely Lilliput-AE (Alexandre Adomnicai, Thierry P. Berger, Christophe Clavier, Julien Francq, Paul Huynh, Virginie Lallemand, Kévin LeGouguec, Marine Minier, Léo Reynaud and Gaël Thomas) and ForkAE (Elena Andreeva, Virginie Lallemand, Antoon Purnal, Reza Reyhanitabar, Arnab Roy and Damian Vizár). ForkAE made it to the second round, but unfortunately a weak point has been detected in the design of Lilliput-AE.
- Organization of regular cryptanalysis meetings with other french cryptographers. Since the publication of the 56 proposals, four meetings have been held and some tangible results have already been achieved. As an example, the meeting participants found a practical differential forgery attack against the proposal named *Quartet*. The details have been made public on the NIST mailing list and they made the NIST remove this candidate from consideration.

7.5.6. Cryptanalysis of SKINNY in the Framework of the SKINNY 2018-2019 Cryptanalysis Competition

Participant: Virginie Lallemand.

Together with Patrick Derbez (University of Rennes) and Aleksei Udovenko (University of Luxembourg) we investigated in [12] the security of the SKINNY tweakable block cipher, a lightweight symmetric cipher proposed at Crypto in 2016. Our setting was the one of the SKINNY 2018-2019 Cryptanalysis Competition, that is we looked for attacks that can be run in practical time and that succeed with a data set reduced to the provided set of 2^{20} (plaintext, ciphertext). We solved the challenges (meaning that we experimentally recovered the 128-bit key) for up to 10-round SKINNY-128-128 and 12-round SKINNY-64-128. To this day these are the best results reported in this setting.

7.6. E-voting

7.6.1. Belenios: a Simple Private and Verifiable Electronic Voting System

Participant: Pierrick Gaudry.

In [9], written with Véronique Cortier and Stéphane Glondu, we have summarized the current state of our voting platform Belenios. It was the occasion to put in a single place the description of several sub-parts of the protocol that are otherwise spread in many articles. We also made statistics regarding the use of the platform during the year 2018, and discussed how security features were or were not activated by the users.

7.6.2. A Simple Alternative to Benaloh Challenge for the Cast-as-Intended Property in Helios/Belenios

Participant: Pierrick Gaudry.

In a short note [17] written with Véronique Cortier, Jannik Dreier, and Mathieu Turuani from the PESTO team, we propose a simple technique that can be added to an Helios-like e-voting protocol, so that the voter can check whether their potentially infected computer has not silently changed their vote.

7.6.3. Breaking the Encryption Scheme of the Moscow Internet Voting System

Participant: Pierrick Gaudry.

In [19], written in collaboration with Alexander Golovnev (Harvard), we explain the vulnerabilities we have found in an Internet voting system used for the election for the representatives of the Moscow Duma that took place in September 2019. The weaknesses in the encryption scheme (based on the discrete logarithm problem in finite fields) were found in the source code that was made available in July 2019 as part of a public testing.

COAST Project-Team

7. New Results

7.1. Trustworthy Collaboration

Participants: Claudia-Lavinia Ignat, Hoang Long Nguyen, Olivier Perrin.

In order to test user acceptance of a collaboration model based on automatic trust assessment, we designed an experiment relying on the trust game. In the trust game money exchange is entirely attributable to the existence of trust between users. Our experimental design [7] tested variations of the trust game: with and without showing the partner identity and with and without explicit computation of partner trust values based on the computational trust model we previously proposed. We organized a user study with 30 participants that confirmed that the availability of this trust metric improves user cooperation and that it predicts participants future behavior. We showed that trust score availability has the same effect as an identity to improve cooperation. Our study suggests that trust score could function as an enhancement or even replacement of traditional identity systems and has the advantage of scalability.

In the scope of Hoang Long Nguyen's PhD thesis, we proposed the architecture of ÔBlock, an open ecosystem for quick development of transparent applications based on consortium blockchain.

7.2. Undo in Collaborative Editing

Participants: Victorien Elvinger, Claudia-Lavinia Ignat.

In collaborative editors a selective undo allows a user to undo an earlier operation, regardless of when, where and by which user the operation was generated. In most existing collaborative editors such as GoogleDrive, selective undo is not integrated and users can only undo their own operations but not the ones generated by the other users. There is currently no generally applicable undo support as stated in the manifesto on CRDTs [17]. We presented a generic support of selective undo for CRDTs by proposing an abstraction that captures the semantics of concurrent undo and redo operations through equivalence classes. The abstraction is a natural extension of undo and redo in sequential applications and is straightforward to implement in practice [9].

7.3. Mitigating the Cost of Identifiers in Sequence CRDT

Participants: Matthieu Nicolas, Gérald Oster, Olivier Perrin.

To achieve high availability, large-scale distributed systems have to replicate data and to minimise coordination between nodes. The literature and industry increasingly adopt Conflict-free Replicated Data Types (CRDTs) to design such systems. CRDTs are data types which behave as traditional ones, e.g. the Set or the Sequence. However, compared to traditional data types, they are designed to support natively concurrent modifications. To this end, they embed in their specification a conflict-resolution mechanism.

To resolve conflicts in a deterministic manner, CRDTs usually attach identifiers to elements stored in the data structure. Identifiers have to comply with several constraints such as uniqueness or being densely ordered according to the kind of CRDT. These constraints may prevent the identifiers' size from being bounded. As the number of the updates increases, the size of identifiers grows. This leads to performance issues, since the efficiency of the replicated data structure decreases over time.

To address this issue, we propose a new CRDT for Sequence which embeds a renaming mechanism. It enables nodes to reassign shorter identifiers to elements in an uncoordinated manner. Obtained experiment results demonstrate that this mechanism decreases the overhead of the replicated data structure and eventually limits it.

To validate the proposed renaming mechanism, we performed an experimental evaluation to measure its performances on several aspects: (i) the size of the data structure ; (ii) the integration time of the rename operation ; (iii) the integration time of insert and remove operations. In cases (i) and (iii), we use LogootSplit as the baseline data structure to compare results. The results we obtained are very encouraging, as the integration time is far shorter with the renaming mechanism, even with the time spent to apply the rename operation.

7.4. Social Networks as Collaboration Support

Participants: Quentin Laporte Chabasse, Gérald Oster, François Charoy.

Safe peer to peer collaborative services requires a trusted peer to peer network in order to be effective. We started to investigate how to leverage social networks underlying inter organizational collaboration to support such collaboration. To reach this goal, we need to analyze collaborative graphs. They are a relevant sources of information to understand behavioural tendencies of groups of individuals. Exponential Random Graph Models (ERGMs) are commonly used to analyze such social processes including dependencies between members of the group. Our approach considers a modified version of ERGMs, modeling the problem as an edge labelling one. The main difficulty is inference since the normalizing constant involved in classical Markov Chain Monte Carlo approaches is not available in an analytic closed form.

The main contribution is to use the recent ABC Shadow algorithm [20]. This algorithm is built to sample from posterior distributions while avoiding the previously mentioned drawback. The proposed method is illustrated on real data sets provided by the HAL [0] platform and provides new insights on self-organized collaborations among researchers[11]

7.5. Secure Collaborative Editing

Participants: Mohammed Riyadh Abdmeziem, François Charoy.

Collaborative edition allows a group of entities to simultaneously edit and share the content of a document in real time. To provide the required keying materials, group key management protocols are usually considered in order to secure and encrypt the exchanged data. Indeed, existing fully distributed protocols induce significant overhead. Instead, centralized solutions are preferred for their high efficiency. Nevertheless, these centralized solutions present two main issues. The first issue is related to the broken end-to-end property, considering the central entity has access to the established credentials. The second issue is related to the single point of failure problem. In fact, if the central entity fails, the key establishment process fails too. To address these challenges, we proposed a simple, and yet efficient approach which enhances central-based protocols with both fault tolerance and end-to-end properties. To do so, we considered the group key as composed of two sub-keys. The first sub-key is only known to the members of the group, excluding the central entity, while the second sub-key is distributed and updated by the central entity following membership changes[3], [4]. Our initial assessment shows that the overall complexity of rekeying operations is not negatively impacted. In addition, our approach is backward compatible with existing solutions in the literature.

7.6. Trust and Data Sharing in Crisis Management

Participants: François Charoy, Béatrice Linot.

Sharing information between responders is important during crisis management response. Tools and platforms are eagerly developed for that purpose. They are supposed to support people and help them to build a shared situation awareness. However as the scale of crisis increases and as more and more organizations are involved, people get reluctant to use them to share their data. They prefer to rely on one to one communication tools like phones or text. This is why we are studying how these collaborative platforms impact the work of responders positively or negatively. We want to know why most of the time they don't want to use them for their original purpose. We studied reports on past incidents and conducted extensive analysis of the use of existing systems (e.g. the French platform CRISORSEC) through interviews, observation and data analysis. Early results show that participant have problems sharing written information for different kind of reason including its persistence, the time taken to produce the message and the lack of knowledge regarding who may access this information. This informs us on the requirement for future collaborative platforms.

[0] https://hal.inria.fr/

7.7. Identification and Selection of Services from Cloud Providers

Participants: Anis Ahmed Nacer, François Charoy, Olivier Perrin.

We continued our work on providing a framework to compare plans for services from cloud providers in order to help architects to select the best composition given the required criteria (both both functional and non-functional requirements) for an application. This year, we have made progress in two directions: the first is how to identify the key elements to be considered when architects want to compare the different plans, and the second one is a methodology to compute the best composition, given partial information provided in service description (based on the WOWA method).

In order to gather the key elements of the comparison that met the architects' requirements and the relationship between these key elements of the comparison, we reviewed the service providers' plans and previous works on benchmarks. Finally, to ensure that the list of key elements of the comparison and their relationship was complete for the service selection process, we conducted an empirical study with the architects.

Regarding the second part, we use the WOWA (Weighted Ordered Weighted Averaging) operator to solve this decision problem. This operator provides an aggregation function that uses both the simultaneous advantage of the OWA method to allow compensation between high and low values and the weighted average method to consider the importance of the suppliers who provide the information. WOWA uses two sets of weights: one corresponds to source significance, and the other corresponds to value significance.

Our evaluations are encouraging, and we are now ready to submit our proposals to conferences.

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

Participants: Amina Ahmed Nacer, Claude Godart, Guillaume Rosinosky, 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 users to identify the different pieces of 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, the paper [5] synthesizes our trust-aware deployment method.

7.9. Priority based events management in IoT-BPM architecture

Participants: Khalid Benali, Abir Ismaili-Alaoui.

BPM allows organizations to evolve their performance and achieve their goals, as it helps them to have a clear vision of their business. Several research works have been done in this area and aimed at improving business processes, by focusing on the optimization of business processes issues at build-time and at run-time, from different perspectives: control-flow perspective, data and event data perspective, and scheduling and event management perspective. Business process instances scheduling and event management are considered as a crucial step in the journey of business process improvement. However, this step becomes more challenging especially when the events are triggered by IoT devices. The main objective of our research consists on scheduling business process instances based on the priority of events that trigger these instances, taking into consideration historical data gathered from previous business process instances. We proposed a clustering approach based on the K-Means algorithm that we apply on a set of event sources, as to classify these sources on different clusters using a score calculated for each event source. This score is based on the frequency and the critically of previous events. The main objective of this approach was to create clusters of priorities. These clusters are used to estimate the criticality level of incoming events, and then the priority level of incoming process instances. However, there is always a degree of uncertainty regarding the criticality/priority level of events generated from sources that belong to the same cluster. This issue can be addressed by using fuzzy logic. In fact, the integration of a Fuzzy Inference System (FIS) in our IoT-BPM architecture, helps us to handle uncertainties regarding the criticality level of events, especially when these events are generated by sources that may have the same characteristics [8].

GAMBLE Project-Team

7. New Results

7.1. Non-Linear Computational Geometry

Participants: Laurent Dupont, Nuwan Herath Mudiyanselage, George Krait, Sylvain Lazard, Viviane Ledoux, Guillaume Moroz, Marc Pouget.

7.1.1. Clustering Complex Zeros of Triangular Systems of Polynomials

This work, presented at the CASC'19 Conference [23], gives the first algorithm for finding a set of natural ϵ-clusters of complex zeros of a regular triangular system of polynomials within a given polybox in $\mathbb{C}^n$, for any given $\epsilon > 0$. Our algorithm is based on a recent near-optimal algorithm of Becker et al (2016) for clustering the complex roots of a univariate polynomial where the coefficients are represented by number oracles. Our algorithm is based on recursive subdivision. It is local, numeric, certified and handles solutions with multiplicity. Our implementation is compared to well-known homotopy solvers on various triangular systems. Our solver always gives correct answers, is often faster than the homotopy solvers that often give correct answers, and sometimes faster than the ones that give sometimes correct results.

In collaboration with R. Imbach and C. Yap (Courant Institute of Mathematical Sciences, New York University, USA).

7.1.2. Numerical Algorithm for the Topology of Singular Plane Curves

We are interested in computing the topology of plane singular curves. For this, the singular points must be isolated. Numerical methods for isolating singular points are efficient but not certified in general. We are interested in developing certified numerical algorithms for isolating the singularities. In order to do so, we restrict our attention to the special case of plane curves that are projections of smooth curves in higher dimensions. In this setting, we show that the singularities can be encoded by a regular square system whose isolation can be certified by numerical methods. This type of curves appears naturally in robotics applications and scientific visualization. This work was presented at the EuroCG'19 Conference [24].

7.1.3. Reliable Computation of the Singularities of the Projection in $\mathbb{R}^3$ of a Generic Surface of $\mathbb{R}^4$

Computing efficiently the singularities of surfaces embedded in $\mathbb{R}^3$ is a difficult problem, and most state-of-the-art approaches only handle the case of surfaces defined by polynomial equations. Let F and G be C^∞ functions from $\mathbb{R}^4$ to $\mathbb{R}$ and $\mathcal{M} = \{(x, y, z, t) \in \mathbb{R}^4 \,|\, F(x, y, z, t) = G(x, y, z, t) = 0\}$ be the surface they define. Generically, the surface $\mathcal{M}$ is smooth and its projection Ω in $\mathbb{R}^3$ is singular. After describing the types of singularities that appear generically in Ω, we design a numerically well-posed system that encodes them. This can be used to return a set of boxes that enclose the singularities of Ω as tightly as required. As opposed to state-of-the art approaches, our approach is not restricted to polynomial mappings, and can handle trigonometric or exponential functions for example. This work was presented at the MACIS'19 Conference [19].

In collaboration with Sény Diatta (University Assane Seck of Ziguinchor, Senegal)

7.1.4. Evaluation of Chebyshev polynomials on intervals and application to root finding

In approximation theory, it is standard to approximate functions by polynomials expressed in the Chebyshev basis. Evaluating a polynomial f of degree n given in the Chebyshev basis can be done in $O(n)$ arithmetic operations using the Clenshaw algorithm. Unfortunately, the evaluation of f on an interval I using the Clenshaw algorithm with interval arithmetic returns an interval of width exponential in n. We describe a variant of the Clenshaw algorithm based on ball arithmetic that returns an interval of width quadratic in n for an interval of small enough width. As an application, our variant of the Clenshaw algorithm can be used to design an efficient root finding algorithm. This work was presented at the MACIS'19 Conference [21].

7.1.5. Using Maple to analyse parallel robots

We present the SIROPA Maple Library which has been designed to study serial and parallel manipulators at the conception level. We show how modern algorithms in Computer Algebra can be used to study the workspace, the joint space but also the existence of some physical capabilities w.r.t. to some design parameters left as degree of freedom for the designer of the robot. This work was presented at the Maple Conference 2019 [18].

In collaboration with Philippe Wenger, Damien Chablat (Laboratoire des Sciences du Numérique de Nantes, UMR CNRS 6004) and Fabrice Rouillier (project team OURAGAN *)*

7.2. Non-Euclidean Computational Geometry

Participants: Vincent Despré, Yan Garito, Elies Harington, Benedikt Kolbe, Georg Osang, Monique Teillaud, Gert Vegter.

7.2.1. Flipping Geometric Triangulations on Hyperbolic Surfaces

We consider geometric triangulations of surfaces, i.e., triangulations whose edges can be realized by disjoint locally geodesic segments. We prove that the flip graph of geometric triangulations with fixed vertices of a flat torus or a closed hyperbolic surface is connected. We give upper bounds on the number of edge flips that are necessary to transform any geometric triangulation on such a surface into a Delaunay triangulation [28].

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

7.2.2. Computing the Geometric Intersection Number of Curves

The geometric intersection number of a curve on a surface is the minimal number of self-intersections of any homotopic curve, i.e. of any curve obtained by continuous deformation. Given a curve c represented by a closed walk of length at most ℓ on a combinatorial surface of complexity n we describe simple algorithms to compute the geometric intersection number of c in $O(n + \ell^2)$ time, construct a curve homotopic to c that realizes this geometric intersection number in $O(n + \ell^4)$ time, decide if the geometric intersection number of c is zero, i.e. if c is homotopic to a simple curve, in $O(n + \ell \log(\ell))$ time [14].

In collaboration with Francis Lazarus (University of Grenoble).

7.3. Probabilistic Analysis of Geometric Data Structures and Algorithms

Participants: Olivier Devillers, Charles Duménil, Xavier Goaoc, Fernand Kuiebove Pefireko, Ji Won Park.

7.3.1. Expected Complexity of Routing in Θ6 and Half-Θ6 Graphs

We study online routing algorithms on the Θ6-graph and the half-Θ6-graph (which is equivalent to a variant of the Delaunay triangulation). Given a source vertex s and a target vertex t in the Θ6-graph (resp. half-Θ6-graph), there exists a deterministic online routing algorithm that finds a path from s to t whose length is at most 2 st (resp. 2.89 st) which is optimal in the worst case [Bose et al., SIAM J. on Computing, 44(6)]. We propose alternative, slightly simpler routing algorithms that are optimal in the worst case and for which we provide an analysis of the average routing ratio for the Θ6-graph and half-Θ6-graph defined on a Poisson point process. For the Θ6-graph, our online routing algorithm has an expected routing ratio of 1.161 (when s and t random) and a maximum expected routing ratio of 1.22 (maximum for fixed s and t where all other points are random), much better than the worst-case routing ratio of 2. For the half-Θ6-graph, our memoryless online routing algorithm has an expected routing ratio of 1.43 and a maximum expected routing ratio of 1.58. Our online routing algorithm that uses a constant amount of additional memory has an expected routing ratio of 1.34 and a maximum expected routing ratio of 1.40. The additional memory is only used to remember the coordinates of the starting point of the route. Both of these algorithms have an expected routing ratio that is much better than their worst-case routing ratio of 2.89 [27].

In collaboration with Prosenjit Bose (University Carleton) and JeanLou De Carufel (University of Ottawa)

7.3.2. A Poisson sample of a smooth surface is a good sample

The complexity of the 3D-Delaunay triangulation (tetrahedralization) of n points distributed on a surface ranges from linear to quadratic. When the points are a deterministic good sample of a smooth compact generic surface, the size of the Delaunay triangulation is $O(n \log n)$. Using this result, we prove that when points are Poisson distributed on a surface under the same hypothesis, whose expected number of vertices is λ, the expected size is $O(\lambda \log_2 \lambda)$ [22].

7.3.3. On Order Types of Random Point Sets

Let P be a set of n random points chosen uniformly in the unit square. We examine the typical resolution of the order type of P. First, we show that with high probability, P can be rounded to the grid of step $\frac{1}{n^{3+\epsilon}}$ without changing its order type. Second, we study algorithms for determining the order type of a point set in terms of the number of coordinate bits they require to know. We give an algorithm that requires on average $4n \log_2 n + O(n)$ bits to determine the order type of P, and show that any algorithm requires at least $4n \log_2 n - O(n \log \log n)$ bits. Both results extend to more general models of random point sets [29].

In collaboration with Philippe Duchon (Université de Bordeaux) and Marc Glisse (project team DATASHAPE *).*

7.3.4. Randomized incremental construction of Delaunay triangulations of nice point sets

Randomized incremental construction (RIC) is one of the most important paradigms for building geometric data structures. Clarkson and Shor developed a general theory that led to numerous algorithms that are both simple and efficient in theory and in practice. Randomized incremental constructions are most of the time space and time optimal in the worst-case, as exemplified by the construction of convex hulls, Delaunay triangulations and arrangements of line segments. However, the worst-case scenario occurs rarely in practice and we would like to understand how RIC behaves when the input is nice in the sense that the associated output is significantly smaller than in the worst-case. For example, it is known that the Delaunay triangulations of nicely distributed points on polyhedral surfaces in $\mathbb{E}^3$ has linear complexity, as opposed to a worst-case quadratic complexity. The standard analysis does not provide accurate bounds on the complexity of such cases and we aim at establishing such bounds. More precisely, we will show that, in the case of nicely distributed points on polyhedral surfaces, the complexity of the usual RIC is $O(n \log n)$ which is optimal. In other words, without any modification, RIC nicely adapts to good cases of practical value. Our proofs also work for some other notions of nicely distributed point sets, such as (ϵ, κ)-samples. Along the way, we prove a probabilistic lemma for sampling without replacement, which may be of independent interest [16], [26].

In collaboration with Jean-Daniel Boissonnat, Kunal Dutta and Marc Glisse (project team DATASHAPE *).*

7.3.5. Random polytopes and the wet part for arbitrary probability distributions

We examine how the measure and the number of vertices of the convex hull of a random sample of n points from an arbitrary probability measure in $\mathbb{R}^d$ relates to the wet part of that measure. This extends classical results for the uniform distribution from a convex set [Bárány and Larman 1988]. The lower bound of Bárány and Larman continues to hold in the general setting, but the upper bound must be relaxed by a factor of $\log n$. We show by an example that this is tight [25].

In collaboration with Imre Barany (Rényi Institute of Mathematics) Matthieu Fradelizi (Laboratoire d'Analyse et de Mathématiques Appliquées) Alfredo Hubard (Laboratoire d'Informatique Gaspard-Monge) Günter Rote (Institut für Informatik, Berlin)

7.4. Discrete Geometric structures

Participants: Xavier Goaoc, Galatée Hemery Vaglica.

7.4.1. Shatter functions with polynomial growth rates

We study how a single value of the shatter function of a set system restricts its asymptotic growth. Along the way, we refute a conjecture of Bondy and Hajnal which generalizes Sauer's Lemma. [12]

7.4.2. The discrete yet ubiquitous theorems of Caratheodory, Helly, Sperner, Tucker, and Tverberg

We discuss five discrete results: the lemmas of Sperner and Tucker from combinatorial topology and the theorems of Carathéodory, Helly, and Tverberg from combinatorial geometry. We explore their connections and emphasize their broad impact in application areas such as game theory, graph theory, mathematical optimization, computational geometry, etc. [13]

7.4.3. Shellability is NP-complete

We prove that for every $d \geq 2$, deciding if a pure, d-dimensional, simplicial complex is shellable is NP-hard, hence NP-complete. This resolves a question raised, e.g., by Danaraj and Klee in 1978. Our reduction also yields that for every $d \geq 2$ and $k \geq 0$, deciding if a pure, d-dimensional, simplicial complex is k-decomposable is NP-hard. For $d \geq 3$, both problems remain NP-hard when restricted to contractible pure d-dimensional complexes. Another simple corollary of our result is that it is NP-hard to decide whether a given poset is CL-shellable. [15]

7.4.4. An Experimental Study of Forbidden Patterns in Geometric Permutations by Combinatorial Lifting

We study the problem of deciding if a given triple of permutations can be realized as geometric permutations of disjoint convex sets in $\mathbb{R}^3$. We show that this question, which is equivalent to deciding the emptiness of certain semi-algebraic sets bounded by cubic polynomials, can be "lifted" to a purely combinatorial problem. We propose an effective algorithm for that problem, and use it to gain new insights into the structure of geometric permutations. [20]

7.5. Classical Computational Geometry

Participants: Olivier Devillers, Sylvain Lazard, Leo Valque.

7.5.1. Rounding Meshes

Let $\mathcal{P}$ be a set of n polygons in $\mathbb{R}^3$, each of constant complexity and with pairwise disjoint interiors. We previously proposed [5] a rounding algorithm that maps $\mathcal{P}$ to a simplicial complex $\mathcal{Q}$ whose vertices have integer coordinates such that 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 built 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. We developed [30] the first implementation of this algorithm, which is also the first implementation for rounding a mesh on a grid (whose fineness is independent of the input size) while preserving reasonable geometric and topological properties. We also provided some insight that this algorithm and implementation have practical average complexity in $O(n\sqrt{n})$ on “real data", which has to be compared to its $O(n^{15})$ worst-case time complexity. Our implementation is still too slow to be used in practice but it provides a good proof of concept.

7.5.2. Hardness results on Voronoi, Laguerre and Apollonius diagrams

We show that converting Apollonius and Laguerre diagrams from an already built Voronoi diagram of a set of n points in 2D requires at least $\Omega(n \log n)$ computation time. We also show that converting an Apollonius diagram of a set of n weighted points in 2D from a Laguerre diagram and vice-versa requires at least $\Omega(n \log n)$ computation time as well. Furthermore , we present a very simple randomized incremental construction algorithm that takes expected $O(n \log n)$ computation time to build an Apollonius diagram of non-overlapping circles in 2D [17].

In collaboration with Kevin Buchin (TU Eindhoven), Pedro de Castro (University Pernanbuco), and Menelaos Karavelas (University Heraklion).

LARSEN Project-Team

7. New Results

7.1. Lifelong autonomy

7.1.1. Motion planning for robot audition

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

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

Robot audition refers to a range of hearing capabilities which help robots explore and understand their environment. Among them, sound source localization is the problem of estimating the location of a sound source given measurements of its angle of arrival with respect to a microphone array mounted on the robot. In addition, robot motion can help quickly solve the front-back ambiguity existing in a linear microphone array. In this work, we focus on the problem of exploiting robot motion to improve the estimation of the location of an intermittent and possibly moving source in a noisy and reverberant environment. We first propose a robust extended mixture Kalman filtering framework for jointly estimating the source location and its activity over time. Building on this framework, we then propose a long-term robot motion planning algorithm based on Monte Carlo tree search to find an optimal robot trajectory according to two alternative criteria: the Shannon entropy or the standard deviation of the estimated belief on the source location. Experimental results show the robustness of the proposed estimation framework to false angle of arrival measurements within $\pm 20°$ and 10% false source activity detection rate. The proposed robot motion planning technique achieves an average localization error 48.7% smaller than a one-step-ahead method.

Publication: [10]

7.1.2. Addressing Active Sensing Problems through Monte-Carlo Tree Search (MCTS)

Participants: Vincent Thomas, Gabriel Belouze, Sylvain Geiser, Olivier Buffet.

The problem of active sensing is of paramount interest for building self awareness in robotic systems. It consists in planning actions in a view to gather information (*e.g.*, measured through the entropy over certain state variables) in an optimal way. In the past, we have proposed an original formalism, ρ-POMDPs, and new algorithms for representing and solving such active sensing problems [24] by using point-based algorithms, assuming either convex or Lipschitz-continuous criteria. More recently, we have developed new approaches based on Monte-Carlo Tree Search (MCTS), and in particular Partially Observable Monte-Carlo Planning (POMCP), which provably converge only assuming the continuity of the criterion. We are now going towards algorithms more suitable to certain robotic tasks by allowing for continuous state and observation spaces.

Publication: [20]

7.1.3. Heuristic Search for (Partially Observable) Stochastic Games

Participants: Olivier Buffet, Vincent Thomas.

Collaboration with Jilles Dibangoye (INSA-Lyon, Inria team CHROMA) and Abdallah Saffidine (University of New South Wales (UNSW), Sydney, Australia).

Many robotic scenarios involve multiple interacting agents, robots or humans, *e.g.*, security robots in public areas. We have mainly worked in the past on the collaborative setting, all agents sharing one objective, in particular through solving Dec-POMDPs by (i) turning them into occupancy MDPs and (ii) using heuristic search techniques and value function approximation [2]. A key idea is to take the point of view of a central planner and reason on a sufficient statistic called *occupancy state*. We are now working on applying similar approaches in the important 2-player zero-sum setting, *i.e.*, with two competing agents. As a preliminary step, we have proposed and evaluated an algorithm for (fully observable) stochastic games, which does not require any problem transformation. Then we have proposed an algorithm for partially observable stochastic games, here turning the problem into an occupancy Markov game.

[This line of research will be pursued through Jilles Dibangoye's ANR JCJC PLASMA.]

7.1.4. Interpretable Action Policies

Participant: Olivier Buffet.

Collaboration with Iadine Chadès and Jonathan Ferrer Mestres (CSIRO, Brisbane, Australia), and Thomas G. Dietterich (Oregon State University, USA).

Computer-aided task planning requires providing user-friendly plans, in particular, plans that make sense to the user. In probabilistic planning (in the MDP formalism), such interpretable plans can be derived by constraining action policies (if X happens, do Y) to depend on a reduced subset of (abstract) states or state variables. We have (i) formalized the problem of finding a set of at most K abstract states (forming a partition of the original state space) such that any optimal policy of the induced abstract MDP is as close as possible to optimal policies of the original MDP, and (ii) proposed 3 solution algorithms with theoretical and empirical evaluations.

7.1.5. Perspective: hierarchical quality diversity, from materials to machines

Participant: Jean-Baptiste Mouret.

Collaboration with CSIRO (Australia) and Vrije Universiteit Amsterdam (Netherlands).

Natural lifeforms specialize to their environmental niches across many levels, from low-level features such as DNA and proteins, through to higher-level artefacts including eyes, limbs and overarching body plans. We propose 'multi-level evolution', a bottom-up automatic process that designs robots across multiple levels and niches them to tasks and environmental conditions. Multi-level evolution concurrently explores constituent molecular and material building blocks, as well as their possible assemblies into specialized morphological and sensorimotor configurations. Multi-level evolution provides a route to fully harness a recent explosion in available candidate materials and ongoing advances in rapid manufacturing processes. We outline a feasible architecture that realizes this vision, highlight the main roadblocks and how they may be overcome, and show robotic applications to which multi-level evolution is particularly suited. By forming a research agenda to stimulate discussion between researchers in related fields, we hope to inspire the pursuit of multi-level robotic design all the way from material to machine.

Publication: [5]

7.1.6. Improving Embodied Evolutionary Robotics

Participant: Amine Boumaza.

Multi-robots learning is a hard still unsolved problem. When framed into the machine learning theoretical setting, it suffers from a high complexity when seeking optimal solutions. On the other hand, when sub-optimal solutions are acceptable Embodied Evolutionary Robotics, can provide solutions that perform well in practice. Improving these algorithms in terms of run-time or solution quality is an important research question.

It has been long known from the theoretical work on evolution strategies, that recombination improves convergence towards better solution and improves robustness against selection error in noisy environment. We propose to investigate the effect of recombination in online embodied evolutionary robotics, where evolution is decentralized on a swarm of agents. We hypothesize that these properties can also be observed in these algorithms and thus could improve their performance. We introduce the $(\mu/\mu, 1)$-On-line Embedded Evolutionary Algorithm (EEA) which uses a recombination operator inspired from evolution strategies and apply it to learn three different collective robotics tasks, locomotion, item collection and item foraging. Different recombination operators are investigated and compared against a purely mutative version of the algorithm. The experiments show that, when correctly designed, recombination improves significantly the adaptation of the swarm in all scenarios.

Publication: [13] [12]

7.1.7. Multi-robot exploration of an unknown environment

Participants: Nicolas Gauville, François Charpillet.

Different approaches exist for multi-robot autonomous exploration. These include frontier approaches, where robots are assigned to unexplored areas of the map, which provide good performance but require sharing the map and centralizing decision-making. The Brick and Mortar approaches, on the other hand, use a ground marking with local decision-making, but give much lower performance. The algorithm developped by Nicolas Gauville during his pre-thesis period is a trade-off between these two approaches, allowing local decision-making and, surprisingly, performances are closed to centralized frontier approaches. We also propose a comparative study of the performance of the three different approaches : *Brick & Mortar*, *Global Frontiers* and *Local Frontiers*. Our local algorithm is also complete for the exploration problem and can be easily distributed on robots with a minor loss of performance. This work follows the *Cart-O-Matic* project in which our team participated, which aimed to explore and map a building while recognizing specific objects inside with a team of 5 mobile robots.

Publication: [16]

7.2. Natural Interaction with Robotics Systems

Thanks to the arrival of Pauline Maurice and the AnDy H2020 project, our activities about interaction are currently focused on ergnonomic interaction, which requires good foundations in motion analysis.

7.2.1. Digital human modeling for collaborative robotics

Participant: Pauline Maurice.

Collaboration with Vincent Padois (Inria Bordeaux and Sorbonne Université), Yvan Measson (CEA-LIST) and Philippe Bidaud (ONERA and Sorbonne Université).

Work-related musculoskeletal disorders in industry represent a major and growing health problem in many developed countries. Collaborative robotics, which allows the joint manipulation of objects by both a robot and a person, is a possible solution provided that it is possible to assess the ergonomic benefit they offer. Using a digital human model (DHM) can cut down the development cost and time by replacing the physical mock-up by a virtual one easier to modify. The first part of this work details the challenges of digital ergonomic assessment for collaborative robotics. State-of-the-art work on DHM simulations with collaborative robots is reviewed to identify which questions currently remain open. The second part of this work focuses on a specific use case and presents a DHM-based method to optimize design parameters of a collaborative robot for an industrial task.

Publication: [21]

7.2.2. Probabilistic decision making for collaborative robotics

Participants: Yang You, Vincent Thomas, Olivier Buffet, François Charpillet, Francis Colas.

Collaboration with Rachid Alami (LAAS, France).

This work is part of the ANR Flying Co-Worker project and focuses on high-level decision making for collaborative robotics. When a robot has to assist a human worker, it does not have direct access to his current intention or his preferences but has to adapt its behaviour to help the human completing his task. To achieve this, we followed what has been proposed by [31] to model a situation of interaction as a Partially Observable Markov Decision Process (POMDP) by assuming that (i) the robot and the human act sequentially, one after another, and that (ii) the human is rational and makes his decision without considering the future robot's action.

7.2.3. Ativity recognition and prediction

Participants: François Charpillet, Francis Colas, Serena Ivaldi, Niyati Rawal, Vincent Thomas.

This work is part of the ANR Flying Co-Worker project and focuses on activity recognition and long-term prediction for collaborative robotics. Recognizing and predicting human activities is fundamental for a robot to help a human. Previous work in the team on activity recognition [6] rely on Hidden Markov Models (HMM) with, in particular, the Markov assumption stating that the distribution on the next state is independent from former states given the current state. This assumption, at the heart of the recurrent expression of the inference in HMM, has the unfortunate consequence to constrain the a priori distribution on the duration in each state to exponential distributions. However, it can be observed in datasets that this is not the case for many activities, which have a typical duration. This discrepancy is negligeable for recognition where HMM models achieve good performance thanks to the observations, but prevents longer-term activity prediction.

In the master project of Niyati Rawal, we investigated a slightly different model, Explicit Duration Hidden Markov Model (EDHMM), in which the duration of the activity can be modeled more finely. Preliminary results show that the recognition performance was similar to HMM but with a better prediction performance.

7.2.4. Humanoid Whole-Body Movement Optimization from Retargeted Human Motions

Participants: Waldez Azevedo Gomes Junior, Vishnu Radhakrishnan, Luigi Penco, Valerio Modugno, Jean-Baptiste Mouret, Serena Ivaldi.

Motion retargeting and teleoperation are powerful tools to demonstrate complex whole-body movements to humanoid robots: in a sense, they are the equivalent of kinesthetic teaching for manipulators. However, retargeted motions may not be optimal for the robot: because of different kinematics and dynamics, there could be other robot trajectories that perform the same task more efficiently, for example with less power consumption. We propose to use the retargeted trajectories to bootstrap a learning process aimed at optimizing the whole-body trajectories w.r.t. a specified cost function. To ensure that the optimized motions are safe, i.e., they do not violate system constraints, we used constrained optimization algorithms. We compared both global and local optimization approaches, since the optimized robot solution may not be close to the demonstrated one. We evaluated our framework with the humanoid robot iCub on an object lifting scenario, initially demonstrated by a human operator wearing a motion-tracking suit. By optimizing the initial retargeted movements, we can improve robot performance by over 40%.

Publication: [14]

7.2.5. Tele-operation of Humanoids

Participants: Luigi Penco, Waldez Gomes, Valerio Modugno, Serena Ivaldi.

We envision a world where robots can act as physical avatars and effectively replace humans in hazardous scenarios by means of teleoperation, which we see as a particular way of interacting with a robot. However, teleoperating humanoids is a challenging task because of differences in kinematics (e.g., structure and joint limits) and dynamics (e.g., mass distribution, inertia) are still significant. Another crucial issue is ensuring the dynamic balance of the robot while trying to imitate the human motion. We propose a multi-mode teleoperation framework for controlling humanoid robots for loco-manipulation tasks that address the aforementioned challenges by using two levels of teleoperation: a low-level for manipulation, realized via whole-body teleoperation, and a high-level for locomotion, based on the generation of reference velocities that are then tracked by the humanoid. We believe that this combination of different modes of teleoperation will considerably ease the burden of controlling humanoids, ultimately increasing their adaptability to complex situations which cannot be handled satisfactorily by fully autonomous systems.

Publication: [11]

7.2.6. Activity Recognition for Ergonomics Assessment of Industrial Tasks with Automatic Feature Selection

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

In industry, ergonomic assessment is currently performed manually based on the identification of postures and actions by experts. We aim at proposing a system for automatic ergonomic assessment based on activity recognition. In this work, we define a taxonomy of activities, composed of four levels, compatible with items evaluated in standard ergonomic worksheets. The proposed taxonomy is applied to learn activity recognition models based on Hidden Markov Models. We also identify dedicated sets of features to be used as input of the recognition models so as to maximize the recognition performance for each level of our taxonomy. We compare three feature selection methods to obtain these subsets. Data from 13 participants performing a series of tasks mimicking industrial tasks are collected to train and test the recognition module. Results show that the selected subsets allow us to successfully infer ergonomically relevant postures and actions.

Publication: [6]

7.2.7. Human movement and ergonomics: An industry-oriented dataset for collaborative robotics

Participants: Pauline Maurice, Adrien Malaisé, Serena Ivaldi.

With the participation of Clélie Amiot, Nicolas Paris and Guy-Junior Richard, interns from Université de Lorraine during the summer 2018.

Improving work conditions in industry is a major challenge that can be addressed with new emerging technologies such as collaborative robots. Machine learning techniques can improve the performance of those robots, by endowing them with a degree of awareness of the human state and ergonomics condition. The availability of appropriate datasets to learn models and test prediction and control algorithms, however, remains an issue. This work presents a dataset of human motions in industry-like activities, fully labeled according to the ergonomics assessment worksheet EAWS, widely used in industries such as car manufacturing. Thirteen participants performed several series of activities, such as screwing and manipulating loads under different conditions, resulting in more than 5 hours of data. The dataset contains the participants' whole-body kinematics recorded both with wearable inertial sensors and marker-based optical motion capture, finger pressure force, video recordings, and annotations by three independent annotators of the performed action and the adopted posture following the EAWS postural grid. Sensor data are available in different formats to facilitate their reuse. The dataset is intended for use by researchers developing algorithms for classifying, predicting, or evaluating human motion in industrial settings, as well as researchers developing collaborative robotics solutions that aim at improving the workers' ergonomics. The annotation of the whole dataset following an ergonomics standard makes it valuable for ergonomics-related applications, but we expect its use to be broader in the robotics, machine learning, and human movement communities.

Publication: [8]

7.2.8. Objective and Subjective Effects of a Passive Exoskeleton on Overhead Work

Participants: Pauline Maurice, Serena Ivaldi.

Collaboration with Jernej Čamernik, Daša Gorjan and Jan Babič (Jozef Stefan Institute, Ljubljana, Slovenia), with Benjamin Schirrmeister and Jonas Bornmann (Otto Bock SE & Co. KGaA, Duderstadt, Germany), with Luca Tagliapietra, Claudia Latella and Daniele Pucci (Istituto Italiano di Tecnologia, Genova, Italy), and with Lars Fritzsche (IMK Automotive, Chemitz, Germany).

Overhead work is a frequent cause of shoulder work-related musculoskeletal disorders. Exoskeletons offering arm support have the potential to reduce shoulder strain, without requiring large scale reorganization of the workspace. Assessment of such systems however requires to take multiple factors into consideration. This work presents a thorough in-lab assessment of PAEXO, a novel passive exoskeleton for arm support during overhead work. A list of evaluation criteria and associated performance metrics is proposed to cover both objective and subjective effects of the exoskeleton, on the user and on the task being performed. These metrics are measured during a lab study, where 12 participants perform an overhead pointing task with and without the exoskeleton, while their physical, physiological and psychological states are monitored. Results show that using PAEXO reduces shoulder physical strain as well as global physiological strain, without increasing low back strain nor degrading balance. These positive effects are achieved without degrading task performance.

Importantly, participant' opinions of PAEXO are positive, in agreement with the objective measures. Thus, PAEXO seems a promising solution to help prevent shoulder injuries and diseases among overhead workers, without negatively impacting productivity.

Publication: [7], [19]

7.2.9. Assessing and improving human movements using sensitivity analysis and digital human simulation

Participant: Pauline Maurice.

Collaboration with Vincent Padois (Inria Bordeaux and Sorbonne Université), Yvan Measson (CEA-LIST) and Philippe Bidaud (ONERA and Sorbonne Université).

Enhancing the performance of technical movements aims both at improving operational results and at reducing biomechanical demands. Advances in human biomechanics and modeling tools allow to evaluate human performance with more and more details. Finding the right modifications to improve the performance is, however, still addressed with extensive time consuming trial-and-error processes. This work presents a framework for easily assessing human movements and automatically providing recommendations to improve their performances. An optimization-based whole-body controller is used to dynamically replay human movements from motion capture data, to evaluate existing movements. Automatic digital human simulations are then run to estimate performance indicators when the movement is performed in many different ways. Sensitivity indices are thereby computed to quantify the influence of postural parameters on the performance. Based on the results of the sensitivity analysis, recommendations for posture improvement are provided. The method is successfully validated on a drilling activity.

Publication: [9]

7.2.10. Human Motion analysis for assistance

Participants: François Charpillet, Jessica Colombel.

Collaboration with David Daney (Inria Bordeaux, Auctus Team)

Different sort of sensors can be used for rehabilitation at home. This year we have evaluated the usabily of a Kinect 2. The proposed approach is to improve joint angle estimates. It is based on a constrained extended Kalman Filter that tracks inputted measured joint centers. Since the proposed approach uses a biomechanical model, it allows to obtain physically consistent constrained joint angles and constant segment lengths. A practical method, that is not sensor specific, for the optimal tuning of the extended Kalman filter covariance matrices is provided. It uses reference data obtained from a stereophotogrammetric system but it has to be tuned only once since it is task specific only. The improvement of optimal tuning over classical methods for setting the covariance matrices is shown with a statistical parametric mapping analysis. The proposed approach was tested with six healthy subjects performing 4 rehabilitation tasks. Joint estimates accuracy was assessed with a reference stereophotogrammetric system. Even if some joints such as the internal/external rotations were not well estimated, the proposed optimized algorithm reached a satisfactory average root mean square difference of 9.7deg and a correlation coefficient 0.86 of for all joints. Our results show that affordable RGB-D sensor can be used for simple in-home rehabilitation when using a constrained biomechanical model.

A work carried out this year, takes the search for a sensor for personal assistance a step further with the study of the new Kinect Azure. Human-robot interaction requires a robust estimate of human motion in real-time. This work presents a fusion algorithm for joint center positions tracking from multiple depth cameras to improve human motion analysis accuracy. The proposed algorithm is based on body tracking measurements fusion with an extended Kalman filter and anthropomorphic constraints. However, the effectiveness and robustness of such algorithm depends on the A direct comparison of joint center positions estimated with a reference stereophotogrammetric system and the ones estimated with the new Kinect 3 (Azure Kinect) sensor and its older version the Kinect 2 (Kinect for Windows) has been made. The proposed approach improves body tracker data even for Kinect 3 which has not the same characteristics than Kinect 2. This study shows also the importance of defining good heuristics to merge data depending on how the body tracking works. Thus, with

proper heuristics, the joint center position estimates are improved by at least 14.6 %. Finally, we propose an additional comparison between Kinect 2 and Kinect 3 exhibiting the pros and cons of the two sensors. This study is now in submission for an international conference.

Finally, a state of the art on biological motion was realized. The purpose of this study is to understand and develop methods for decomposing motion. The EWalk dataset (http://gamma.cs.unc.edu/GAIT/#EWalk) will allow us to test emotion recognition from simple decompositions and classifiers. Then, we will extend the methods to other cognitive parameters.

7.2.11. Reliable localization of pedestrians in a smart home using multi-sensor data fusion

Participants: François Charpillet, Lina Achaji.

Collaboration with Maan Badaoui EL Najjar(Cristal Laboratory Lille, DiCOT Team), Mohamad Daher (the Lebanese University Faculty of technology, Tripoli)

One objective of the Larsen team is to develop technologies allowing older people to live independently as long as possible in their own homes instead of in specialized institutions. However, elderly people face physical problems that reduce their autonomy, and consequently their capacity to achieve daily activities. The integration of environmental or body sensors in what is called nowadays smart habitats is a solution that is appealing to provide a better quality of life with safer conditions. Localization and tracking of people in indoor environments are one of the primary services to be developed to follow them up at home, permitting to evaluate their physical states through the observation of their Activities of Daily Living (ADL). We proposed during the internship of Lina Achaji to localize and track the center of pressure (CoP) of people (one or two) in a smart home using a load sensing floor equipped with around 400 load sensors as well as wearable sensors. The data fusion is made using an informational filter where an inverted pendulum bio-mechanical model is introduced. The obtained results are very promising and were validated using a motion tracking system and force plates.

Publication: [4]

7.2.12. Ambient assisting living

Participants: François Charpillet, Yassine El Khadiri.

Collaboration with Cedric Rose from Diatelic compagny.

The ageing of the population confronts modern societies with an unprecedented demographic transformation. These include the imbalance in our pension systems and the cost of caring for the elderly. On this last point, apart from the economic aspects, the placement of elderly people is often only a choice of reason and can be quite badly experienced by people. One response to this societal problem is the development of technologies that make it easier to keep elderly people at home. The state of the art in this field abounds with upstream projects that are moving in this direction. Many of them are seeking to develop home monitoring systems. Their objectives are to detect and even prevent the occurrence of worrying or critical situations and to assess the physical condition or even fragility of the people being monitored. It is within this framework that this contribution is made. In this work, we have focused on the particular problem of monitoring the quality of sleep as well as the detection of nocturnal waking of a person living alone at home. The home is equipped with simple ambient sensors such as binary motion detectors. We have developed a Bayesian inference method that allows our solution to be flexible and robust enough for different types of installations and apartment configurations while maintaining a prediction accuracy of 0.94. This solution is currently being deployed on several dozen apartments in Lorraine by Diatelic and Pharmagest compagnies.

Publication: [15]

MAGRIT Team

7. New Results

7.1. Matching and localization

Participants: Marie-Odile Berger, Vincent Gaudilliere, Gilles Simon, Frédéric Sur, Matthieu Zins.

7.1.1. View synthesis for efficient and accurate pose computation

Estimating the pose of a camera from a scene model 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. 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-centred scenes, or need an initial pose guess. A new method based on viewpoint simulation is presented in [15]. In this article, we show that view synthesis dramatically improves pose computation and that both the synthesis process and pose computation can be done in a very efficient way. Two major problems are especially addressed: the positioning of the virtual viewpoints with respect to the scene, and the synthesis of geometrically consistent patches. Experiments show that patch synthesis dramatically improves the accuracy of the pose in case of difficult registration, with a limited computational cost.

7.1.2. Localization from objects

We are interested in AR applications which take place in man-made GPS-denied environments, such as industrial or indoor scenes. In such environments, relocalization may fail due to repeated patterns and large changes in appearance which occur even for small changes in viewpoint. During this year, we have investigated a new method for relocalization which operates at the level of objects and takes advantage of the impressive progress realized in object detection. Recent works have opened the way towards object oriented reconstruction from elliptic approximation of objects detected in images. We have gone beyond that and have proposed a new method for pose computation based on ellipse/ellipsoid correspondences. In [18], we have proved that a closed form estimate of the translation can be uniquely inferred from the rotation matrix of the pose. When two or more correspondences are available, the rotation matrix is deduced through an optimization problem with three degrees of freedom. However, the pose cannot be uniquely computed from one correspondence. In [19], we consider the practical common case where an initial guess of the rotation matrix of the pose is known, for instance with an inertial sensor or from the estimation of orthogonal vanishing points [10]. The translation is recovered as in [18], [24]. We proved the effectiveness of the method on real scenes from a set of object detections generated by YOLO [33]. Globally, considering pose at the level of objects allows us to avoid common failures due to repeated structures. In addition, due to the small combinatorics induced by object correspondences, our method is well suited to fast rough localization even in large environments.

A patent was filed on this method in May 2019 [27]. An Inria technological transfer action (ATT) on the subject of object based localization will start in January 2020 with the aim to produce a demonstrator for industrial maintenance in complex environments.

7.2. Handling non-rigid deformations

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

7.2.1. Compliance-based non rigid registration

Within J. Guevara's PhD thesis, we are investigating non rigid registration methods which exploit the matching of the vascular trees and are able to cope with large deformations of the organ. This year, we have developed a matching method which is entirely based on the mechanical properties of the organ. We thus avoid tedious parameter tuning which is required by many methods and instead use parameters whose values are known or can be measured. Our method makes use of an advanced biomechanical model which handles heterogeneities and anisotropy due to vasculature. The main originality of the method lies in the definition of a better and novel metric for generating improved graph-matching hypotheses, based on the notion of compliance, the inverse of stiffness. This method reduces the computation time by predicting first the most plausible matching hypotheses on a mechanical basis and reduces the sensitivity on the search space parameters. These contributions improve the registration quality and meet intra-operative timing constraints. Experiments have been conducted on ten realistic synthetic datasets and two real porcine datasets which where automatically segmented. This work was recently accepted in the journal *Annals of Biomedical Engineering* [9], [11].

7.2.2. Individual-specific heart valve modeling

Recent works on computer-based models of mitral valve behavior rely on manual extraction of the complex valve geometry, which is tedious and requires a high level of expertise. On the contrary, in the context of D. Panicheva's PhD thesis, we are investigating methods to segment the chordae with little human supervision which produce mechanically-coherent simulations of the mitral valve.

Valve chordae are generalized cylinders: Instead of being limited to a line, the central axis is a continuous curve; instead of a constant radius, the radius varies along the axis. Most of the time, chordae sections are flattened ellipses and classical model-based methods commonly used for vessel enhancement or vessel segmentation fail. We have exploited the fact that there are no other generalized cylinders than the chordae in the CT scan and we have proposed a topology-based method for chordae extraction. This approach is flexible and only requires the knowledge of an upper bound of the maximum radius of the chordae. The method has been tested on three CT scans. Overall, non-chordae structures are correctly identified and detected chordae ending points match up with actual chordae attachment points [21].

We then worked on evaluating the effectiveness of our approach. The valve behavior was simulated with a biomechanical framework based on the Finite Element Method. A structural model with no fluid-structure interaction was used. Physiological behavior was simulated by mechanical forces such as blood pressure, contact forces and tension forces applied from chordae tensions. The chordae segmentation was validated by comparing the simulation results to those obtained with manually segmented chordae [22].

7.2.3. Image-based biomechanical simulation of the diaphragm during mechanical ventilation

When intensive care patients are subjected to mechanical ventilation, the ventilator causes damage to the muscles that govern the normal breathing, leading to Ventilator Induced Diaphragmatic Dysfunction (VIDD). The INVIVE project aims to study the mechanics of respiration through numerical simulation in order to learn more about the onset of VIDD. We have worked during this year on how to compute solutions of the static linear elasticity equation using last year's work on the diaphragm geometry [26]. Since obtaining an analytical formulation of the boundary conditions in 3D is complex, we have worked on adapting our method to implicit geometries built from 2D data of the diaphragm. The idea is to have an analytical formulation of both the geometry and the boundary conditions to validate our radial basis framework. It is based on points belonging to a cross-section that has been chosen in the middle of the diaphragm. Points are gathered in groups inside rectangles based on a K-means classification. Rectangle dimensions are set so as to ensure cross-coverage. Curve patches are then computed for each rectangle using radial basis functions. A list of local curves is obtained from both the thoracic and abdomen zones and by combining them it is possible to evaluate the global implicit curve of the diaphragm.

7.2.4. 3D catheter navigation from monocular images

In interventional radiology, the 3D shape of the micro-tool (guidewire, micro-catheter or micro-coil) can be very difficult, if not impossible to infer from fluoroscopy images. We consider this question as a single view

3D curve reconstruction problem. Our aim is to assess whether, and under which conditions, a sophisticated physics-based model can be effective to compensate for the incomplete data in this ill-posed problem.

Raffaella Trivisonne started her PhD thesis in November 2015 (co-supervised by Stéphane Cotin, from MIMESIS team in Strasbourg) to address this research topic. An unscented Kalman filter is used as a fusion mechanism, in a non-rigid shape-from-motion approach: the observations are image data (opaque markers placed along the device), and the model is implemented through interactive physics-based simulation. Our contribution is to handle contacts, which introduce discontinuities in the first and second order derivatives of motion (resp. velocity and forces). Extensive validation on both synthetic and phantom-based data has been carried out this year [30], and various state vector parametrizations have been investigated, in particular in a view to achieve data assimilation of mechanical parameters to improve the predictability of simulation.

In this context, validation is made very complex by the need to acquire ground truth 3D curve shapes that are subjected to contacts and demonstrate highly transient dynamic deformations (e.g. stick and slip transitions after contact). Thomas Mangin was hired on a 1-year engineer contract (started in March 2019) to design and develop an experimental platform to acquire such ground truth data. The catheter is inserted in a translucent, silicon vascular phantom to generate contacts with no visual occlusion of the catheter shape. It is reconstructed from images acquired by a stereo rig made of two orthogonal high speed cameras. The motion is fully controlled by an original 3D-printed active device that induces accurate translation and rotation motions to the micro-tool. Monte-Carlo simulations are currently being carried out to certify the accuracy of the ground truth data produced by this system.

7.3. Image processing

Participants: Marie-Odile Berger, Fabien Pierre, Frédéric Sur.

7.3.1. Computational photomechanics

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

We have investigated the optimization of the pattern marking the specimen [13], which is the topic of several recent papers. Checkerboard is the optimized pattern in terms of sensor noise propagation when the signal is correctly sampled, but its periodicity causes convergence issues with DIC. The consequence is that checkerboards are not used in DIC applications although they are optimal in terms of sensor noise propagation. We have shown that it is possible to use LSA to estimate displacement and strain fields from checkerboard images, although LSA was originally designed to process 2D grid images. A comparative study of checkerboards and grids shows that, under similar experimental conditions, the noise level in displacement and strain maps obtained with checkerboards is lower than that obtained with classic 2D grids. A patent on this topic was filed [28].

Another scientific contribution concerns the restoration of displacement and strain maps. DIC and LSA both provide displacement fields equal to the actual one convoluted by a kernel known a priori. The kernel indeed corresponds to the Savitzky-Golay filter in DIC, and to the analysis window of the windowed Fourier transform used in LSA. While convolution reduces noise level, it also gives a systematic measurement error. We have proposed a deconvolution method to retrieve the actual displacement and strain fields from the output of DIC or LSA [12]. The proposed algorithm can be considered as a variant of Van Cittert deconvolution, based on the small strain assumption. It is demonstrated that it allows enhancing fine details in displacement and strain maps, while improving spatial resolution.

7.3.2. Cartoon-texture decomposition

Decomposing an image as the sum of geometric and textural components is a popular problem of image analysis. In this problem, known as cartoon and texture decomposition, the cartoon component is piecewise smooth, made of the geometric shapes of the images, and the texture component is made of stationary or quasi-stationary oscillatory patterns filling the shapes. Microtextures being characterized by their power spectrum, we propose to extract cartoon and texture components from the information provided by the power spectrum of image patches. The contribution of texture to the spectrum of a patch is detected as statistically significant spectral components with respect to a null hypothesis modeling the power spectrum of a non-textured patch. The null-hypothesis model is built upon a coarse cartoon representation obtained by a basic yet fast filtering algorithm of the literature. The coarse decomposition is obtained in the spatial domain and is an input of the proposed spectral approach. We thus design a "dual domain" method. The statistical model is also built upon the power spectrum of patches with similar textures across the image. The proposed approach therefore falls within the family of non-local methods. Compared to variational methods or fast filers, the proposed non-local dual-domain approach [16] is shown to achieve a good compromise between computation time and accuracy. Matlab code is publicly available.

7.3.3. Variational methods for image processing

The work described in [20] aims to couple the powerful prediction of the convolutional neural network (CNN) to the accuracy at pixel scale of the variational methods. We have focused on a CNN which is able to compute a statistical distribution of the colors for each pixel of the image based on a learning stage on a large color image database. A variational method able to select a color candidate among a given set while performing regularization of the result is combined with a CNN, to design a fully automatic image colorization framework with an improved accuracy in comparison with CNN alone. To solve the proposed model, we have proposed in [17] a novel accelerated alternating optimization scheme to solve block biconvex nonsmooth problems whose objectives can be split into smooth (separable) regularizers and simple coupling terms. The proposed method performs a Bregman distance-based generalization of the well-known forward–backward splitting for each block, along with an inertial strategy which aims at getting empirical acceleration. We discuss the theoretical convergence of the proposed scheme and provide numerical experiments on image colorization.

MFX Project-Team

7. New Results

7.1. Star-shaped Metrics for Mechanical Metamaterial Design

Participants: Jonàs Martínez, Mélina Skouras, Christian Schumacher, Samuel Hornus, Sylvain Lefebvre, Bernhard Thomaszewski.

Digital manufacturing technologies such as 3D printing and laser cutting enable us to fabricate designs with great geometric detail. One particular way of exploiting this capability is to create patterned sheet materials whose geometric structures can be tailored to control their macro-mechanical behavior.

A typical approach to model and analyze structured sheet materials is centered around the concept of a representative element—a tile—which is repeated, transformed, and laid out so as to generate a regular spatial tiling. Changing the shape of the representative tile allows to control macro-mechanical properties such as isotropy or negative Poisson's ratios. Generalizing this material design principle from a single representative tile to *families* of tiles that can be combined in a spatially-varying manner opens the door to structures with progressively-graded material properties.

At SIGGRAPH 2019 we have presented a method for designing mechanical metamaterials [14]. It is based on the novel concept of Voronoi diagrams induced by star-shaped metrics. As one of its central advantages, our approach supports interpolation between arbitrary metrics (see Figure 1). This capability opens up a rich space of tile geometries with interesting aesthetics and a wide range of mechanical properties. They include isotropic, tetragonal, orthotropic, as well as smoothly graded materials. We have validated the mechanical properties predicted by simulation through tensile tests on a set of physical prototypes. An open source C++ implementation of the technique can be found at https://github.com/mfx-inria/starshaped2d

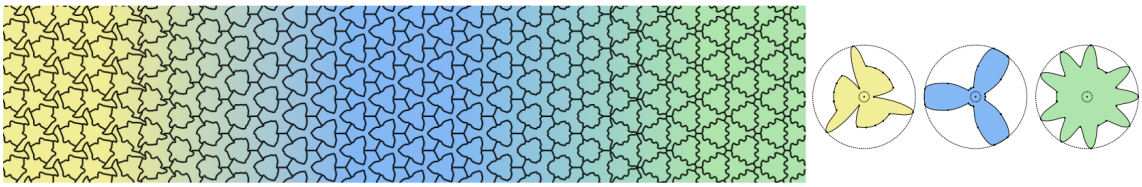


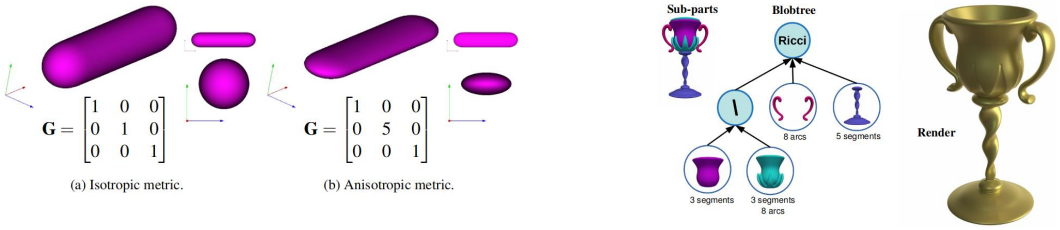
Figure 1. Our method generates a smoothly-graded pattern (left) when interpolating between three star-shaped distance functions (regular) on a regular honeycomb lattice. Each distance function is compactly parameterized with polar coordinates, allowing for simple interpolation in metric space as indicated by color-coding.

7.2. Anisotropic convolution surfaces

Participants: Alvaro Javier Fuentes Suárez, Evelyne Hubert, Cédric Zanni.

Skeletons, as a set of curves and/or surfaces centered inside a shape, provide a compact representation of the shape structure. Due to this property, skeletons have proved useful in many applications that range from shape analysis to 3D modeling and deformation. Convolution surfaces associate radii information to the skeleton and provide a simple way for users to rapidly define a shape. A convolution surface is an implicit surface defined as a level set of a scalar field, the convolution field, that is obtained by integrating a kernel function over the skeleton. This technique allows us to build a complex shape by modeling parts that combine into a smooth surface, independently of the smoothness of the skeleton itself. They also represent a volume with the convolution surface as its boundary and can therefore be combined with other composition operators from implicit modeling frameworks.

We have introduced anisotropic convolution surfaces [12], an extension that increases the modeling freedom by providing ellipse-like normal sections around 1D skeletons. It increases the diversity of shapes that can be generated from 1D skeletons, and reduces the need for 2D skeletons, while it still retains smoothness. We achieve anisotropy not just in the normal sections but also in the tangential direction. This allows sharper and steeper radius variation, and the control of thickness at skeleton endpoints (see Figure 2). The method is applied to skeletons represented by bi-arcs. It allows us to control precisely the orientation of anisotropy thanks to rotation minimizing frames. This work was presented at Shape Modeling International 2019.



(a) Isotropic metric. (b) Anisotropic metric.

Figure 2. Our method, based on an anisotropic metric, allows us to generate an ellipse-like cross section around 1D skeletons (segments, bi-arcs). The thickness around the skeleton can be controlled precisely both in the orthogonal cross-section and in the tangential direction giving finer control to the user. The density field generated can then be used in a classical implicit modeling framework.

7.3. Procedural Phasor Noise

Participants: Thibault Tricard, Semyon Efremov, Cédric Zanni, Fabrice Neyret, Jonàs Martínez, Sylvain Lefebvre.

Procedural pattern synthesis is a fundamental tool of Computer Graphics. In 2019 we introduced a new formulation that generates a wide range of patterns that could not be produced by other techniques. Our procedural *phasor noise* is based on a prior technique called Gabor noise, which creates oscillating patterns with accurate control over their frequency content (power spectrum). Gabor noise achieves this by summing a large number of Gabor kernels — Gaussian weighted sinewaves — distributed pseudo-randomly in space. Unfortunately Gabor noise suffers from local loss of contrast and lacks control over the shape of the oscillations (which always have a sinewave profile).

Our method solves these limitations by reformulating Gabor noise to expose its instantaneous phase. Once the phase obtained we can directly remap a periodic profile function onto it, to obtain an oscillating pattern of constant contrast and controlled profile geometry, while retaining all desirable properties of Gabor noise (see Figure 3). This unlocks two main applications. The first is in texture synthesis for computer graphics, to generate color, displacement and normal fields. The second is in additive manufacturing, where our method can be applied in 3D to generate a wide range of microstructures.

This work was done in collaboration with Fabrice Neyret (MAVERICK, Inria) and has been published in ACM Transactions on Graphics, in 2019 [17]. Thibault Tricard and Semyon Efremov did a joint presentation at ACM SIGGRAPH 2019.

7.4. Ribbed support vaults for 3D printing of hollowed objects

Participants: Thibault Tricard, Frédéric Claux, Sylvain Lefebvre.

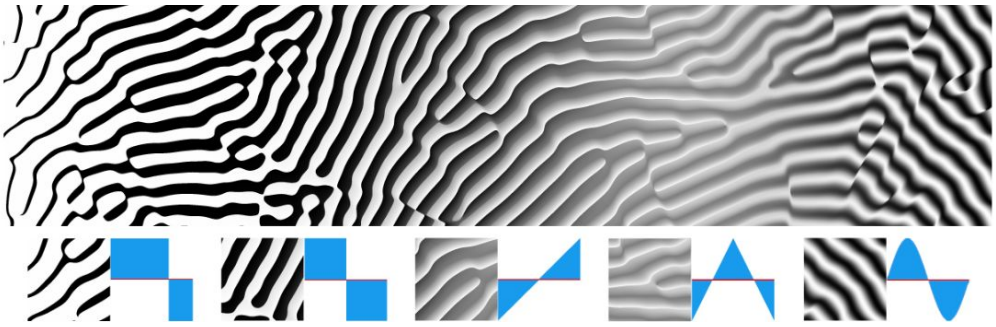


Figure 3. Phasor noise is a novel procedural function generating strongly oriented, coherent stripe patterns. The profiles of the oscillations are controlled (here: square, triangular, sine).

In additive manufacturing, and in particular with the popular filament-based fabrication, the printing time remains a major constraint. In a typical print, most of the time is spent filling the interior of the object. Based on this observation we explored how to print an object as empty as possible. The difficulty is that any deposited material has to be supported from below to prevent the object from collapsing.

We developed a simple, yet very efficient, algorithm that generates a lightweight ribbed support vault infill (see Figure 4). Our algorithm sweeps once through the slices from top to bottom, detects non-supported points, and connects them with a segment to the closest already supported points. The endpoints of open segments are eroded from one slice to the next. This process generates hierarchical ribbed support vaults that quickly retract and merge with the enclosing walls, while offering printability guarantees.

Our approach greatly reduces material usage (reaching parts as empty as 98%) and thus strongly reduces print time. Nevertheless it guarantees printability, and scales to very large inputs.

This work originated from the University of Limoges and was the master topic of Thibault Tricard, under the supervision of Frédéric Claux and in collaboration with Sylvain Lefebvre. The work was published in Computer Graphics Forum in June 2019 [16].

Figure 4. A 3D bunny model printed with our internal ribbed supports. It is mostly empty, with the ribbed vaults providing just enough support to prevent filament to fall during manufacturing.

7.5. CurviSlicer: Slightly curved slicing for 3-axis printers

Participants: Jimmy Étienne, Nicolas Ray, Daniele Panozzo, Samuel Hornus, Charlie C.l. Wang, Jonàs Martínez, Sara Mcmains, Marc Alexa, Brian Wyvill, Sylvain Lefebvre.

Most additive manufacturing processes fabricate objects by stacking planar layers of solidified material. As a result, produced parts exhibit a so-called staircase effect, which results from sampling slanted surfaces with parallel planes. Using thinner slices reduces this effect, but it always remains visible where layers almost align with the input surfaces. In this research we exploit the ability of some additive manufacturing processes to deposit material slightly out of plane to dramatically reduce these artifacts.

We focused in particular on the widespread Fused Filament Fabrication (FFF) technology, since most printers in this category can deposit along slightly curved paths, under deposition slope and thickness constraints. Our algorithm curves the layers, making them either follow the natural slope of the input surface or on the contrary, make them intersect the surfaces at a steeper angle thereby improving the sampling quality.

Rather than directly computing curved layers, our algorithm deforms the input model before slicing it with a standard planar approach. The deformation is optimized for reproduction accuracy. We demonstrate that this approach enables us to encode all fabrication constraints, including the guarantee of generating collision-free toolpaths, in a convex optimization that can be solved using a QP solver.

This work emerged from a problem solving session between its co-authors at the 17th international Bellairs Workshop on Computational Geometry (2018). It was then pursued during 2019 in the context of the PhD thesis of Jimmy Étienne and as a collaboration with Nicolas Ray (PIXEL, Inria). The work was published in ACM Transactions on Graphics in 2019 [11] and presented at ACM SIGGRAPH 2019 by Jimmy Étienne.

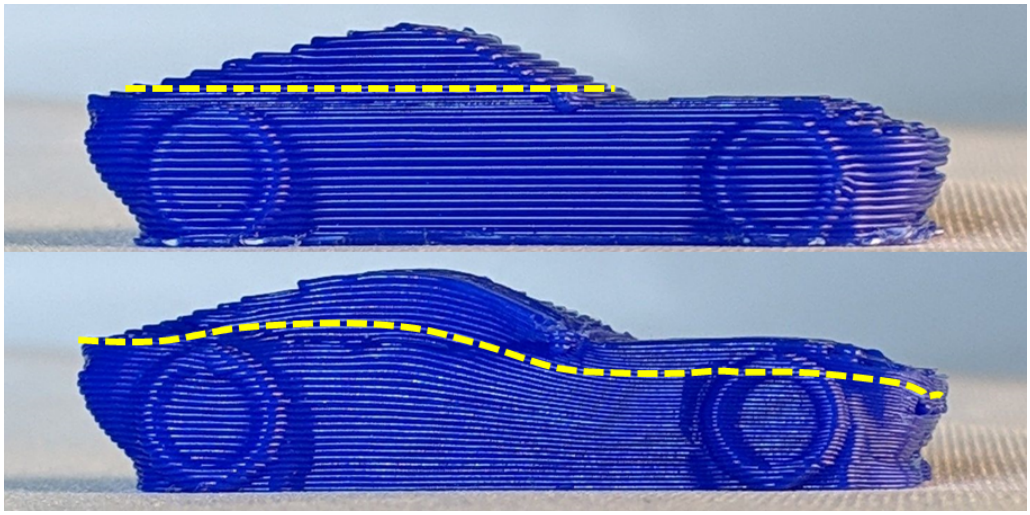


Figure 5. A 3D model printed with our technique. The algorithm automatically generates curved slices (right) to better reproduce the slanted surfaces, removing the staircase defect created by standard planar layers (top-left: standard, bottom-left: curved).

7.6. Extrusion-Based Ceramics Printing with Strictly-Continuous Deposition

Participants: Jean Hergel, Kevin Hinz, Bernhard Thomaszewski, Sylvain Lefebvre.

3D printing with extruded ceramic paste induces constraints that deviate significantly from standard thermoplastic materials. In particular existing path generation methods for thermoplastic materials rely on transfer moves to navigate between different print paths in a given layer. However, when printing with clay, these transfer moves can lead to severe artifacts and failure.

We explored how to eliminate transfer moves altogether by generating deposition paths that are continuous within and across layers. In each layer, we optimize a continuous support path with respect to length, smoothness, and distance to the model. Comparisons to existing path generation methods designed for thermoplastic materials show that our method substantially improves print quality and often makes the difference between success and failure.

This work was primarily done at the University of Montréal in collaboration with Sylvain Lefebvre. It was published in ACM Transactions on Graphics 2019 [13], and presented at SIGGRAPH Asia 2019 by Jean Hergel.

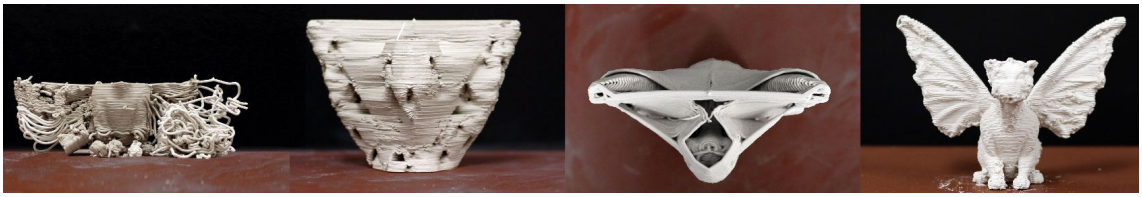


Figure 6. Our technique greatly improves the reliability of 3D printing with extruded clay.

MIMESIS Team

7. New Results

7.1. Real-time simulation of hyperelastic materials using Deep Learning

Participants: Andrea Mendizabal, Pablo Márquez-Neila, Stéphane Cotin.

The finite element method (FEM) is among the most commonly used numerical methods for solving engineering problems. Due to its computational cost, various ideas have been introduced to reduce computation times, such as domain decomposition, parallel computing, adaptive meshing, and model order reduction. In this work we propose the U-Mesh: a data-driven method based on a U-Net architecture that approximates the non-linear relation between a contact force and the displacement field computed by FE algorithm. We show that deep learning, one of the latest machine learning methods based on artificial neural networks, can enhance computational mechanics through its ability to encode highly non-linear models in a compact form. Our method is applied to three benchmark examples: a cantilever beam, an L-shape and a liver model subject to moving punctual loads. A comparison between our method and proper orthogonal decomposition (POD) is done. The results show that U-Mesh can perform very fast simulations on various geometries and topologies, mesh resolutions and number of input forces with very small errors. results were published in the Journal of Medical Image Analysis [23] (impact factor 8.5).

7.2. FEM-based confidence assessment of non-rigid registration

Participants: Paul Baksic, Hadrien Courtecuisse, Matthieu Chabanas, Bernard Bayle.

Non-rigid registration is often used for 3D representations during surgical procedures. It needs to provide good precision in order to guide the surgeon properly. We proposed in [25] a method that allows the computation of a local upper bound of the registration confidence over the whole organ volume. Using a bio-mechanical model, we apply tearing forces over the whole organ to compute the upper bound of the degrees of freedom left by the registrations constraints. Confrontation of our method with experimental data shows promising results to estimate the registration confidence. Indeed, the computed maximum error appears to be a real upper bound(see figure 4). A more advanced method was submitted at IPCAI 2020.

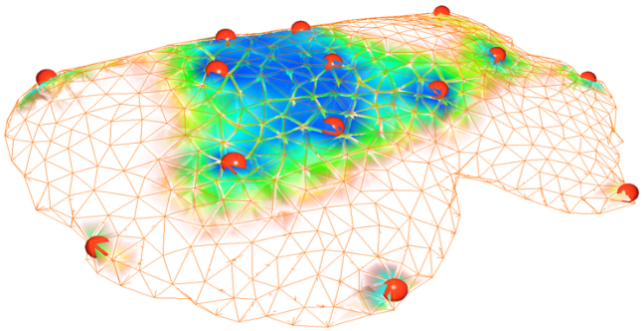


Figure 4. This is an example of confidence map given by our method on a registration of a lamb liver. The red dots are the registration constraint given by sensors. High confidence area are presented in blue. The area where the confidence is below the one needed for the surgery appears transparent.

7.3. Physics-based Deep Neural Network for Augmented Reality

Participants: Jean-Nicolas Brunet, Andrea Mendizabal, Antoine Petit, Nicolas Golse, Eric Vibert, Stéphane Cotin.

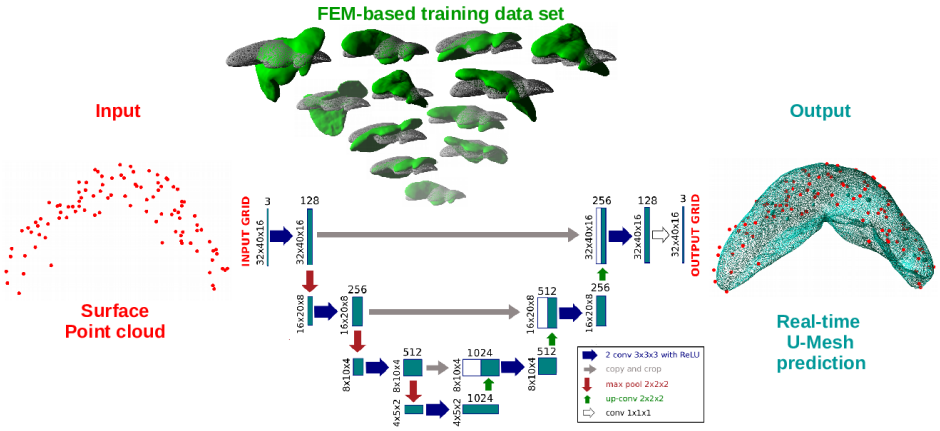


Figure 5. The U-Mesh framework allows for extremely fast simulations of soft tissues accounting for large non linear deformations.

We propose an approach combining a finite element method and a deep neural network to learn complex elastic deformations with the objective of providing augmented reality during hepatic surgery. Derived from the U-Net architecture, our network is built entirely from physics-based simulations of a preoperative segmentation of the organ (see figure 5). These simulations are performed using an immersed-boundary method, which offers several numerical and practical benefits, such as not requiring boundary-conforming volume elements. We perform a quantitative assessment of the method using synthetic and *ex vivo* patient data. Results show that the network is capable of solving the deformed state of the organ using only a sparse partial surface displacement data and achieve similar accuracy as a FEM solution, while being about 100x faster. When applied to an *ex vivo* liver example, we achieve the registration in only $3\,ms$ with a mean target registration error (TRE) of $2.9\,mm$. This results were presented at MICCAI 2019 [22].

7.4. Estimation of boundary conditions for patient-specific liver simulation during augmented surgery

Participants: Sergei Nikolaev, Stéphane Cotin.

Augmented liver surgery is an active research area that aims at improving the surgical outcome by enhancing the view of internal structures. However, to precisely estimate the position of these, an accurate model of the liver is required. While researchers have focused on proposing new models, algorithms for real-time computation or estimation of the tissue properties, very few have addressed the question of boundary conditions. Yet, they play a key role in the computation of the deformation. Boundary conditions mainly result from ligaments connecting the liver to its surrounding anatomy and limiting its motion. Unfortunately, ligaments' shapes and properties cannot be identified with preoperative imaging. We propose to estimate both the location and stiffness of ligaments by using a combination of a statistical atlas, numerical simulation, and Bayesian inference (fig. 6). Ligaments are modeled as polynomial springs connected to a liver finite element model. Their original location on the liver is based on an anatomical atlas, and their initial stiffness is taken

from the literature. These characteristics are then corrected using a reduced order unscented Kalman filter based on observations taken from the laparoscopic image stream. Our approach is evaluated using synthetic data and phantom data. Results show that our estimation of the boundary conditions improves the accuracy of the simulation by 75% when compared to typical methods involving Dirichlet boundary conditions. The results were submitted for a presentation at IPCAI 2020

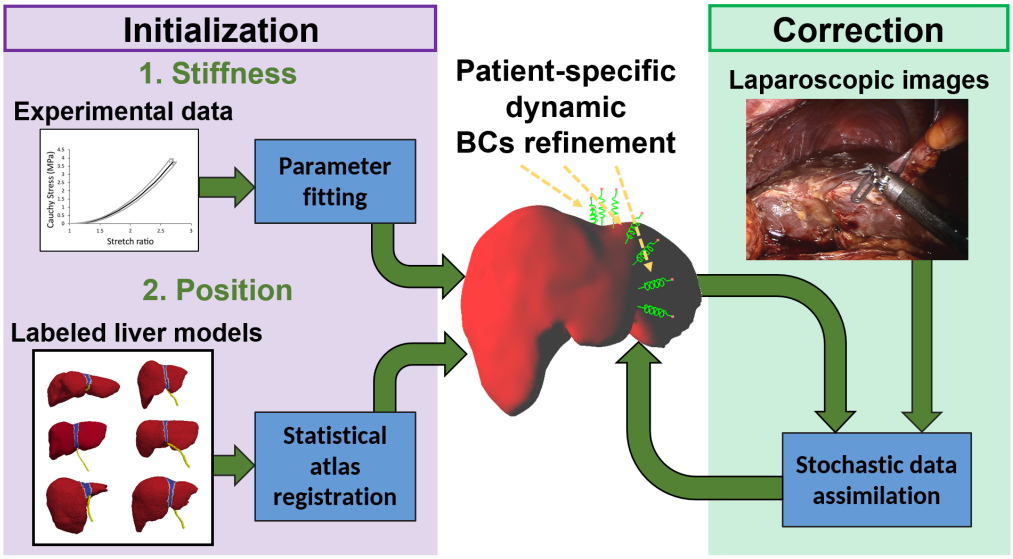


Figure 6. Overview of the boundary condition identification process. It contains two main steps. 1 - Initial approximation based on statistics from the processed model database and experimental data. 2 - Identification based on intraoperative patient-specific images.

7.5. Corotated meshless implicit dynamics for deformable bodies

Participants: Jean-Nicolas Brunet, Vincent Magnoux, Benoît Ozell, Stéphane Cotin.

We proposed a fast, stable and accurate meshless method to simulate geometrically non-linear elastic behaviors. To address the inherent limitations of finite element (FE) models, the discretization of the domain is simplified by removing the need to create polyhedral elements. The volumetric locking effect exhibited by incompressible materials in some linear FE models is also completely avoided. Our approach merely requires that the volume of the object be filled with a cloud of points (see figure 7). To minimize numerical errors, we constructed a corotational formulation around the quadrature positions that is well suited for large displacements containing small deformations. The equations of motion was integrated in time following an implicit scheme. The convergence rate and accuracy were validated through both stretching and bending case studies. Finally, results were presented using a set of examples that show how we can easily build a realistic physical model of various deformable bodies with little effort spent on the discretization of the domain. We presented our work at WSCG 2019 [21]. (Fig. 7).

7.6. The effect of discretization on parameter identification. Application to patient-specific simulations

Participants: Nava Schulmann, Igor Peterlik, Stéphane Cotin.

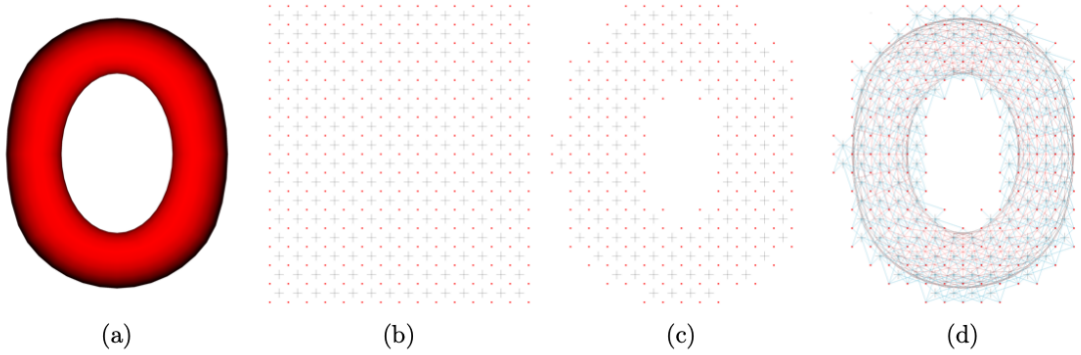


Figure 7. Volumetric discretizations of a 3D surface. (a) Surface mesh provided by the user. (b) Background grid where the grid's cubes are used to place the DOFs(degrees of freedom) and the integration points. (c) DOFs and integration points are cropped to fit the surface mesh. (d) A neighborhood of the closest particles is built around each integration point.

Identifying the elastic parameters of a finite element model from a dynamically acquired set of observations is a fundamental challenge in many data-driven medical applications, from soft surgical robotics to image-guided per-operative simulations. While various strategies exist to tackle the parameter-identification inverse problem [29], the effect of sub-optimal discretization, as often required in real-time applications, is largely overlooked. Indeed, the need to tune the parameter values in order to account for discretization-induced stiffening in specific models is reported in different works (e.g. [Chen et al., 2015, Anna et al., 2018]). However, to the best of our knowledge, no systematic study of this phenomenon exists to date, nor has any strategy to select optimal effective values been developed. Our work addresses the issue of parameter identification in coarsened meshes with special attention to the dynamical nature of the identification. We focus on the estimation of Young's moduli in simplified systems and show that the estimated stiffnesses are underestimated in a systematic manner when reducing the number of degrees of freedom. We also show that the effective stiffness of a given coarse mesh, when associated with an undersampled mesh discretization, is not constant but strongly depends on the prescribed deformations. These results show that the estimated parameters should not be considered as the true parameter value of the organ or tissue but instead are model-dependent values. We argue that Bayesian methods present a clear advantage w.r.t. classical minimization methods by their ability to efficiently adapt the local parameter values. The results were presented at CMBBE 2019 [26].

7.7. Elastic registration based on biomechanical graph matching

Participants: Jaime Garcia Guevara, Igor Peterlik, Marie-Odile Berger, Stéphane Cotin.

An automatic elastic registration method suited for vascularized organs is proposed. The vasculature in both the preoperative and intra-operative images is represented as a graph. A typical application of this method is the fusion of pre-operative information onto the organ during surgery, to compensate for the limited details provided by the intra-operative imaging modality (e.g. CBCT) and to cope with changes in the shape of the organ. Due differences in image modalities and organ deformation, each graph has a different topology and shape. The Adaptive Compliance Graph Matching (ACGM) method presented does not require any manual initialization, handles intra-operative nonrigid deformations of up to 65 mm and computes a complete displacement field over the organ from only the matched vasculature. ACGM is better than the previous Biomechanical Graph Matching method [3] (BGM) because it uses an efficient biomechanical vascularized liver model to compute the organ's transformation and compliance of vessel bifurcations. It allows to efficiently find the best graph matches with a novel compliance-based adaptive search. These contributions are evaluated

on ten realistic synthetic and two real porcine automatically segmented datasets. ACGM obtains better target registration error (TRE) than BGM, with an average TRE in the real datasets of 4.2 mm compared to 6.5 mm, respectively. It also is up to one order of magnitude faster, less dependent on the parameters used and more robust to noise. Figure 8 depicts the large deformation and the registered porcine CBCT and CTA data. Results were published in Annals of Biomedical Engineering (2019) [4].

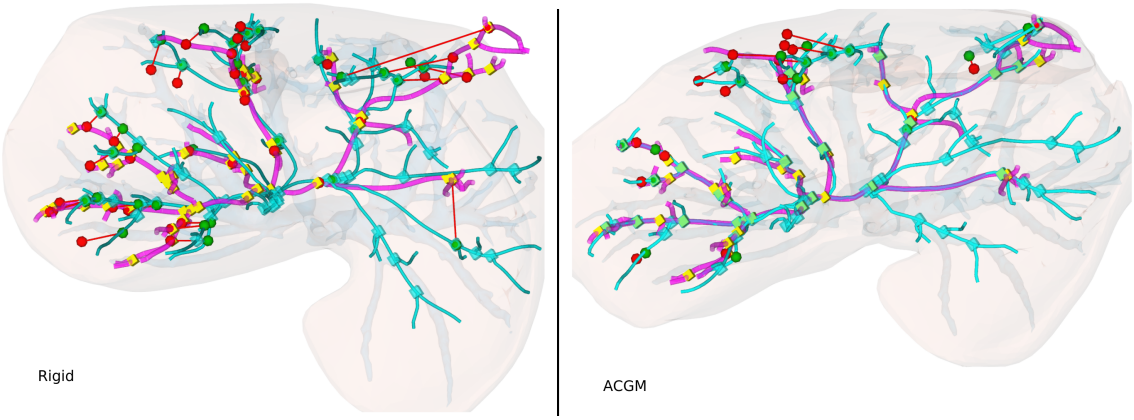


Figure 8. Registration between CTA and CBCT images. The target CBCT (in pink) and source CTA (in cyan) portal vein graphs are rendered with tubular structures. The graph nodes (bifurcations) are shown as cubic markers (in yellow for the target, cyan for the source and green for the matched). The augmented hepatic vein, which was only visible in the CTA image, is in transparent blue behind the portal veins graphs. The 37 target evaluation landmarks (red spheres) and their corresponding connected source landmarks (green spheres) and the liver structures are rigidly aligned and show the large intra-operative deformation (left image). The result of the registration process (right) shows the 16 registered landmarks.

MOCQUA Team

7. New Results

7.1. Semicomputable points in Euclidean spaces

- Participants: Mathieu Hoyrup, Donald Stull

Many natural problems/objects from theoretical computer science and logic are not decidable/computable, but semidecidable/semicomputable only: the halting problem, provability, domino problem, attractors of dynamical systems, etc. We pursue our program to study semicomputable objects in a systematic way. In this work, we focus on objects that can be described by finitely many real numbers, in particular polynomials and disks in the plane. Such objects can be identified with points of Euclidean spaces. We therefore introduce and study a notion of semicomputable point in Euclidean spaces, providing a multi-dimensional analog of a well-known unidimensional notion. The study involves ideas from linear algebra, convex analysis and computability. This work was presented at MFCS 2019 [27].

7.2. Computability on quasi-Polish spaces

- Participants: Mathieu Hoyrup, Cristobal Rojas, Victor Selivanov, Donald Stull

Descriptive Set Theory (DST) is a branch of topology which interacts very nicely with computability and logic. Indeed, these three theories involve measuring the complexity of describing objects in different ways (respectively as combinations of open sets, by programs, by formulae), which are intimately related. However, DST is traditionally developed on spaces relevant to mathematical analysis (Polish spaces), but not to theoretical computer science. The recently introduced quasi-Polish spaces are a much broader class of spaces including for instance Scott domains, important in functional programming. However, how to compute in such spaces is still not well-understood. In particular, quasi-Polish spaces can be characterized in many ways, so one has to choose the right definition to start with. We compare the computable versions of some of them, proving their non-equivalence, and focus on one of them, providing evidence that this notion is probably the right one. This work was presented at DCFS 2019 [26].

7.3. Degree spectra of Polish spaces

- Participants: Mathieu Hoyrup, Takayuki Kihara, Victor Selivanov

Mathematical objects can encode information. An obvious example is given by subsets of the plane: a text printed on a sheet of paper is a subset of the plane conveying information. However, when the object is submitted to deformations, what information can still be conveyed? What information is invariant under such deformations?

It is the core question in computable structure theory: for instance, what can be encoded in an infinite graph, which can be decoded from the structure itself and not from a particular presentation of the graph? Mathematically, what information is robust under graph isomorphism? It happens that much information can be encoded, for instance by using the lengths of the cycles in the graph.

Albegraic structures have been thoroughly studied from this perspective. However, the study of topological structures is almost inexistant, and more difficult (they are continuous while algebraic structures are often discrete). For instance, what information can be encoded in a subset of the plane, which is stable under continuous deformations (homeomorphisms)?

We have tackled this question during the visit of Takayuki Kihara and Victor Selivanov, and obtained many interesting results. For instance, we have proved that no direct information can be encoded (for instance, no infinite binary sequence can be extracted by an algorithm, unless the sequence is already computable). However, limit information can be encoded (for instance, a binary sequence can be encoded in such a way that a double-sequence converging to it can be extracted from the object by an algorithm). It is still open whether a single limit is possible.

A paper is still in preparation.

7.4. Computable SFTs

- Participants: Emmanuel Jeandel and Pascal Vanier

Previous works by the two participants have shown that there is a striking similarity between subshifts of finite type (tilings, coloring of the plane that do not contain a given set of patterns) and finitely presented groups (finitely generated groups with a finite number of equations).

This analogy can be described intuitively as follows: colors in subshifts corresponds to the generators of the groups, forbidden patterns correspond to the equations. Finite type is the same as finite presentation, and minimal subshifts correspond to simple groups.

The article [29] develops this analogy to computable objects: It is well known by the Higman-Thompson theorem that a finitely generated group is computable iff it is a subgroup of a simple group which is itself a subgroup of a finitely presented group. In this article, we give an equivalent for subshifts : a subshift is computable iff it is the restriction of a minimal subshift which is itself the restriction of a subshift of finie type.

7.5. Probabilistic cellular automata for problem solving

- Participants: Nazim Fatès, Irène Marcovici

Directly related to the theme exposed in Sec. 4.3 , we examined the problem of self-stabilisation, as introduced by Dijkstra in the 1970's, in the context of cellular automata [33]. More precisely, we examined how to stabilise k-colourings, that is, infinite grids which are coloured with k distinct colours in such a way that adjacent cells have different colours. The idea is that if, for any reason (e.g., noise, previous usage, tampering by an adversary), the colours of a finite number of cells in a valid k-colouring are modified, thus introducing errors, we can correct the system into a valid k-colouring by using local rules only. In other words, we designed cellular automaton rules which, starting from any finite perturbation of a valid k-colouring, reach a valid k-colouring in finite time. We discussed the different cases depending on the number of colours k, and propose some deterministic and probabilistic rules which solve the problem for $k \neq 3$. We also explained why the case $k = 3$ is more delicate. Finally, we proposed some insights on the more general setting of this problem, passing from k-colourings to other tilings (subshifts of finite type).

In the same spirit, we addressed the problem of detecting failures in a distributed network [30]. Our question is: if some components can break down over time, how can we detect that the failure rate has exceeded a given threshold without any central authority? We want to estimate the global state of the network, only through local interactions of components with their neighbours. In particular, we wish to reach a consensus on an alert state when the failure rate exceeds a given threshold. We used a cellular automaton in order to propose solutions in the case of a network with a grid structure. We compared three methods of self-organisation that are partly inspired by physical and biological phenomena. As an application, we envisioned sensor networks or any type of decentralised system with a great number of components.

Concerning the fundamental properties of asynchronous cellular automata, we presented a tutorial on the convergence properties of the 256 Elementary Cellular Automata under the fully asynchronous updating, that is, when only one cell is updated at each time step. We regrouped the results which have been presented in different articles and exposed a full analysis of the behaviour of finite systems with periodic boundary conditions. Our classification relies on the scaling properties of the average convergence time to a fixed point. We presented the different scaling laws that can be found, which fall in one of the following classes: logarithmic, linear, quadratic, exponential and non-converging. The techniques for quantifying this behaviour rely mainly on Markov chain theory and martingales. Most behaviours can be studied analytically but there are still many rules for which obtaining a formal characterisation of their convergence properties is still an open problem.

Our article on the global synchronisation problem was finally published [21]. In this problem, one is asked to find a cellular automaton which has the property that every initial condition evolves into a homogeneous blinking state. We studied this simple inverse problem for the case of one-dimensional systems with periodic boundary conditions. Two paradoxical observations were made: (a) despite the apparent simplicity of finding rules with good statistical results, there exist no perfect deterministic solutions to this problem, (b) if we allow the use of randomness in the local rule, constructing "perfect" stochastic solutions is easy. For the stochastic case, we give some rules for which the mean time of synchronisation varies quadratically with the number of cells and ask if this result can be improved. To explore more deeply the deterministic rules, we code our problem as a SAT problem and use SAT solvers to find rules that synchronise a large set of initial conditions.

7.6. Diagrammatic quantum computing

- Participants: Titouan Carette, Dominic Horsman, Emmanuel Jeandel, Simon Perdrix, Renaud Vilmart.

This year, we have contributed in several ways to the foundations and the applications of the ZX-calculus, a diagrammatic language for quantum computing.

Emmanuel Jeandel, Simon Perdrix, and Renaud Vilmart have introduced a general normal form for ZX-diagrams implying completeness results for various (almost all) fragments of quantum mechanics [28]. Renaud Vilmart has also introduced the simple, meaningful axiomatisation of the full ZX-calculus [31]. This two papers have been published at LICS'19.

Titouan Carette, Emmanuel Jeandel, Simon Perdrix, and Renaud Vilmart, have introduced a new simple categorical construction allowing to deal with non pure quantum evolutions (i.e. involving quantum measurements, discard of quantum systems, and probability mixtures). Wen this new construction coincides with the existing constructions, it provides simpler axiomatisation. For instance, this construction provides a complete equational theory for an extension of the ZX-calculus for arbitrary (non necessary pure) quantum evolutions. This result has been published at ICALP'19 [24].

Titouan Carette, Dominic Horsman (form LIG Grenoble) and Simon Perdrix have provided an axiomatisation for a scalable ZX-calculus where each wire represents a register of qubits, instead of a single qubit in the standard ZX-calculus. The scalable ZX-calculus allows compact representation of quantum algorithms, protocols and quantum codes. This result has been published at MFCS'19 [23]

7.7. Causal Graph Dynamics

- Participants: Pablo Arrighi, Simon Martiel, Simon Perdrix.

Causal Graph Dynamics extend Cellular Automata to arbitrary time-varying graphs of bounded degree. The whole graph evolves in discrete time steps, and this global evolution is required to have a number of symmetries: shift-invariance (it acts everywhere the same) and causality (information has a bounded speed of propagation). Pablo Arrighi (LIS, Marseille), Simon Martiel (Atos-Bull) and Simon Perdrix have considered a natural physics-like symmetry, namely reversibility. In particular, they extended two fundamental results on reversible cellular automata, by proving that the inverse of a causal graph dynamics is a causal graph dynamics, and that these reversible causal graph dynamics can be represented as finite-depth circuits of local reversible gates. These results have been published in the journal Natural Computing [16].

7.8. Contextuality in multipartite pseudo-telepathy graph games

- Participants: Anurag Anshu, Peter Høyer, Mehdi Mhalla, and Simon Perdrix.

Analyzing pseudo-telepathy graph games, Anurag Anshu, Peter Høyer, Mehdi Mhalla, and Simon Perdrix proposed a way to build contextuality scenarios exhibiting the quantum advantage using graph states. A new tool, called multipartiteness width, is introduced to investigate which scenarios are hard to decompose and to show that there exist graphs generating scenarios with a linear multipartiteness width. These results have been published in the Journal of Computer and System Science [15].

MULTISPEECH Project-Team

7. New Results

7.1. Beyond black-box supervised learning

Participants: Emmanuel Vincent, Denis Jouvet, Antoine Deleforge, Vincent Colotte, Irène Illina, Romain Serizel, Imran Sheikh, Pierre Champion, Adrien Dufraux, Ajinkya Kulkarni, Manuel Pariente, Georgios Zervakis, Zaineb Chelly Dagdia, Mehmet Ali Tugtekin Turan, Brij Mohan Lal Srivastava.

This year marked a significant increase in our research activities on domain-agnostic challenges relating to deep learning, such as the integration of domain knowledge, data efficiency, or privacy preservation. Our vision was illustrated by a keynote [18] and several talks [19], [17] on the key challenges and solutions.

7.1.1. Integrating domain knowledge

7.1.1.1. Integration of signal processing knowledge

State-of-the-art methods for single-channel speech enhancement or separation are based on end-to-end neural networks including learned real-valued filterbanks. We tackled two limitations of this approach. First, to ensure that the representation properly encodes phase properties as the short time Fourier transform and other conventional time-frequency transforms, we designed complex-valued analytic learned filterbanks and defined corresponding representations and masking strategies which outperformed the popular ConvTasNet algorithm [59]. Second, in order to allow generalization to mixtures of sources not seen together in training, we explored the modeling of speech spectra by variational autoencoders (VAEs), which are a variant of the probabilistic generative models classically used in source separation before the deep learning era. The VAEs are trained separately for each source and used to infer the source signals underlying a given mixture. Compared with existing iterative inference algorithms involving Gibbs sampling or gradient descent, we proposed a computationally efficient variational inference method based on an analytical derivation in which the encoder of the pre-learned VAE can be used to estimate the variational approximation of the true posterior [42], [55].

7.1.2. Learning from little/no labeled data

7.1.2.1. Learning from noisy labels

ASR systems are typically trained in a supervised fashion using manually labeled data. This labeling process incurs a high cost. Classical semi-supervised learning and transfer learning approaches to reduce the transcription cost achieve limited performance because the amount of knowledge that can be inferred from unlabeled data is intrinsically lower. We explored the middle ground where the training data are neither accurately labeled nor unlabeled but a not-so-expensive "noisy" transcription is available instead. We proposed a method to learn an end-to-end ASR model given a noise model and a single noisy transcription per utterance by adapting the auto segmentation criterion (ASG) loss to account for several possible transcriptions. Because the computation of this loss is intractable, we used a differentiable beam search algorithm that samples only the best alignments of the best transcriptions [32].

7.1.2.2. Transfer learning

We worked on the disentanglement of speaker, emotion and content in the acoustic domain for transferring expressivity information from one speaker to another one, particularly when only neutral speech data is available for the latter. In [36], we proposed to transfer the expressive characteristics through layer adaptation during the learning step. The obtained results highlighted that there is a difficult trade-off between speaker's identity to remove and the expressivity to transfer. We are now working on an approach relying on multiclass N-pair based deep metric learning in recurrent conditional variational autoencoder (RCVAE) for implementing a multispeaker expressive text-to-speech (TTS) system. The proposed approach conditions the text-to-speech system on speaker embeddings, and leads to a clustering with respect to emotion in a latent space. The deep

metric learning helps to reduce the intra-class variance and increase the inter-class variance. We transfer the expressivity by using the latent variables for each emotion to generate expressive speech in the voice of a different speaker for which no expressive speech is available. The performance measured shows the model's capability to transfer the expressivity while preserving the speaker's voice in synthesized speech.

7.1.3. Preserving privacy

Speech signals involve a lot of private information. With a few minutes of data, the speaker identity can be modeled for malicious purposes like voice cloning, spoofing, etc. To reduce this risk, we investigated speaker anonymization strategies based on voice conversion. In contrast to prior evaluations, we argue that different types of attackers can be defined depending on the extent of their knowledge. We compared three conversion methods in three attack scenarios, and showed that these methods fail to protect against an attacker that has extensive knowledge of the type of conversion and how it has been applied, but may provide some protection against less knowledgeable attackers [64]. As an alternative, we proposed an adversarial approach to learn representations that perform well for ASR while hiding speaker identity. Our results demonstrate that adversarial training dramatically reduces the closed-set speaker classification accuracy, but this does not translate into increased open-set speaker verification error [45]. We are currently organizing the 1st Voice Privacy Challenge in which these and other approaches will be further assessed and compared.

7.2. Speech production and perception

7.2.1. Articulatory modeling

Participants: Denis Jouvet, Anne Bonneau, Dominique Fohr, Yves Laprie, Vincent Colotte, Slim Ouni, Agnes Piquard-Kipffer, Elodie Gauthier, Manfred Pastatter, Théo Biasutto-Lervat, Sara Dahmani, Ioannis Douros, Amal Houidhek, Lou Lee, Shakeel Ahmad Sheikh, Anastasiia Tsukanova, Louis Delebecque, Valérian Girard, Thomas Girod, Seyed Ahmad Hosseini, Mathieu Hu, Leon Rohrbacher, Imene Zangar.

7.2.1.1. Articulatory speech synthesis

A number of simplifying assumptions have to be made in articulatory synthesis to enable the speech signal to be generated in a reasonable time. They mainly consist of approximating the propagation of the sound in the vocal tract as a plane wave and approximating the 3D vocal tract shape from the mid-sagittal shape [30], and also simplifying the vocal tract topology by removing small cavities [29]. The posture of the subject in the MRI machine was also investigated [31]. Vocal tract resonances were evaluated from the 3D acoustic simulation computed with the K-wave Matlab package from the complete 3D vocal tract shape recovered from MRI and compared to those of real speech [27].

We also developed an approach for using articulatory features for speech synthesis. The approach is based on a deep feed-forward neural network-based speech synthesizer trained with the standard recipe of Merlin on the audio recorded during real-time MRI (RT-MRI) acquisitions: denoised (and yet containing a residual noise of the MRI machine) speech in French and force-aligned state labels encoding phonetic and linguistic information [26]. The synthesizer was augmented with eight parameters representing articulatory information (lips opening and protrusion, distances between the tongue and the velum, between the velum and the pharyngeal wall, and between the tongue and the pharyngeal wall) that were automatically extracted from the captures and aligned with the audio signal and the linguistic specification.

7.2.1.2. Dynamics of vocal tract and glottal opening

The problem of creating a 3D dynamic atlas of the vocal tract that captures the dynamics of the articulators in all three dimensions has been addressed [28]. The core steps of the method are using 2D real time MRI in several sagittal planes and, after temporal alignment, combine them using adaptive kernel regression. As a preprocessing step, a reference space was created to be used in order to remove anatomical information of the speakers and keep only the variability in speech production for the construction of the atlas. Using adaptive kernel regression makes the choice of atlas time points independent of the time points of the frames that are used as an input for the atlas construction.

We started the development of a database of realistic glottal gestures which will be used to design the glottal opening dynamics in articulatory synthesis paradigms. Experimental measurements of glottal opening dynamics in VCV and VCCV sequences uttered by real subjects have been achieved thanks to a specifically designed external photoglottographic device (ePGG) [33]. The existence of different patterns of glottal opening is evidenced according to the class of the consonant articulated.

7.2.1.3. Multimodal coarticulation modeling

We have investigated labial coarticulation to animate a virtual face from speech. We experimented a sequential deep learning model, bidirectional gated recurrent networks, that have been used successfully in addressing the articulatory inversion problem. We have used phonetic information as input to ensure speaker independence. The initialization of the last layers of the network has greatly eased the training and helped to handle coarticulation. It relies on dimensionality reduction strategies, allowing injecting knowledge of useful latent representation of the visual data into the network. We have trained and evaluated the model with a corpus consisting of 4 hours of French speech, and we got a good average RMSE (Root Mean Square Error) close to 1.3 mm [21].

7.2.1.4. Identifying disfluency in stuttered speech

Within the ANR project BENEPHIDIRE, the goal is to automatically identify typical kinds of stuttering disfluency using acoustic and visual cues for their automatic detection. This year, we started analyzing existing stuttering acoustic speech datasets to characterize the kind of data.

7.2.2. Multimodal expressive speech

7.2.2.1. Arabic speech synthesis

We have continued working on Modern Standard Arabic speech synthesis with ENIT (École Nationale d'Ingénieurs de Tunis, Tunisia), using HMM and NN based approaches. This year we investigated the modeling of the fundamental frequency for Arabic speech synthesis with feedforward and recurrent DNN, and using specific linguistic features for Arabic like vowel quantity and gemination [50].

7.2.2.2. Expressive audiovisual synthesis

After acquiring a high quality expressive audio-visual corpus based on fine linguistic analysis, motion capture, and naturalistic acting techniques, we have analyzed, processed, and phonetically aligned it with speech. We used conditional variational autoencoders (CVAE) to generate the duration, acoustic and visual aspects of speech without using emotion labels. Perceptual experiments have confirmed the capacity of our system to generate recognizable emotions. Moreover, the generative nature of the CVAE allowed us to generate well-perceived nuances of the six emotions and to blend different emotions together [23].

7.2.2.3. Lipsync - synchronization of lips movements with speech

In the ATT Dynalips-2, we have developed an English version of the system which allows us having a full multilingual lipsync system. During this ATT, we also worked on the business aspects (business plan, funding, investment, search for clients, etc.) with the goal of creating a startup, spinoff of the laboratory, during 2020.

7.2.3. Categorization of sounds and prosody

7.2.3.1. Non-native speech production

We analysed voicing in sequences of obstruents with French as L1 and German as L2, that is languages characterized by strong differences in the voicing dimension, including assimilation direction. To that purpose, we studied the realizations of two sequences of obstruents, where the first consonant, in final position, was fortis, and the second consonant, in initial position, was either a lenis stop or a lenis fricative. These sequences lead to a possible anticipation of voicing in French, a direction not allowed in German given German phonetics and phonology. Highly variable realizations were observed: progressive and regressive assimilation, and absence of assimilation, often accompanied by an unexpected pause [22].

We also started investigating non-native phoneme productions of French learners of German in comparison to phoneme productions by native German speakers. A set of research questions has been developed for which a customized French/German corpus was designed, and recorded by one reference native speaker of German so far. Based on these initial recordings and according to the targeted research questions, analysis strategies and algorithms have been elaborated and implemented, and are ready to be employed onto a larger data set. By means of these methods we expect to access phonetic and phonological grounds of recurrently occurring mis-pronunciation.

7.2.3.2. Language and reading acquisition by children having some language impairments

We continued examining the schooling experience of 170 children, teenagers and young adults with specific language impairment (dysphasia, dyslexia, dysorthographia) facing severe difficulties in learning to read. The phonemic discrimination, phonological and phonemic analysis difficulties faced in their childhoods had raised reading difficulties, which the pupils did not overcome. With 120 of these young people, we explored the presence of other neuro-developmental disorders. We also studied their reading habits to achieve better understanding of their difficulties.

We continued investigating the acquisition of language by hard-of-hearing children via cued speech (i.e. augmenting the audiovisual speech signal by visualizing the syllables uttered via a code of hand positions). We have used a digital book and a children's picture book with 3 hard-of-hearing children in order to compare scaffolding by the speech therapist or the teacher in these two situations.

We started to examine language difficulties and related problems with children with autism and to work with their parents with a view to creating an environment conducive to their progress [39].

7.2.3.3. Computer assisted language learning

In the METAL project, experiments are planned to investigate the use of speech technologies for foreign language learning and to experiment with middle and high school students learning German. This includes tutoring aspects based on a talking head to show proper articulation of words and sentences; as well as using automatic tools derived from speech recognition technology, for analyzing student pronunciations. The web application is under development, and experiments have continued for analyzing the performance of an automatic detection of mispronunciations made by language learners.

The ALOE project deals with children learning to read. In this project, we are also involved with tutoring aspects based on a talking head, and with grapheme-to-phoneme conversion which is a critical tool for the development of the digitized version of ALOE reading learning tools (tools which were previously developed and offered only in a paper form).

7.2.3.4. Prosody

The keynote [15] summarizes recent research on speech processing and prosody, and presents the extraction of prosodic features, as well as their usage in various tasks. Prosodic correlates of discourse particles have been investigated further. It was found that occurrences of different discourse particles with the same pragmatic value have a great tendency to share the same prosodic pattern; hence, the question of their commutability have been studied [37].

7.3. Speech in its environment

Participants: Denis Jouvet, Antoine Deleforge, Dominique Fohr, Emmanuel Vincent, Md Sahidullah, Irène Illina, Odile Mella, Romain Serizel, Tulika Bose, Guillaume Carbajal, Diego Di Carlo, Sandipana Dowerah, Ashwin Geet Dsa, Adrien Dufraux, Raphaël Duroselle, Mathieu Fontaine, Nicolas Furnon, Mohamed Amine Menacer, Mauricio Michel Olvera Zambrano, Lauréline Perotin, Sunit Sivasankaran, Nicolas Turpault, Nicolas Zampieri, Ismaël Bada, Yassine Boudi, Mathieu Hu, Stephane Level.

7.3.1. Acoustic environment analysis

We are constantly surrounded by ambient sounds and rely heavily on them to obtain important information about our environment. Deep neural networks are useful to learn relevant representations of these sounds. Recent studies have demonstrated the potential of unsupervised representation learning using various flavors of the so-called triplet loss (a triplet is composed of the current sample, a so-called positive sample from the same class, and a negative sample from a different class), and compared it to supervised learning. To address real situations involving both a small labeled dataset and a large unlabeled one, we combined unsupervised and supervised triplet loss based learning into a semi-supervised representation learning approach and compared it with supervised and unsupervised representation learning depending on the ratio between the amount of labeled and unlabeled data [49].

Pursuing our involvement in the community on ambient sound recognition, we co-organized a task on large-scale sound event detection as part of the Detection and Classification of Acoustic Scenes and Events (DCASE) 2019 Challenge [48]. It focused on the problem of learning from audio segments that are either weakly labeled or not labeled, targeting domestic applications. We also published a summary of the outcomes of the DCASE 2017 Challenge, in which we had organized the first version of that task [7] and a detailed analysis of the submissions to that task in 2018 [16] and 2019 [61].

7.3.2. Speech enhancement and noise robustness

7.3.2.1. Sound source localization and counting

In multichannel scenarios, source localization, counting and separation are tightly related tasks. Concerning deep learning based speaker localization, we introduced the real and imaginary parts of the acoustic intensity vector in each time-frequency bin as suitable input features. We analyzed the inner working of the neural network using layerwise relevance propagation [9]. We also defined alternative regression-based approaches for localization and compared them to the usual classification-based approach on a discrete grid [43]. Lauréline Perotin successfully defended her PhD on this topic [2]. In [24], we proposed the first deep-learning based method for blindly estimating early acoustic echoes. We showed how estimates of these echoes enable 2D sound source localization with only two microphones near a reflective surface, a task normally impossible with traditional methods. Finally, we published our former work on motion planning for robot audition [8].

We organized the IEEE Signal Processing Cup 2019, an international competition aimed at teams of undergraduate students [5]. The tasks we proposed were on sound source localization using an array embedded in a flying drone for search and rescue application. Submissions to the first phase of the competition were opened from November 2018 to March 2019, and the final took place on May the 13th at the international conference ICASSP in Brighton. 20 teams of undergraduate students from 18 universities in 11 countries participated, for a total of 132 participants. The drone-embedded sound source localization dataset we recorded for the challenge was made publically available after the competition and has received over 1,000 file downloads as of December 2019.

7.3.2.2. Speech enhancement

We investigated the effect of speaker localization accuracy on deep learning based speech enhancement quality. To do so, we generated a multichannel, multispeaker, reverberated, noisy dataset inspired from the well studied WSJ0-2mix and evaluated enhancement performance in terms of the word error rate. We showed that the signal-to-interference ratio between the speakers has a higher impact on the ASR performance than the angular distance [62]. In addition, we proposed a deflation method which estimates the sources iteratively. At each iteration, we estimate the location of the speaker, derive the corresponding time-frequency mask and remove the estimated source from the mixture before estimating the next one [63].

In parallel, we introduced a method for joint reduction of acoustic echo, reverberation and noise. This method models the target and residual signals after linear echo cancellation and dereverberation using a multichannel Gaussian modeling framework and jointly represents their spectra by means of a neural network. We developed an iterative block-coordinate ascent algorithm to update all the filters. The proposed approach outperforms in terms of overall distortion a cascade of the individual approaches and a joint reduction approach which does not rely on a spectral model of the target and residual signals [53], [57].

In the context of ad-hoc acoustic antennas, we proposed to extend the distributed adaptive node-specific signal estimation approach to a neural networks framework. At each node, a local filtering is performed to send one signal to the other nodes where a mask is estimated by a neural network in order to compute a global multi-channel Wiener filter. In an array of two nodes, we showed that this additional signal can be efficiently taken into account to predict the masks and leads to better speech enhancement performances than when the mask estimation relies only on the local signals [58].

We have been pursuing our work on non-Gaussian heavy-tail models for signal processing, and notably investigated whether such models could be of use to devise new cost functions for the training of deep generative models for source separation [34]. In the case of speech enhancement, it turned out that the related log-likelihood functions could advantageously replace the more constraining squared-error and lead to significant performance gains.

We have also been pursuing our theoretical work on multichannel alpha-stable models, devising two new multichannel filtering methods that are adequate for processing multivariate heavy-tailed vectors. The related work is presented in Mathieu Fontaine's PhD manuscript [1].

7.3.2.3. Robust speech recognition

Achieving robust speech recognition in reverberant, noisy, multi-source conditions requires not only speech enhancement and separation but also robust acoustic modeling. In order to motivate further work by the community, we created the series of CHiME Speech Separation and Recognition Challenges in 2011. We are now organizing the 6th edition of the Challenge, and released the French dataset for ambient assisted living applications previously collected as part of the FUI VOICEHOME project [4].

7.3.2.4. Speaker recognition

Automatic speaker recognition systems give reasonably good recognition accuracy when adequate amount of speech data from clean conditions are used for enrollment and test. However, performance degrades substantially in real-world noisy conditions as well as due to the lack of adequate speech data. Apart from these two practical limitations, speaker recognition performance also degrades in presence of spoofing attacks [51] where playback voice or synthetic speech generated with voice conversion or speech synthesis methods are used by attackers to access a system protected with voice biometrics.

We have explored a new speech quality measure for quality-based fusion of speaker recognition systems. The quality metric is formulated with the zero-order statistics estimated during i-vector extraction. The proposed quality metric is shown to capture the speech duration information, and it has outperformed absolute-duration based quality measures when combining multiple speaker recognition systems. Noticeable improvement over existing methods have been observed specifically for the short-duration conditions [10].

We have also participated in speaker recognition evaluation campaigns NIST SREs and VoxSRC. For the NIST SREs [54], the key problem was to recognize speakers from low-quality telephone conversations. In addition, the language mismatch between system development and data under test made the problem more challenging. In VoxSRC, on the other hand, the main problem was to recognize speakers speaking short sentences of about 10 sec where the speech files are extracted from Youtube video clips. We have explored acoustic feature extraction, domain adaptation, parameter optimization and system fusion for these challenges. For VoxSRC, our system has shown substantial improvement over baseline results.

We also introduced a statistical uncertainty-aware method for robust i-vector based speaker verification in noisy conditions, that is the first one to improve over simple chaining of speech enhancement and speaker verification on the challenging NIST-SRE corpus mixed with real domestic noise and reverberation [44].

Robust speaker recognition is an essential component of speaker diarization systems. We have participated in the second DIHARD challenge where the key problem was the diarization of speech signals collected from diverse real-world conditions. We have explored speech activity detection, domain grouping, acoustic features, and speech enhancement for improved speaker recognition. Our proposed system has shown considerable improvement over the Kaldi-based baseline system provided by the challenge organizer [60].

We have co-organized the ASVspoof 2019 challenge, as an effort to develop next-generation countermeasures for automatic detection of spoofed/fake audio [46]. This involved creating the audio dataset, designing experiments, evaluating and analyzing the results. 154 teams or individuals participated in the challenge. The database is available for research and further exploration from Edinburgh DataShare, and has been downloaded/viewed more than a thousand times so far.

We have also analyzed whether target speaker selection can help in attacking speaker recognition systems with voice impersonation [35]. Our study reveals that impersonators were not successful in attacking the systems, however, the speaker similarity scores transfer well from the attacker's system to the attacked system [12]. Though there were modest changes in F0 and formants, we found that the impersonators were able to considerably change their speaking rates when mimicking targets.

7.3.2.5. Language identification

State-of-the-art spoken language identification systems are constituted of three modules: a frame level feature extractor, a segment level embedding extractor and a classifier. The performance of these systems degrades when facing mismatch between training and testing data. Although most domain adaptation methods focus on adaptation of the classifier, we have developed an unsupervised domain adaptation of the embedding extractor. The proposed approach consists in a modification of the loss of the segment level embedding extractor by adding a regularisation term. Experiments were conducted with respect to transmission channel mismatch between telephone and radio channels using the RATS corpus. The proposed method is superior to adaptation of the classifier and obtain the same performance as published language identification results but without using labelled data from the target domain.

7.3.3. Linguistic and semantic processing

7.3.3.1. Transcription, translation, summarization and comparison of videos

Within the AMIS project, we studied different subjects related to the processing of videos. The first one concerns the machine translation of Arabic-English code-switched documents [41]. Code-switching is defined as the use of more than one language by a speaker within an utterance. The second one deals with the summarization of videos into a target language [11]. This exploits research carried on in several areas including video summarization, speech recognition, machine translation, audio summarization and speech segmentation. One of the big challenges of this work was to conceive a way to evaluate objectively a system composed of several components given that each of them has its limits and that errors propagate through the components. A third aspect was a method for extracting text-based summarization of Arabic videos [40]. The automatic speech recognition system developed to transcribe the videos has been adapted to the Algerian dialect, and additional modules were developed for segmenting the flow of recognized word into sentences, and for summarization. Finally the last aspect concerns the comparison of the opinions of two videos in two different languages [20]. Evaluations have been carried on comparable videos extracted from a corpus of 1503 Arabic and 1874 English videos.

7.3.3.2. Detection of hate speech in social media

The spectacular expansion of the Internet led to the development of a new research problem in natural language processing, the automatic detection of hate speech, since many countries prohibit hate speech in public media. In the context of the M-PHASIS project, we proposed a new approach for the classification of tweets, aiming to predict whether a tweet is abusive, hate or neither. We compare different unsupervised word representations and DNN classifiers, and study the robustness of the proposed approaches to adversarial attacks when adding one (healthy or toxic) word. We are evaluating the proposed methodology on the English Wikipedia Detox corpus and on a Twitter corpus.

7.3.3.3. Introduction of semantic information in an automatic speech recognition system

In current state-of-the-art automatic speech recognition systems, N-gram based models are used to take into account language information. They have a local view and are mainly based on syntax. The introduction of semantic information and longer term information in a recognition system should make it possible to remove some ambiguities and reduce the error rate of the system. Within the MMT project, we are proposing and

evaluating methods for integrating semantic information into our speech recognition system through the use of various word embeddings.

7.3.3.4. Music language modeling

Similarly to speech, language models play a key role in music modeling. We represented the hierarchical structure of a temporal scenario (for instance, a chord progression) via a phrase structure grammar and proposed a method to automatically induce this grammar from a corpus and to exploit it in the context of machine improvisation [6].

NEUROSYS Project-Team

7. New Results

7.1. From the microscopic to the mesoscopic scale

Participants: Laure Buhry, Amélie Aussel, Nathalie Azevedo Carvalho, Dominique Martinez (CNRS), Radu Ranta (Univ. Lorraine, CRAN).
In collaboration with Harry Tran (Univ. Lorraine, CRAN), Louise Tyvaert (Univ. Lorraine, CRAN, CHRU Nancy), Olivier Aron (Univ. Lorraine, CRAN, CHRU Nancy), Sylvain Contassot-Vivier (Univ. Lorraine),

7.1.1. Hippocampal oscillatory activity

7.1.1.1. Healthy hippocampus

We proposed a detailed anatomical and mathematical model of the hippocampal formation for the generation of healthy hippocampal activity, especially sharp-wave ripples and theta-nested gamma oscillations [24], [25]. Indeed, the mechanisms underlying the broad variety of oscillatory rhythms measured in the hippocampus during the sleep-wake cycle are not yet fully understood. We proposed a computational model of the hippocampal formation based on a realistic topology and synaptic connectivity, and we analyzed the effect of different changes on the network, namely the variation of synaptic conductances, the variations of the CAN channel conductance and the variation of inputs. By using a detailed simulation of intracerebral recordings, we showed that this model is able to reproduce both the theta-nested gamma oscillations that are seen in awake brains and the sharp-wave ripple complexes measured during slow-wave sleep. The results of our simulations support the idea that the functional connectivity of the hippocampus, modulated by the sleep-wake variations in Acetylcholine concentration, is a key factor in controlling its rhythms [24].

We further extended this work with an extensive study of the parameter range of the healthy hippocampus activity and showed that the "healthy model" was unable to reproduce pathological hippocampal oscillations observed in temporal lobe epilepsy.

7.1.1.2. Modeling LFP measures

The development of this model was also the opportunity to extend our model of the measure of the local field potential (LFP) and to study the contribution of spikes (not only synaptic currents) to the generation of the LFP. Indeed, simulating extracellular recordings of neuronal populations is a challenging task for understanding the nature of extracellular field potentials (LFPs), investigating specific brain structures and mapping cognitive functions. In general, it is assumed that extracellular recording devices (micro and/or macro-electrodes) record a mixture of low frequency patterns, mainly attributed to the synaptic currents and high-frequency components reflecting action potential (APs) activity. Simulating such signals often requires a high computational burden due to the multicompartmental neuron models used. Therefore, different LFP proxies coexist in the literature, most of them only reproducing some of the features of experimental signals. This may be an issue in producing and validating computational models of phenomena where the fast and slow components of neural activity are equally important, such as hippocampal oscillations. In this part of the work, we proposed an original approach for simulating large-scale neural networks efficiently while computing a realistic approximation of the LFP signal including extracellular signatures of both synaptic and action potentials [26]. We applied this method on the hippocampal network we developed earlier and compared the simulated signal with intracranial measurements from human patients.

7.1.1.3. Epilepsy of the mesial temporal lobe

The model described above has then been extended to include pathological changes observed in temporal lobe epilepsy, the future goal being to better understand the generation and propagation of epileptic activity throughout the brain, and therefore to investigate new potential therapeutic targets.

The mechanisms underlying the generation of hippocampal epileptic seizures and interictal events during the sleep-wake cycle are not yet fully understood. In this article, based on our previous computational modeling work of the hippocampal formation based on realistic topology and synaptic connectivity, we study the role of network specificity and channel pathological conditions of the epileptic hippocampus in the generation and maintenance of seizures and interictal oscillations. Indeed, the epilepsies of the mesial temporal lobe are associated with hippocampal neuronal and axonal loss, mossy fiber sprouting and channelopathies, namely impaired potassium and chlore dynamics. We show, through the simulations of hippocampal activity during slow-wave sleep and wakefulness that: (i) both mossy fiber sprouting and sclerosis account for epileptic seizures, (ii) high hippocampal sclerosis with low sprouting suppresses seizures, (iii) impaired potassium and chloride dynamics have little influence on the generation of seizures, (iv) but do have an influence on interictal spikes that decreases with high mossy fiber sprouting. A manuscript is in preparation for the Journal of Neuroscience.

7.1.2. Synchronization phenomena in neuronal network models

From a more computational point of view, we got interested in interneuronal gamma oscillations and synchronization in hippocampus-like networks via different models, especially in adaptive exponential integrate-and-fire neurons. Fast neuronal oscillations in gamma frequencies are observed in neocortex and hippocampus during essential arousal behaviors. Through a four-variable Hodgkin–Huxley type model, Wang and Buzsáki have numerically demonstrated that such rhythmic activity can emerge from a random network of GABAergic interneurons via minimum synaptic inputs. In this case, the intrinsic neuronal characteristics and network structure act as the main drive of the rhythm. We investigate inhibitory network synchrony with a low complexity, two-variable adaptive exponential integrate-and-fire (AdEx) model, whose parameters possess strong physiological relevances, and provide a comparison with the two-variable Izhikevich model and Morris–Lecar model. Despite the simplicity of these three models, the AdEx model shares two important results with the previous biophysically detailed Hodgkin–Huxley type model: the minimum number of synaptic inputs necessary to initiate network gamma-band rhythms remains the same, and this number is weakly dependent on the network size. Meanwhile, Izhikevich and Morris–Lecar neurons demonstrate different results in this study. We further investigated the necessary neuronal, synaptic and connectivity properties, including gap junctions and shunting inhibitions, for AdEx model leading to sparse and random network synchrony in gamma rhythms and nested theta gamma rhythms. These findings suggest a computationally more tractable framework for studying synchronized networks in inducing cerebral gamma band activities.

7.1.3. Event-driven simulation of large scale neural models with on-demand connectivity generation

Accurate simulations of brain structures is a major problem in neuroscience. Many works are dedicated to design better models or to develop more efficient simulation schemes. In this work, we propose to combine time-stepping numerical integration of Hodgkin-Huxley type neurons with event-driven updating of the synaptic currents. A spike detection method was also developed to determine the spike time more precisely in order to preserve the second-order Runge-Kutta methods. This hybrid approach allows us to regenerate the outgoing connections at each event, thereby avoiding the storage of the connectivity. Consequently, memory consumption and execution time are significantly reduced while preserving accurate simulations, especially spike times of detailed point neuron models. The efficiency of the method has been demonstrated on the simulation of 10^6 interconnected MSN neurons with Parkinson disease (an article has been submitted to Frontiers in Neuroinformatics) [23].

7.2. From the Mesoscopic to the Macroscopic Scale

Participants: Laurent Bougrain, Sébastien Rimbert, Oleksii Avilov, Rahaf Al-Chwa, Anais Coster, Elina Ortega Herrera, Nicolas Rault, Radu Ranta (univ. Lorraine).
In collaboration with Stéphanie Fleck (Univ. Lorraine)

7.2.1. On source space resolution in EEG brain imaging for motor imagery

Brain source localization accuracy is known to be dependent on the EEG sensor placement over the head surface. In Brain-Computer Interfaces (BCI), according to the paradigm used, Motor Imagery (MI) and Steady-State Visual Evoked Potential (SSVEP) in particular, electrodes are not well distributed over the head, and their number is not standardized as in classical clinical applications. We proposed a method for quantifying the expected quality of source localization with respect of the sensor placement, known as EEG montage. Our method, based on a subspace correlation metric, can be used to assess which brain sources can be distinguished (as they generate sufficiently different potentials on the electrodes), and also to identify regions/volumes in which precise source localization is impossible (i.e. all sources inside those regions could generate similar electrode potentials). In particular, for a MI dedicated montage, we show that source localization is less precise than for standard montages, although the local density of electrodes over the areas of interest is higher [13].

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

Hundreds of millions of general anesthesia are performed each year on patients all over the world. Among these patients, 0.2 to 1.3% are victims of Accidental Awareness during General Anesthesia (AAGA), i.e. an unexpected awakening of the patient during a surgical procedure under general anesthesia. This terrifying experience may be very traumatic for the patient and should be avoided by the anesthesiologists. Out of all the techniques used to reduce these awakenings, there is currently no solution based on the EEG signal to detect this phenomenon efficiently. Since the first reflex for a patient during an AAGA is to move, a passive BCI based on the intention of movement is conceivable. However, the challenge of using such BCI is that the intention to move from the waking patient is not initiated by a trigger that could be used to guide a classifier. We proposed a solution based on Median Nerve Stimulation (MNS), which causes specific modulations in the motor cortex and can be altered by an intention of movement. We showed that MNS may provide a foundation for an innovative BCI that would allow the detection of an AAGA [15], [7].

Moreover the way in which propofol (i.e., an anesthetic commonly used for the general anesthesia induction) affects motor brain activity within the electroencephalographic (EEG) signal has been poorly investigated and is not clearly understood. For this reason, a detailed study of the motor activity behavior with a step-wise increasing dose of propofol is required and would provide a proof of concept for such an innovative BCI. We started a study to highlight the occurrence of movement attempt patterns, mainly changes in oscillations called event-related desynchronization (ERD) and event-related synchronization (ERS), in the EEG signal over the motor cortex, in healthy subjects, without and under propofol sedation, during four different motor tasks [8], [12].

7.2.3. Can a subjective questionnaire be used as a brain-computer interface performance predictor?

Predicting a subject's ability to use a Brain Computer Interface (BCI) is one of the major issues in the BCI domain. Relevant applications of forecasting BCI performance include: the ability to adapt the BCI to the needs and expectations of the user; assessing the efficiency of BCI use in stroke rehabilitation; and finally, homogenizing a research population. A limited number of recent studies have proposed the use of subjective questionnaires, such as, the Motor Imagery Questionnaire Revised-Second Edition (MIQ-RS). However, further research is necessary to confirm the effectiveness of this type of subjective questionnaire as a BCI performance estimation tool. We aimed to answer the following questions: can the MIQ-RS be used to estimate the performance of an MI-based BCI? If not, can we identify different markers that could be used as performance estimators? To answer these questions, we recorded EEG signals from 35 voluntary healthy subjects during BCI use. The subjects previously had completed the MIQ-RS questionnaire. We conducted an offline analysis to assess the correlation between the questionnaire scores related to Kinesthetic and Motor imagery tasks and the performances of four classification methods. Our results show no significant correlation between BCI performance and the MIQ-RS scores. However, we revealed that BCI performance is correlated to habits and frequency of practicing manual activities [6].

7.2.3.1. Hypnotic State Modulates Sensorimotor Beta Rhythms During Real Movement and Motor Imagery

Hypnosis techniques are currently used in the medical field and directly influence the patient's state of relaxation, perception of the body, and its visual imagination. There is evidence to suggest that a hypnotic state may help patients to better achieve tasks of motor imagination, which is central in the rehabilitation protocols after a stroke. However, the hypnosis techniques could also alter activity in the motor cortex. To the best of our knowledge, the impact of hypnosis on the EEG signal during a movement or an imagined movement is poorly investigated. In particular, how event-related desynchronization (ERD) and event-related synchronization (ERS) patterns would be modulated for different motor tasks may provide a better understanding of the potential benefits of hypnosis for stroke rehabilitation. To investigate this purpose, we recorded EEG signals from 23 healthy volunteers who performed real movements and motor imageries in a closed eye condition. Our results suggest that the state of hypnosis changes the sensorimotor beta rhythm during the ERD phase but maintains the ERS phase in the mu and beta frequency band, suggesting a different activation of the motor cortex in a hypnotized state [14], [9].

ORPAILLEUR Project-Team

7. New Results

7.1. Mining of Complex Data

Participants: Nacira Abbas, Guilherme Alves Da Silva, Alexandre Bazin, Alexandre Blansché, Lydia Boudjeloud-Assala, Quentin Brabant, Brieuc Conan-Guez, Miguel Couceiro, Adrien Coulet, Sébastien Da Silva, Alain Gély, Laurine Huber, Nyoman Juniarta, Florence Le Ber, Tatiana Makhalova, Jean-François Mari, Pierre Monnin, Amedeo Napoli, Laureline Nevin, Abdelkader Ouali, François Pirot, Frédéric Pennerath, Justine Reynaud, Claire Theobald, Yannick Toussaint, Laura Alejandra Zanella Calzada, Georgios Zervakis.

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 and possibly large data. FCA, which usually applies to binary data-tables, can be adapted to work on more complex data. In this way, we have contributed to some main extensions of FCA, namely Pattern Structures, Relational Concept Analysis and application of the "Minimum Description Length" (MDL) within FCA. Pattern Structures (PS [80], [85]) allow building a concept lattice from complex data, e.g. numbers, sequences, trees and graphs. Relational Concept Analysis (RCA [90]) is able to analyze objects described both by binary and relational attributes and can play an important role in text classification and text mining. Many developments were carried out in pattern mining and FCA for improving data mining algorithms and their applicability, and for solving some specific problems such as information retrieval, discovery of functional dependencies and biclustering.

We got several results in the discovery of approximate functional dependencies [29], the mining of RDF data, the visualization of the discovered patterns, and redescription mining. Moreover, based on Relational Concept Analysis, we worked also on the discovery and representation of n-ary relations in the framework of FCA [3]. In the same way, reusing ideas form subgroup discovery, we have initiated a whole line of research on the covering of the pattern spaces based on the "Minimum Description Length" (MDL) principle and we are working on the adaptation of MDL within the FCA framework [36] [7].

We are also working on designing hybrid mining methods, based on mining methods able to deal with symbolic and numerical data in parallel. In the context of the GEENAGE project, we are interested in the identification, in biomedical data, of biomarkers that are predictive of the development of diseases in the elderly population. Actually, the data are issued from a preceding study on metabolomic data for the detection of diabetes of type 2 [23]. The problem can be viewed as a classification problem where features which are predictive of a class should be identified. This leads us to study the notions of prediction and discrimination in classification problems. Combining numerical machine learning methods such as random forests, neural networks, and SVM, then multicriteria decision making methods (Pareto fronts), and pattern mining methods (including FCA), we developed a hybrid mining approach for selecting the features which are the most predictive and/or discriminant. Then the selected features are organized within a concept lattice to be presented to the analyst together with the reasons for their selection. The concept lattice makes more easy and natural the understanding of the feature selection. As such, this approach can also be seen as an explicable mining method, where the output includes the reasons for which features are selected in terms of prediction and discrimination.

In the framework of the CrossCult European Project about cultural heritage, we worked on the mining of visitor trajectories in a museum or a touristic site. We presented a theoretical and practical research work about the characterization of visitor trajectories and the mining of these trajectories as sequences [83], [84]. The mining process is based on two approaches in the framework of FCA. We focused on different types of sequences and more precisely on subsequences without any constraint and frequent contiguous subsequences. We also introduced a similarity measure allowing us to build a hierarchical classification which is used for interpretation and characterization of the trajectories. A natural extension of this research work

on the characterization of trajectories is related to recommendation, i.e. based on an actual trajectory, how to recommend next items to be visited? Biclustering is a good candidate for designing recommendation methods and we especially worked on this topic this current year. In particular, we worked on several aspects of biclustering in the framework of FCA and we also tried to build a generic and unified framework from which several biclustering methods can be derived [34], [52].

7.1.2. Redescription Mining

Redescription mining is one of the pattern mining methods developed in the team. This method aims at finding distinct common characterizations of the same objects and, reciprocally, at identifying sets of objects having multiple shared descriptions [89]. This is motivated by the idea that in scientific investigations data oftentimes have different nature. For example, 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 existing between these different aspects. A practical example in biology consists in finding geographical areas having two characterizations, one in terms of their climatic profile and one in terms of the occupying species. Discovering such redescriptions can contribute to better understand the influence of climate over species distribution. Besides biology, redescription mining can be applied in many concrete domains.

Following this way, we applied redescription mining for analyzing and mining RDF data in the web of data with the objective of discovering definitions of concepts and as well disjunctions (incompatibilities) of concepts, for completing knowledge bases in a semi-automated way [41] [10]. Redescription mining is well adapted to the task as a definition is naturally based on two sides of an equation, a left-hand side and a right-hand side.

7.1.3. Text Mining

The research work in text mining is mainly based on two ongoing PhD theses. The first research subject is related to the study of discourse and argumentation structures in a text based on tree mining and redescription mining [33], while the second research work is related to the mining of Pubmed abstracts about rare diseases. In the first research line, we investigate the similarities existing between discourse and argumentation structures by aligning subtrees in a corpus where texts are annotated. Contrasting related work, here we focus on the comparison of substructures within the text and not only the matching of relations. Based on data mining techniques such as tree mining and redescription mining, we are able to show that the structures underlying discourse and argumentation can be (partially) aligned. There the annotations related to discourse and argumentation allow us to derive a mapping between the structures. In addition, the approach enables the study of similarities between diverse discourse structures, and as well the differences in terms of expressive power.

In the second research line, the objective is to discover features related to rare diseases, e.g. symptoms, related diseases, treatments, and possible disease evolution or variations. The texts to be analyzed are from Pubmed, i.e. a platform collecting millions of publications in the medical domain. This research project aims at developing new methods and tools for supporting knowledge discovery in textual data by combining methods from Natural Language Processing (NLP) and Knowledge Discovery in Databases (KDD). Here a key idea is to design an interacting and convergent process where NLP methods are used for guiding text mining and KDD methods are used for analyzing textual documents. In this way, NLP is based on extraction of general and temporal information, while KDD methods are especially based on pattern mining, FCA, and graph mining.

7.1.4. Consensus, Aggregation Functions and Multicriteria Decision Aiding Functions

Aggregation and consensus theory study processes dealing with the problem of merging or fusing several objects, e.g., numerical or qualitative data, preferences or other relational structures, into a single or several objects of similar type and that best represents them in some way. Such processes are modeled by so-called aggregation or consensus functions [81], [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 the context of the project RHU "Fighting Heart Failure" (that aims to identify and describe relevant bio-profiles of patients suffering from heart failure) we are dealing with biomedical data, highly complex and heterogeneous, that include, among other, sociodemographical aspects, biological and clinical features, drugs taken by the patients, etc. One of our main challenges is to define relevant aggregation operators on this heterogeneous patient data that lead to a clustering of the patients. Each cluster should correspond to a bio-profile, i.e. a subgroup of patients sharing the same form of the disease and thus the same diagnosis and medical care strategy. We are working on ways for comparing and clustering patients, namely, by defining multidimensional similarity measures on this complex and heterogeneous biomedical data. To this end, we recently proposed a novel approach, that we named "unsupervised extremely randomized trees" (UET), that is inspired by the frameworks of unsupervised random forests (URF) and of extremely randomized trees (ET). The empirical study of UET showed that it outperforms existing methods (such as URF) in running time, while giving better clustering. However, UET was implemented for numerical data only, and this is a drawback when dealing with biomedical data.

To overcome this limitation we have recently proposed an adaptation of UET [63] that is agnostic to variable types –numerical, symbolic or both–, that is robust to noise, to correlated variables, and to monotone transformations, thus drastically limiting the need for preprocessing. In addition, this provides similarity measures for clustering purposes that show outperforming results compared to state-of-the-art clustering methodologies.

Also, motivated by current trends in graph clustering for applications in the semantic web, and community identification in computer and social networks, we recently proposed a novel graph clustering method, i.e. GraphTrees [61], that is based on random decision trees to compute pairwise dissimilarities between vertices in vertex-attributed graphs. Unlike existing methodologies, it applies directly to graphs whose vertex-attributes are heterogeneous without preprocessing, and with promising results in benchmark datasets that are competitive with best known methods.

In the context of the project ISIPA, we mainly focused on the utility-based preference model in which preferences are represented as an aggregation of preferences over different attributes, structured or not, both in the numerical and qualitative settings. In the latter case, the Sugeno integral is widely used in multiple criteria decision making and decision under uncertainty, for computing global evaluations of items based on local evaluations (utilities). The combination of a Sugeno integral with local utilities is called a Sugeno utility functional (SUF). A noteworthy property of SUFs is that they represent multi-threshold decision rules. However, not all sets of multi-threshold rules can be represented by a single SUF. We showed how to represent any set of multi-threshold rules as a combination of SUFs. Moreover, we studied their potential advantages as a compact representation of large sets of rules, as well as an intermediary step for extracting rules from empirical datasets [51]. We also proposed a novel method [58] for learning sets of decision rules that optimally fit the training data and that favors short rules over long ones. This is a competitive alternative to other methods for monotonic classification as in [78].

7.2. Knowledge Discovery in Healthcare and Life Sciences

Participants: Alexandre Bazin, Miguel Couceiro, Adrien Coulet, Sébastien Da Silva, Florence Le Ber, Jean-François Mari, Pierre Monnin, Amedeo Napoli, Abdelkader Ouali, Yannick Toussaint.

7.2.1. Ontology-based Clustering of Biological Data

Biomedical objects can be characterized by ontology annotations. For example, Gene Ontology annotations provide information on the functions of genes, while Human Phenotype Ontology (HPO) annotations provide information about phenotypes associated with diseases. It is usual to consider such annotations in the analysis of biomedical data, most of the time annotations from only one single ontology. However, complex objects such as diseases can be annotated at the same time w.r.t. different ontologies, making clear distinct dimensions. We are investigating how annotations from several ontologies may be cooperating in disease classification. In particular, we classified Genetic Intellectual Disabilities, on the basis of their HPO annotations and of Gene Ontology annotations of genes known for being responsible for these diseases [88]. We used clustering algorithms based on semantic similarities that enable us to compare sets of annotations. In particular, this experiment illustrates the fact that considering several ontologies provides better results in clustering, while selecting the best set of ontologies to combine is depending on the dataset and on the classification task. This study is still going on.

7.2.2. Validation of Pharmacogenomic Knowledge

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

To carry out this validation, we define a minimal data schema for pharmacogenomic knowledge units (PGxO ontology), which is instantiated with data of different provenance (e.g. biomedical databases, literature and EHRs). The output of this instantiation is a (unique) knowledge graph called PGxLOD (https://pgxlod.loria.fr/). We defined and applied a set of so-called "reconciliation rules" that compare and align whenever possible knowledge units of different provenance [9]. The results of these rule applications are of particular interest since they highlight knowledge units defined in various data and knowledge sources. We are continuing this effort by studying how graph convolutional networks enable us to learn and then to compare the representation of n-ary relationships in the form of graph embeddings [39].

In addition, following our participation in the Biohackathon 2018 in Paris (https://2018.biohackathon-europe.org/), we firstly updated PGxLOD and improved its quality, completeness, and interconnection with other resources. Secondly we mined PGxLOD and searched for explanations about molecular mechanisms of adverse drug responses. Preliminary results where presented at the MedInfo Conference [59].

7.2.3. Mining Electronic Health Records

In the context of the Snowball Inria Associate Team, we studied the use of Electronic Health Records (EHRs) to predict at first prescription the need for a patient to be prescribed with a reduced drug dose [6]. We particularly focused on drugs whose dosage is known to be sensitive and variable. We used data from the Stanford Hospital to construct cohorts of patients that either did or did not need a dose change for each considered drug. After feature selection, we trained Random Forest models which successfully predict whether a new patient will or not require a dose change after being prescribed one of 23 drugs among 22 drug classes. Several of these drugs are related to clinical guidelines that recommend dose reduction exclusively in the case of adverse reaction. For these cases, a reduction in dosage may be considered as a surrogate for an adverse reaction, which our system could help to predict and to prevent.

In collaboration with Stanford University, we continued studying the development of predictive models from EHR data, in particular to evaluate the risk of atherosclerotic cardiovascular diseases (ASCVD). The evaluation of ASCVD risk is crucial for deciding upon the prescription of preventive therapies such as statins and others lipid lowering therapies. The prevalence of these diseases is depending on subgroups in a population, such as African-American and Asian people, which are indeed under-represented in cohorts that were used to fit the model currently used in clinics to evaluate the risk of ASCVD [25]. Due to such under-representation, biases are appearing in the evaluation of the risk when considering these different subgroups in the population. Then we proposed a method and a predictive model that controls, to some extent, the variability in the prediction of ASCVD when considering such "foreign" subgroups [40].

7.3. Knowledge Engineering and Web of Data

Participants: Nacira Abbas, Alexandre Bazin, Miguel Couceiro, Adrien Coulet, Florence Le Ber, Pierre Monnin, Amedeo Napoli, Justine Reynaud, Yannick Toussaint.

A first research topic in this axis relies on knowledge discovery in the web of data. This follows 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 and the quality of data and their potential to provide concept definitions in terms of necessary and sufficient conditions [73], [74]. We have proposed a novel technique based on Formal Concept Analysis which classifies subsets of RDF data into a concept lattice. This allows data exploration as well as the discovery of implication rules which are used to automatically detect possible completions of RDF data and to provide definitions. Experiments on the DBpedia knowledge base show that this kind of approach is well-founded and effective [41] [10]. In addition, it should be noticed that this research work also involves redescription mining, showing the potential complementarity between definition mining and redescription mining.

The second topic in this axis is related to dependencies [77]. In the relational database model, functional dependencies (FDs) indicate a functional relation between sets of attributes: the values of a set of attributes are determined by the values of another set of attributes. FDs can be generalized into relational dependencies, also known as "link keys" in the web of data [76]. For example, link keys may identify the same book or article in different bibliographical data sources, where a link key is a statement of the form: $\{\langle \texttt{auteur}, \texttt{creator}\rangle, \langle \texttt{titre}, \texttt{title}\rangle\}$ $linkkey$ $\langle \texttt{Livre}, \texttt{Book}\rangle$ stating that whenever an instance of the class `Livre` has the same values for properties `auteur` and `titre` as an instance of class `Book` has for properties `creator` and `title`, then they denote the same entity. Such link keys are more complex than FDs in databases in several respects and they raise new problems to solve [2].

One main objective of this research work is to follow the lines initiated in recent papers [29], and to extend to link keys the characterization of FDs and of Similarity Dependencies within FCA and pattern structures. Indeed, this is one of the objective of the ANR ELKER project. Accordingly, one purpose is to extend the initial proposals based on FCA and to provide adapted implementations. This is part of the thesis work of Nacira Abbas initiated at the end of 2018 [26]. Moreover, we are currently investigating possible connections with Relational Concept Analysis and redescription mining. We would like to study the formulation of the discovery of link keys in reusing and extending some construction heuristics that were developed in redescription mining. Actually, redescription mining is a data mining technique which aims at constructing pairs of descriptions, i.e., pairs of logical statements, one for each of two datasets, such that their support sets, i.e., the sets of objects that satisfy each statements of a pair, respectively, are most similar, as measured for example by their Jaccard index.

PESTO Project-Team

7. New Results

7.1. Security protocols

7.1.1. Analysis of Equivalence Properties

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

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

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

- The *Akiss* tool provides good expressivity as it supports a large number of cryptographic primitives (including the XOR primitive, extremely popular in low energy devices such as RFID tags) and protocols with else branches. It allows verification for a bounded number of protocol sessions. The tool is precise for a class of determinate processes, and can approximate equivalence for other protocols. The tool however suffers from efficiency problems when the number of sessions increases. The computation can be partially distributed on different cores. To overcome these efficiency problems of the *Akiss* tool, Gazeau and Kremer completely revisit the theory underlying *Akiss*. Rather than enumerating the possible traces, the new version directly reasons about partial ordered traces. A new implementation is also in progress and the first results seem extremely promising.
- The DEEPSEC tool is a recent tool that allows for user-defined cryptographic primitives that can be modelled as a subterm convergent rewrite system (slightly more restricted than AKISS), but supports the whole applied pi calculus, except for bounding the number of sessions. It is precise, in that it decides equivalence (without any approximations) and has good efficiency (slightly less than SAT-Equiv) for the class of determinate processes (where partial order reductions apply). To improve efficiency for non-determinate processes, Cheval, Kremer and Rakotonirina [21] develop new optimisation techniques. This is achieved through a new, stronger equivalence for which partial-order reductions are sound even for non-determinate processes, as well as new symmetry reductions. They demonstrate that these techniques provide a significant (several orders of magnitude) speed-up in practice, thus increasing the size of the protocols that can be analysed fully automatically. Even though the new equivalence is stronger, it is nevertheless coarse enough to avoid false attacks on most practical examples.
- The SAT-Equiv tool relies on a "small-attack property": if there is an attack against trace equivalence, then there is a well-typed attack, that is an attack where the messages follow some a priori given structure. This allows to dramatically reduce the search space. We have recently extended [11] this approach to a class of equational theory, that encompasses all standard cryptographic primitives (including e.g. randomized encryption) as well as theories that are less considered by automatic tools, such as threshold decryption. This result will allow to further extend the SAT-Equiv tool but can also be used more generally to characterize the form of an attack, independently of the considered tool.

From a more foundational point of view, in collaboration with Erbatur (LMU, Germany) and Marshall (Univ Mary Washington, USA), Ringeissen studies decision procedures for the intruder deduction and the static equivalence problems in combinations of subterm convergent rewrite systems and syntactic theories for which it is possible to apply a mutation principle to simplify equational proofs. As a continuation of a work initially presented at UNIF'18, it has been shown that a matching property is applicable to solve both intruder deduction and static equivalence. This matching property can be satisfied when using a matching algorithm known for syntactic theories [29]. A journal paper reporting this result is currently under review.

7.1.2. Decision Procedures for Equational Theories

Participants: Christophe Ringeissen, Michaël Rusinowitch.

Equational theories and unification procedures are widely used in protocol analyzers to model the capabilities of a (passive) intruder. In the context of protocol analysis, many equational theories of practical interest satisfy the finite variant property. This class of theories is indeed a class of syntactic theories admitting a terminating mutation-based unification algorithm. This mutation-based unification algorithm generalizes the syntactic unification algorithm known for the empty theory. In collaboration with Erbatur (LMU, Germany) and Marshall (Univ Mary Washington, USA), this particular unification algorithm has been applied by Ringeissen to get new non-disjoint combination results for the unification problem [23], [32].

In collaboration with Anantharaman (LIFO, Orléans), Hibbs (SUNY Albany & Google, USA), and Narendran (SUNY Albany, USA), Rusinowitch has studied the unification problem in list theories. Decision procedures for various list theories have been investigated in the literature with applications to automated verification. In [17], it has been shown that the unifiability problem for some list theories with a *reverse* operator is NP-complete. A unifiability algorithm is given for the case where the theories are extended with a *length* operator on lists.

Among theories with the finite variant property, the class of theories presented by subterm convergent rewrite systems is particularly remarkable because it satisfies in addition a locality property. For this class of theories, it is thus possible to get a satisfiability procedure based on a reduction to the empty theory via an instantiation with the finitely many terms occurring in the input problem. As an alternative to locality, Ringeissen has investigated a politeness property, in collaboration with Chocron (Insikt Intelligence, Spain) and Fontaine (Veridis project-team). This approach has led to new non-disjoint combination results for the satisfiability problem modulo data structure theories extended with some bridging functions such as the *length* operator on lists [10], [26].

7.1.3. Recast of ProVerif

Participants: Vincent Cheval, Véronique Cortier.

Motivated by the addition of global states in ProVerif, we have started a major revision of the popular tool ProVerif. This revision goes well beyond global states and is conducted in collaboration with Bruno Blanchet, the original and main developer of ProVerif. One of the first main changes is the addition of ProVerif of the notion of "lemmas" and "axioms" that can be added to either encode additional properties (axioms) or help ProVerif to prove the desired properties. It is indeed now possible to specify lemmas, that will significantly reduce the number of considered clauses in the saturation procedure of ProVerif. These lemmas should of course be proved themselves by ProVerif, possibly by induction thanks to a particular care of the order of literals in the saturation procedure. The new approach provides more flexibility in cases where ProVerif was not able to terminate or yield false attacks (e.g. in the presence of global states).

Moreover, even when ProVerif is able to prove security, the tool is suffering from efficiency issues when applied to complex industrial protocols (up to 1 month running time for the analysis of the NoiseExplorer protocol). One reason is the subsumption procedure: a clause shall not be added if it is subsumed by another one (that is, if there exists a more general clause). This is crucial to avoid running into non termination issues. We have started a major rewrite of the subsumption procedure, taking advantage of the recent progress in this domain, in the automated deduction area. Another reason is the translation of processes into Horn clauses: For each conditional in the process, ProVerif generates a Horn clause for each possible result of this conditional.

On complex protocols with many interleaved conditionals, ProVerif is faced with an exponential blowup in the number of generated clauses. We have improved the generation of Horn clauses by avoiding exploring branches that would directly be subsumed by other conditional branches. The first experimental results show significant speed-up on many examples: On average, ProVerif is now 5 to 10 times faster than its current release, with some examples peaking at 50 to 200 times speedup.

7.1.4. Verification of Protocols with Global States

Participants: Jannik Dreier, Lucca Hirschi.

The *TAMARIN* prover is a state-of-the-art verification tool for cryptographic protocols in the symbolic model. Dreier, in collaboration with Hirschi, Sasse (ETH Zurich), and Radomirovic (Dundee), improved the underlying theory and the tool to deal with an equational theory modeling XOR operations. Exclusive-or (XOR) operations are common in cryptographic protocols, in particular in RFID protocols and electronic payment protocols. Although there are numerous applications, due to the inherent complexity of faithful models of XOR, there is only limited tool support for the verification of cryptographic protocols using XOR. This makes *TAMARIN* the first tool to support simultaneously this large set of equational theories, protocols with global mutable state, an unbounded number of sessions, and complex security properties including observational equivalence. We demonstrated the effectiveness of our approach by analyzing several protocols that rely on XOR, in particular multiple RFID-protocols, where we can identify attacks as well as provide proofs. These results were presented at CSF'18, an extended version was accepted in the Journal of Computer Security [12].

7.1.5. Symbolic Methods in Computational Cryptography Proofs

Participants: Charlie Jacomme, Steve Kremer.

Code-based game-playing is a popular methodology for proving the security of cryptographic constructions and side-channel countermeasures. This methodology relies on treating cryptographic proofs as an instance of relational program verification (between probabilistic programs), and decomposing the latter into a series of elementary relational program verification steps. Barthe (MPI on Security and Privacy, Bochum), Grégoire (Inria SAM), Jacomme, Kremer and Strub (LIX, École Polytechnique) develop principled methods for proving such elementary steps for probabilistic programs that operate over finite fields and related algebraic structures. They focus on three essential properties: program equivalence, information flow, and uniformity. We give characterizations of these properties based on deducibility and other notions from symbolic cryptography. They use (sometimes improve) tools from symbolic cryptography to obtain decision procedures or sound proof methods for program equivalence, information flow, and uniformity. Finally, they evaluate their approach using examples drawn from provable security and from side-channel analysis - for the latter, they focus on the masking countermeasure against differential power analysis. A partial implementation of our approach is integrated in EasyCrypt, a proof assistant for provable security, and in MaskVerif, a fully automated prover for masked implementations. This work was presented at CSF [18].

7.1.6. Analysis of Deployed Protocols

Participants: Sergiu Bursuc, Lucca Hirschi, Steve Kremer.

7.1.6.1. New Privacy Threat on 3G, 4G, and Upcoming 5G AKA Protocols

Mobile communications are used by more than two-thirds of the world population who expect security and privacy guarantees. The 3rd Generation Partnership Project (3GPP) responsible for the worldwide standardization of mobile communication has designed and mandated the use of the AKA protocol to protect the subscribers' mobile services. Even though privacy was a requirement, numerous subscriber location attacks have been demonstrated against AKA, some of which have been fixed or mitigated in the enhanced AKA protocol designed for 5G.

We found and reported [9] a new privacy attack against all variants of the AKA protocol, including 5G AKA, that breaches subscriber privacy more severely than known location privacy attacks do. Our attack exploits a new logical vulnerability we uncovered that would require dedicated fixes. We demonstrate the practical feasibility of our attack using low cost and widely available setups. Finally we conduct a security analysis of the vulnerability and discuss countermeasures to remedy our attack.

Our attack has later been considered to be a *key issue in 5G [38]* by 3GPP [0]. Since then, various vendors[0] have proposed countermeasures, which are currently under discussion.

7.1.6.2. Contingent Payments

Bursuc and Kremer study protocols that rely on a public ledger infrastructure, concentrating on protocols for zero-knowledge contingent payment, whose security properties combine diverse notions of fairness and privacy. They argue that rigorous models are required for capturing the ledger semantics, the protocol-ledger interaction, the cryptographic primitives and, ultimately, the security properties one would like to achieve. Our focus is on a particular level of abstraction, where network messages are represented by a term algebra, protocol execution by state transition systems (e.g. multiset rewrite rules) and where the properties of interest can be analyzed with automated verification tools. They propose models for: (1) the rules guiding the ledger execution, taking the coin functionality of public ledgers such as Bitcoin as an example; (2) the security properties expected from ledger-based zero-knowledge contingent payment protocols; (3) two different security protocols that aim at achieving these properties relying on different ledger infrastructures; (4) reductions that allow simpler term algebras for homomorphic cryptographic schemes. Altogether, these models allow us to derive a first automated verification for ledger-based zero-knowledge contingent payment using the Tamarin prover. Furthermore, our models help in clarifying certain underlying assumptions, security and efficiency tradeoffs that should be taken into account when deploying protocols on the blockchain. This work was presented at ESORICS [20].

7.2. E-voting

7.2.1. Definitions for E-Voting

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

Existing formal (computational) definitions for privacy in electronic voting make the assumption that the bulletin board which collects the votes behaves honestly: the only ballots on the board are created by voters, all ballots are placed without tampering with them, and no ballots are ever removed. This strong assumption is difficult to enforce in practice and whenever it does not hold vote privacy can be broken. As a consequence, voting schemes are proved secure only against an honest voting server while they are designed and claimed to resist a dishonest one. We have proposed a framework for the analysis of electronic voting schemes in the presence of malicious bulletin boards. We identify a spectrum of notions where the adversary is allowed to tamper with the bulletin board in ways that reflect practical deployment and usage considerations. To clarify the security guarantees provided by the different notions we establish a relationship with simulation-based security with respect to a family of ideal functionalities. The ideal functionalities make clear the set of authorised attacker capabilities which makes it easier to understand and compare the associated levels of security. We then leverage this relationship to show that each distinct level of ballot privacy entails some distinct form of individual verifiability. As an application, we have studied three protocols of the literature (Helios, Belenios, and Civitas) and identified the different levels of privacy they offer. This work has appeared as a part of the PhD thesis [8], defended by Joseph Lallemand in November 2019.

Some modern e-voting systems take into account that the platform used for voting may be corrupted, e.g. infected by malware, yet aiming to ensure privacy and integrity of votes even in that case. Bursuc and Kremer, in collaboration with Dragan (Univ of Surrey) propose a new definition of vote privacy, formalized in the cryptographic model as a computational indistinguishability game. The definition captures both known and novel attacks against several voting schemes, and they propose a scheme that is provably secure in this setting. Moreover the proof is formalized and machine-checked in the EasyCrypt theorem prover [40]. This result has been presented at EuroS&P [19].

7.2.2. Design of E-Voting Protocols

Participants: Véronique Cortier, Jannik Dreier, Joseph Lallemand, Mathieu Turuani.

[0] 3rd Generation Partnership Project, responsible for the standardization of 3G, 4G, and 5G mobile networks
[0] Qualcomm, Gemalto, China Mobile, Mobile Thales Thales, Nokia Nokia, ZTE ZTE, and Huawei.

Most existing voting systems either assume trust in the voting device or in the voting server. Filipiak (Orange Labs), Lallemand, and Cortier proposed a novel Internet voting scheme, BeleniosVS, that achieves both privacy and verifiability against a dishonest voting server as well as a dishonest voting device. In particular, a voter does not leak her vote to her voting device and she can check that her ballot on the bulletin board does correspond to her intended vote. Additionally, our scheme guarantees receipt-freeness against an external adversary. A formal proof of privacy, receipt-freeness, and verifiability has been established using the tool ProVerif, covering a hundred cases of threat scenarios. Proving verifiability required the identification of a set of sufficient conditions, that can be handled by ProVerif [42]. This contribution is of independent interest. This work has been presented at CSF'19 [22].

As a part of a contract with Idemia, we are designing a novel electronic voting system tailored to their needs. The system is made for on-site elections, with the use of smart cards. However, the goal is that the trust should not be placed in one single part of the system, hence smart cards can not be trusted. One originality of the approach is the possibility to re-use existing techniques, in conjunction with the use of smart-cards and paper ballots. In this context, we have designed a novel audit technique [36], which can be seen as a variant to the "cast or audit" approach proposed by Josh Benaloh. One significant advantage of our solution is that voters now audit systematically their ballot (instead of choosing whether they should audit or not) and cast the audited ballot.

7.3. Online Social Networks

7.3.1. Privacy Protection in Social Networks

Participants: Bizhan Alipour, Abdessamad Imine, Michaël Rusinowitch.

Social media such as Facebook provides a new way to connect, interact and learn. Facebook allows users to share photos and express their feelings by using comments. However, Facebook users are vulnerable to attribute inference attacks where an attacker intends to guess private attributes (e.g., gender, age, political view) of target users through their online profiles and/or their vicinity (e.g., what their friends reveal). Given user-generated pictures on Facebook, we show in [16] how to launch gender inference attacks on their owners from pictures meta-data composed of: (i) alt-texts generated by Facebook to describe the content of pictures, and (ii) comments posted by friends, friends of friends or regular users. We assume these two meta-data are the only available information to the attacker. Evaluation results demonstrate that our attack technique can infer the gender with an accuracy of 84% by leveraging only alt-texts, 96% by using only comments, and 98% by combining alt-texts and comments. We compute a set of sensitive words that enable attackers to perform effective gender inference attacks. We show the adversary prediction accuracy is decreased by hiding these sensitive words. To the best of our knowledge, this is the first inference attack on Facebook that exploits comments and alt-texts solely. In subsequent work we have investigated the case where comments are reduced to Emojis.

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

Participants: Ahmad Abboud, Michaël Rusinowitch.

In a joint project with the Resist research group at Inria Nancy and Numeryx company, we are working on the design, implementation and evaluation of a double-mask technique for building compressed and verifiable filtering rules in Software Defined Networks with the possibility of distributing the workload processing among several packet filtering devices operating in parallel [33], [34].

RESIST Team

7. New Results

7.1. Monitoring

7.1.1. Encrypted Traffic Analysis

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

Nowadays, most of Web services are accessed through HTTPS. While preserving user privacy is important, it is also mandatory to monitor and detect specific users' actions, for instance, according to a security policy. Our paper [4] presents a solution to monitor HTTP/2 traffic over TLS. It highly differs from HTTP/1.1 over TLS traffic what makes existing monitoring techniques obsolete. Our solution, H2Classifier, aims at detecting if a user performs an action that has been previously defined over a monitored Web service, but without using any decryption. It is thus only based on passive traffic analysis and relies on random forest classifier. A challenge is to extract representative values of the loaded content associated to a Web page, which is actually customized based on the user action. Extensive evaluations with five top used Web services demonstrate the viability of our technique with an accuracy between 94% and 99%.

We were also interested by Internet of Things (IoT) as related devices become widely used and their control is often provided through a cloud-based web service that interacts with an IoT gateway, in particular for individual users and home automation. Therefore, we propose a technique demonstrating that is possible to infer private user information, i.e., actions performed, by considering a vantage point outside the end-user local IoT network. By learning the relationships between the user actions and the traffic sent by the web service to the gateway, we have been able to establish elementary signatures, one for each possible action, which can be then composed to discover compound actions in encrypted traffic. We evaluated the efficiency of our approach on one IoT gateway interacting with up to 16 IoT devices and showed that a passive attacker can infer user activities with an accuracy above 90%. This work has been published in [16] and is related to the H2020 SecureIoT project (section 9.3.1.2).

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

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

The Internet-of-Things has become a reality with numerous protocols, platforms and devices being developed and used to support the growing deployment of smart services. Providing new services requires the development of new functionalities, and the elaboration of complex systems that are naturally a source of potential threats. Real cases recently demonstrated that the IoT can be affected by naïve weaknesses. Therefore, security is of paramount importance.

In that context, we have proposed a process mining approach, that is capable to cope with a variety of devices and protocols, for supporting IoT predictive security [14]. We have described the underlying architecture and its components, and have formalized the different phases related to this solution, from the building of behavioral models to the detection of misbehaviors and potential attacks. The pre-processing identifies the states characterizing the IoT-based system, while process mining methods elaborate behavioral models that are compatible with the heterogeneity of protocols and devices [26]. These models are then exploited to analyze monitoring data at runtime and detect misbehaviors and potential attacks preventively. Based on a proof-of-concept prototype, we have quantified the detection performances, as well as the influence of time splitting and clustering techniques. The experimental results clearly show the benefits of our solution combining process mining and clustering techniques. As future work, we are interested in comparing it to other alternative learning techniques, as well as in evaluating to what extent the generated alerts can be exploited to drive the activation of counter-measures.

This work has been achieved in the context of the H2020 SecureIoT project (section 9.3.1.2).

7.1.3. Monitoring of Blockchains' Networking Infrastructure

Participants: Thibault Cholez [contact], Jean-Philippe Eisenbarth, Olivier Perrin.

With the raise of blockchains, their networking infrastructure becomes a critical asset as more and more money and services are made on top of them. However, they are largely undocumented and may be prone to performance issues and severe attacks so that the question of the resiliency of their overlay network arises. With regard to the state of the art on P2P networks security, the fact that a service infrastructure is distributed is not sufficient to assess its reliability, as many bias (for instance, if nodes are concentrated in a given geographical location) and attacks (eclipse, Sybil or partition attacks) are still possible and may severely disturb the network.

Overall, according to the scientific literature, the security provided by the proof of work consensus and the huge size of the main public blockchains seem to protect them well from large scale attacks (51% attack, selfish mining attack, etc.) whose cost to be successful becomes prohibitive and often exceeds the expected gain. However, rather than only focusing on the application level, an attacker could rather try to disturb the underlying P2P network to weaken the consensus in some specific parts of the blockchain network to gain advantage. Our current work uses a third-party crawler to get an accurate view of the Bitcoin overlay network. We are currently analyzing the data with graph theory metrics to identify possible anomalies or flaws that could be exploited by attackers.

7.1.4. Quality of Experience Monitoring

Participants: Isabelle Chrisment [contact], Antoine Chemardin, Frédéric Beck, Lakhdar Meftah [University of Lille], Romain Rouvoy [University of Lille].

We carried on our collaboration with the SPIRALS team (Inria/Université de Lille). Even though mobile crowdsourcing allows industrial and research communities to build realistic datasets, it can also be used to track participants' activity and to collect insightful reports from the environment (e.g., air quality, network quality). While data anonymization for mobile crowdsourcing is commonly achieved *a posteriori* on the server side, we have proposed a decentralized approach, named Fougere [19], which introduces an *a priori* data anonymization process. In order to validate our privacy preserving proposal, two testing frameworks (ANDROFLEET and PEERFLEET [20]) have been designed and implemented. They allows developers to automate reproducible testing of nearby peer-to-peer (P2P) communications.

In the context of both ANR BottleNet (section 9.2.1.1) and IPL BetterNet (section 9.2.5.1) projects, we continued to work on our open measurement platform for the quality of mobile Internet access (i.e., setup and manage the backend infrastructure for data collection and analysis). This platform is hosted by the High Security Laboratory [0] located at Inria Nancy Grand-Est. A collect campaign has been performed with a small set of volunteer users selected by the INSEAD-Sorbonne Université Behavioural Lab [0].

7.2. Experimentation

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

7.2.1. Grid'5000 Design and Evolutions

Participants: Benjamin Berard [SED], Luke Bertot, Alexandre Merlin, Lucas Nussbaum [contact], Nicolas Perrin, Patrice Ringot [SISR LORIA], Teddy Valette [SED].

The team was again heavily involved in the evolutions and the governance of the Grid'5000 testbed.
Technical team management. Since the beginning of 2017, Lucas Nussbaum serves as the *directeur technique* (CTO) of Grid'5000 in charge of managing the global technical team (10 FTE). He is also a member of the *Bureau* of the GIS Grid'5000.

[0] https://lhs.loria.fr
[0] https://www.insead.edu/centres/insead-sorbonne-universite-lab-en

SILECS project. We are also heavily involved in the ongoing SILECS project, that aims to create a new infrastructure on top of the foundations of Grid'5000 and FIT in order to meet the experimental research needs of the distributed computing and networking communities.
SLICES ESFRI proposal. At the European level, we are involved in a ESFRI proposal submission. We submitted a *Design Study* project in November 2019, and are in the final stages of submitting the ESFRI proposal itself in early 2020.
TILECS workshop. We participated in the organization of the TILECS workshop. TILECS (*Towards an Infrastructure for Large-Scale Experimental Computer Science*, https://www.silecs.net/tilecs-2019/) gathered about 80 members (mostly faculty) of the testbeds designers and users community in France, to discuss the future plans for research infrastructures in the networking and distributed computing fields. During that workshop, Lucas Nussbaum presented Grid'5000 [32].
Group storage. A technical contribution from the team is the addition of a *group storage* service that allows groups of users to share data, with improved security and performance compared to what was previously available.
Support for Debian 10. Another notable technical contribution from the team is the work of Teddy Valette on supporting Debian 10 in the set of Grid'5000 system environments made available to users.
New clusters available in Nancy: graffiti, gros, grue. Finally, the team was also heavily involved in the purchase and installation of several new clusters in the Nancy site, gathering funding from CPER LCHN, CPER Entreprises, MULTISPEECH team, LARSEN team. This greatly increases the resources available locally, both for GPUs (graffiti and grue), and for large-scale experiments (gros).

7.2.2. Involvement in the Fed4FIRE Testbeds Federation

Participants: Luke Bertot, Lucas Nussbaum [contact].

In the context of the Fed4FIRE+ project (section 9.3.1.1), Grid'5000 was officially added to the Fed4FIRE federation at the beginning of 2019. In 2019, we implemented on-demand *stitching* between Grid'5000 experiments and other testbeds of the federation (through VLANs provided by GEANT and RENATER), allowing experiments that combine resources from Grid'5000 and other testbeds [27]. We are also improving our implementation of an SFA Aggregate Manager in order to allow the use of Grid'5000 through Fed4FIRE tools, such as the jFed GUI.
We also worked on the issue of classifying and presenting the set of testbeds available in the federation. This was the subject of a presentation at the GEFI collaboration workshop [31].

7.2.3. I/O Emulation Support in Distem

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

We finished the work on adding I/O emulation support in Distem, in order to experiment how Big Data solution can handle degraded situations [22].

7.2.4. Distributing Connectivity Management in Cloud-Edge infrastructures

Participant: Lucas Nussbaum [contact].

In the context of David Espinel's PhD (CIFRE Orange, co-supervised with Adrien Lebre and Abdelhadi Chari), we worked on distributing connectivity management in Cloud-Edge infrastructures [38]. The classic approach of deploying large data centers to provide Cloud services is being challenged by the emerging needs of Internet of Things applications, Network Function Virtualization services or Mobile edge computing. A massively distributed Cloud-Edge architecture could better fit the requirements and constraints of these new trends by deploying on-demand Infrastructure as a Service in different locations of the Internet backbone (i.e, network point of presences). A key requirement in this context is the establishment of connectivity among several virtual infrastructure managers in charge of operating each site. In this work, we analyzed the requirements and challenges raised by the inter-site connectivity management in a Cloud-Edge infrastructure.

7.2.5. NDN Experimentation

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

While ICN is a promising technology, we currently lack experiments carrying real user traffic. This also highlights the difficulty of making the link between the new NDN world and the current IP world. To address this issue, we designed and implemented an HTTP/NDN gateway (composed of ingress and egress gateways) that can seamlessly transport the traffic of regular web users over an NDN island, making them benefit from the good properties of the protocol to deliver content (request mutualization, caching, etc.). The gateway itself is part of a wider architecture that aims to use NFV to deploy NDN and benefit from its orchestration capability to address performance and security issues inherent to new network architectures.

To validate the whole architecture, a testbed involving real users was made. The gateway was used by dozens of users for a few weeks to prove that running a NDN network over NFV is a viable solution to address the transition between both worlds. Users accessed many websites through the NDN network in a very satisfying way. The results have been published in IEEE Communications Magazine [5].

7.3. Analytics

7.3.1. CPS Security Analytics

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

During 2019, we evaluated a novel type of attack, named Measurement as Reference attack (MaR), on the cooperative control and communication layers in microgrids, where the attacker targets the communication links between distributed generators (DGs) and manipulates the reference voltage data exchanged by their controllers. We assessed its impact on reference voltage synchronization at the different control layers of a microgrid. Results and the development of an experimental platform are presented in [18] to demonstrate this attack, in particular the maximum voltage deviation and inaccurate reference voltage synchronization it causes in a microgrid. ML algorithms are also applied on the collected datasets from this platform for the detection of this attack.

7.3.2. Optimal and Verifiable Packet Filtering in Software-Defined Networks

Participants: Abdelkader Lahmadi [contact], Ahmad Abboud, Michael Rusinowitch [Pesto team], Miguel Couceiro [Orpailleur team], Adel Bouhoula [Numeryx].

Packet filtering is widely used in multiple networking appliances and applications, in particular, to block malicious traffic (protection of network infrastructures through firewalls and intrusion detection systems). It is also widely deployed on routers, switches and load balancers for packet classification. This mechanism relies on the packet's header fields to filter such traffic by using range rules of IP addresses or ports. However, the set of packet filters has to handle a growing number of connected nodes and many of them are compromised and used as sources of attacks. For instance, IP filter sets available in blacklists may reach several millions of entries, and may require large memory space for their storage in filtering appliances. In [40], [39], we proposed a new method based on a double mask IP prefix representation together with a linear transformation algorithm to build a minimized set of range rules. We have formally defined the double mask representation over range rules and proved that the number of required masks for any range is at most $2w-4$, where w is the length of a field. This representation makes the network more secure, reliable and easier to maintain and configure. We show empirically that the proposed method achieves an average compression ratio of 11% on real-life blacklists and up to 74% on synthetic range rule sets. Finally, we add support of double mask into a real SDN network.

7.3.3. Port Scans Analysis

Participants: Jérôme François [contact], Frederic Beck, Sofiane Lagraa [University of Luxembourg], Yutian Chen [Telecom Nancy], Laurent Evrard [University of Namur], Jean-Noël Colin [University of Namur].

TCP/UDP port scanning or sweeping is one of the most common technique used by attackers to discover accessible and potentially vulnerable hosts and applications. Although extracting and distinguishing different port scanning strategies is a challenging task, the identification of dependencies among probed ports is primordial for profiling attacker behaviors, with as a final goal to better mitigate them. In [6], we proposed an approach that allows us to track port scanning behavior patterns among multiple probed ports and identify intrinsic properties of observed group of ports. Our method is fully automated and based on graph modeling and data mining techniques including text mining. It provides to security analysts and operators relevant information about services that are jointly targeted by attackers. This is helpful to assess the strategy of the attacker, such that understanding the types of applications or environment she targets. We applied our method to data collected through a large Internet telescope (or Darknet).

In addition, we decided to leverage this knowledge for improving data analysis techniques applied to network traffic monitoring. Network traffic monitoring is primordial for network operations and management for many purposes such as Quality-of-Service or security. However, one major difficulty when dealing with network traffic data (packets, flows...) is the poor semantic of individual attributes (number of bytes, packets, IP addresses, protocol, TCP/UDP port number...). Many attributes can be represented as numerical values but cannot be mapped to a meaningful metric space. Most notably are application port numbers. They are numerical but comparing them as integers is meaningless. In [13], [12], we propose a fine grained attacker behavior-based network port similarity metric allowing traffic analysis to take into account semantic relations between port numbers. The behavior of attackers is derived from passive observation of a Darknet or telescope, aggregated in a graph model, from which a semantic dissimilarity function is defined. We demonstrated the veracity of this function with real world network data in order to pro-actively block 99% of TCP scans.

7.4. Orchestration

7.4.1. Mutualization of Monitoring Functions in Edge Computing

Participants: Jérôme François [contact], Mohamed Abderrahim [Orange Labs], Meryem Ouzzif [Orange Labs], Karine Guillouard [Orange Labs], Adrien Lebre [STACK Inria team, IMT Atlantique], Charles Prud'Homme [IMT Atlantique], Xavier Lorca [IMT Mines Albi, France].

By relying on small sized and massively distributed infrastructures, the edge computing paradigm aims at supporting the low latency and high bandwidth requirements of the next generation services that will leverage IoT devices (e.g., video cameras, sensors). To favor the advent of this paradigm, management services, similar to the ones that made the success of cloud computing platforms, should be proposed. However, they should be designed in order to cope with the limited capabilities of the resources that are located at the edge. In that sense, they should mitigate as much as possible their footprint. Among the different management services that need to be revisited, we investigated in [10] the monitoring one. Monitoring functions tend to become compute-, storage-and network-intensive, in particular because they will be used by a large part of applications that rely on real-time data. To reduce as much as possible the footprint of the whole monitoring service, we proposed to mutualize identical processing functions among different tenants while ensuring their quality-of-service (QoS) expectations. We formalized our approach as a constraint satisfaction problem and show through micro-benchmarks its relevance to mitigate compute and network footprints.

This work has been achieved in the context of the Inria-Orange joint lab (section 9.2.2.1).

7.4.2. Software-Defined Security for Clouds

Participants: Rémi Badonnel [contact], Olivier Festor, Maxime Compastié.

Cloud infrastructures provide new facilities to build elaborated added-value services by composing and configuring a large variety of computing resources, from virtualized hardware devices to software products. They are however further exposed to security attacks than traditional environments. We have pursued our efforts on a software-defined security strategy based on the TOSCA language, in order to support the protection of cloud resources using unikernel techniques [11]. This language enables the specification of cloud services and their orchestration. We have extended it to drive the integration and configuration of security mechanisms

within cloud resources, at the design and operation phases, according to different security levels. We rely on unikernel techniques to elaborate cloud resources using a minimal set of libraries, in order to reduce the attack surface. We have designed a framework to interpret this extended language and to generate and configure protected unikernel virtual machines, in accordance with contextual changes. The adaptation is typically performed through the regeneration of protected unikernel virtual machines in a dynamic manner. We have quantified the benefits and limits of this approach through extensive series of experiments. As future work, we are interested in investigating security issues specifically related to cloud resource migrations, and evaluating to what extent our hardening techniques can be complemented by security chains.

This word has been achieved in the context of the Inria-Orange joint lab (section 9.2.2.1).

7.4.3. Chaining of Security Functions

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

Software-defined networking offers new opportunities for protecting end users and their applications. It enables the elaboration of security chains that combines different security functions, such as firewalls, intrusion detection systems, and services for preventing data leakage. In that context, we have continued our efforts on the orchestration and verification of security chains, in collaboration with Stephan Merz from the VeriDis project-team at Inria Nancy, and concretized with the PhD defense of Nicolas Schnepf in September 2019 [3]. In particular, we have proposed this year an approach for automating the merging of security chains in software-defined networks [24]. This method complements the inference-based generation techniques that we proposed in [9]. The merging algorithms are designed to compose several security chains into a single one, in order to minimize the number of security functions and rules, while preserving the semantics of the initial chains. The algorithms have been implemented in Python and have been integrated into a proof-of-concept prototype that also contains the learning and inference components [23]. The performance of this implementation has been evaluated through extensive experiments. In particular, we have compared different approaches to merging security chains in terms of the complexity of the resulting chains, their accuracy, and the overhead incurred in computing the combined chains. The proposed solution is able to minimize the number of security functions and rules. It also facilitates the building of security chains at runtime, through a decoupling from the generation of individual chains.

7.4.4. Software-Defined Traffic Engineering to Absorb Influx of Network Traffic

Participants: Jérôme François [contact], Abdelkader Lahmadi, Romain Azais [MOSAIC team], Benoit Henry [IMT Lille Douai], Shihabur Chowdhury [University of Waterloo], Raouf Boutaba [University of Waterloo].

Existing shortest path-based routing in wide area networks or equal cost multi-path routing in data center networks do not consider the load on the links while taking routing decisions. As a consequence, an influx of network traffic stemming from events such as distributed link flooding attacks and data shuffle during large scale analytics can congest network links despite the network having sufficient capacity on alternate paths to absorb the traffic. This can have several negative consequences, service unavailability, delayed flow completion, packet losses, among others. In this regard and under the context of NetMSS associate team (section 9.4.1.1), we proposed SPONGE [15], a traffic engineering mechanism for handling sudden influx of network traffic. SPONGE models the network as a stochastic process, takes the switch queue occupancy and traffic rate as inputs, and leverages the multiple available paths in the network to route traffic in a way that minimizes the overall packet loss in the network. We demonstrated the practicality of SPONGE through an OpenFlow based implementation, where we periodically and pro-actively reroute network traffic to the routes computed by SPONGE. Mininet emulations using real network topologies show that SPONGE is capable of reducing packet drops by 20% on average even when the network is highly loaded because of an ongoing link flooding attack.

SEMAGRAMME Project-Team

6. New Results

6.1. Syntax-Semantics Interface

Participants: Philippe de Groote, Sylvain Pogodalla, William Babonnaud.

6.1.1. Abstract Categorial Grammars

We have worked on implementing parsing optimization to the Abstract Categorial Grammar tool kit. These optimizations are based on Datalog program rewriting techniques, in particular a general version of Magic Sets [27], [36]. Theses optimizations rely on the tree isomorphism between derivation trees resulting from parsing with a given abstract categorial grammar, and proofs of facts in a corresponding Datalog program. Because magic rewriting breaks the isomorphism, a transformation of proofs back to derivation trees has been proposed.

6.1.2. Lexical Semantics

The lexicon model underlying Montague semantics is an enumerative model that would assign a meaning to each atomic expression. This model does not exhibit any interesting strucuture. In particular, polysemy problems are considered as homonymy phenomena: a word has as many lexical entries as it has senses, and the semantic relations that might exist between the different meanings of a same word are ignored. To overcome these problems, models of generative lexicons have been proposed in the literature. Implementing these generative models in the realm of the typed λ-calculus necessitates a calculus with notions of subtyping and type coercion. In this context, we have investigated several ways of expressing coercion using record types, and intersection types. In addition, William Babonnaud has shown how the structure of a generative lexicon may be formalized in type theory, using the categorical notion of a topos [10].

6.2. Discourse Dynamics

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

6.2.1. Dynamic Logic

We have enriched our type-theoretic dynamic logic in several directions in order to take into account more dynamic phenomena. In particular, we have continued to study the dynamic properties of determiners in order to systematically capture their semantics by defining an appropriate notion of dynamic generalized quantifier. To this end, Clement Beysson has studied several issues raised by the modeling of plural determiners, which necessitates to introduce plural discourse referents that can be formalized as second-order bound variables.

6.2.2. Dialogue Modeling

Maxime Amblard and Maria Boritchev have developed a dynamic model of dialogue. We have focused on the relation between question and answers and on building a resource based on settlers of Catan game records (the DiNG corpus).

We presented in [12] research on a compositional treatment of questions in a neo-Davidsonian event semantics style. [28] presented a dynamic neo-Davidsonian compositional treatment of declarative sentences. Starting from complex formal examples, we enriched Champollion's framework with ways of handling phenomena specific to question-answer pair representation. Maria Boritchev gave two presentations on these issues [16], [21].

In [9], we presented a taxonomy of questions and answers based on real-life data extracted from spontaneous dialogue corpora. This classification allowed us to build a fine-grained annotation scheme, which we applied to several languages: English, French, Italian and Chinese. In [13], we presented an annotation scheme for classifying the content and discourse contribution of question-answer pairs. We proposed detailed guidelines for using the scheme and applied them to dialogues in English, Spanish, and Dutch. Finally, we have reported on initial machine learning experiments for automatic annotation.

In another direction, Maxime Amblard has started a common work with Chloé Braud on Formal and Statistical Modelling of dialogue. To this end, we have started with Chuyuan Li to design a dialogue model to structure the different necessary linguistic informations for interaction. This model will be implemented in a tool that finely manages interaction through formal and learning strategies.

6.2.3. Pathological Discourse Modelling

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

As part of the work carried out in the SLAM project, Maxime Amblard, Michel Musiol and Manuel Rebuschi (*Archives Henri-Poincaré, Université de Lorraine*) continue to work on modelling interactions with schizophrenic patients. We published an article about the corpus [20]. We are writing a book on these issues, in particular, we wrote a long introduction [19]. Maxime Amblard and Michel Musiol were awarded by an Inria Exploratory Action on this issues ODiM. This year we recruited the project's collaborators. In addition, we started the constitution of a new resource.

6.3. Common Basic Resources

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

6.3.1. Corpus Annotation

The Universal Dependencies project (UD) aims at building a syntactic dependency scheme which allows for similar analyses for several different languages. Bruno Guillaume and Guy Perrier are active in the UD community, and participate to the development and the improvement of the French data in this international initiative. Bruno Guillaume converted a new French treebank into UD: the French Question Bank (FQB), developped by Djamé Seddah and Marie Candito [35]. With the conversion system described in [2], the corpus UD_French-FQB was introduced in version 2.4 of UD in May 2019.

Bruno Guillaume, Marie-Catherine de Marneffe (Ohio State University, Columbus, Ohio, USA) and Guy Perrier improved the consistency of two French corpora annotated with the UD scheme [6]. They improved the annotations of the two French corpora to render them closer to the UD scheme, and evaluated the changes done to the corpora in terms of closeness to the UD scheme as well as of internal corpus consistency.

Bruno Guillaume and Guy Perrier developed and popularized the use of the GREW tool for various language applications and more particularly the pattern matching module Grew-match [22], [26], [17].

SUD is an annotation scheme for syntactic dependency treebanks, that is almost isomorphic to UD (Universal Dependencies). Contrary to UD, it is based on syntactic criteria (favoring functional heads) and the relations are defined on distributional and functional bases. In [14], Kim Gerdes (Sorbonne nouvelle, Paris 3), Bruno Guillaume, Sylvain Kahane (*Université Paris Nanterre*) and Guy Perrier recalled and specified the general principles underlying SUD, presented the updated set of SUD relations, discussed the central question of Multiword Expressions, and introduced an orthogonal layer of deep-syntactic features converted from the deep-syntactic part of the UD scheme.

6.3.2. FR-FraCas

Maxime Amblard, Clement Beysson, Philippe de Groote, Bruno Guillaume, Sylvain Pogodalla and Karën Fort carried on the development of FR-FraCas, a French version of the FraCas test suite [31] which is an inference test suite, in English, for evaluating the inferential competence of different NLP systems and semantic theories. There currently exists a multilingual version of the resource for Farsi, German, Greek, and Mandarin. Sémagramme completed the first translation into French of the test suite. The latter has been publicly released [0]. We also ran an experiment in order to test both the translation and the logical semantics underlying the problems of the test suite. The experiment was run with 18 French native speakers. Such an experiment provides a way of checking the hypotheses made by formal semanticists against the actual semantic capacity of speakers (in the present case, French speakers), and allows us to compare the results we obtained with the ones of similar experiments that have been conducted for other languages [30], [29].

[0] https://gitlab.inria.fr/semagramme-public-projects/resources/french-fracas

SPHINX Project-Team

7. New Results

7.1. Control, stabilization and optimization of heterogeneous systems

Participants: Rémi Buffe, Thomas Chambrion, Eloïse Comte, Arnab Roy, Takéo Takahashi, Jean-François Scheid, Julie Valein.

Control and optimization

The use of measures (instead of functions) as controls is usually referred to as "impulsive control". While the theory is now well understood for finite dimensional dynamics, many questions are still open for the control of PDEs. In [19], Thomas Chambrion and his co-authors discuss the notion of solution for the impulsive control (using measures instead of functions for the control) of the general bilinear Schrödinger equations. The results are adapted in [35] to the case of potentials with high regularity. These techniques have been used to extend the celebrated obstruction to controllability by Ball, Marsden and Slemrod to the case of abstract bilinear equations with bounded potentials [33] and the Klein-Gordon equation [20]. Other obstructions to controllability (preservation of regularity) have been investigated for the Gross-Pitaiewski equation with unbounded potentials [34].

In [15], an optimal control problem for groundwater pollution due to agricultural activities is considered, the objective being the optimization of the trade-off between the fertilizer use and the cleaning costs. The spread of the pollution is modeled by a convection-diffusion-reaction equation. We are interested in the buffer zone around the captation well and we determine its optimal size.

In [44], Eduardo Cerpa, Emmanuelle Crépeau and Julie Valein study the boundary controllability of the Korteweg-de Vries equation on a tree-shaped network, with less controls than equations.

In [27], Jérôme Lohéac and Takéo Takahashi study the locomotion of a ciliated microorganism in a viscous incompressible fluid. They use the Blake ciliated model: the swimmer is a rigid body with tangential displacements at its boundary that allow it to propel in a Stokes fluid. This can be seen as a control problem: using periodical displacements, is it possible to reach a given position and a given orientation? They are interested in the minimal dimension d of the space of controls that allows the microorganism to swim. Their main result states the exact controllability with $d = 3$ generically with respect to the shape of the swimmer and with respect to the vector fields generating the tangential displacements. The proof is based on analyticity results and on the study of the particular case of a spheroidal swimmer.

In [31], Arnab Roy and Takéo Takahashi study the controllability of a fluid-structure interaction system. They consider a viscous and incompressible fluid modeled by the Boussinesq system and the structure is a rigid body with arbitrary shape which satisfies Newton's laws of motion. They assume that the motion of this system is bidimensional in space. They prove the local null controllability for the velocity and temperature of the fluid and for the position and velocity of the rigid body for a control acting only on the temperature equation on a fixed subset of the fluid domain.

Rémi Buffe and Ludovick Gagnon consider N manifolds without boundary that intersect each other. They assume that the speed of propagation on each manifold is different, which implies that the Snell conditions applies at the interface. They give sufficient geometric conditions to ensure the controllability with distributed controls on $N - 1$ manifolds.

Stabilization

In [17], Lucie Baudouin, Emmanuelle Crépeau and Julie Valein study the exponential stability of the nonlinear Korteweg-de Vries equation with boundary time-delay feedback. Two different methods are employed: a Lyapunov functional approach (allowing to have an estimation on the decay rate, but with a restrictive assumption on the length of the spatial domain of the KdV equation) and an observability inequality approach, with a contradiction argument (for any non-critical lengths but without estimation on the decay rate).

In [55], Julie Valein shows the semi-global exponential stability of the nonlinear Korteweg-de Vries equation in the presence of a delayed internal feedback, for any lengths, in the case where the weight of the feedback with delay is smaller than the weight of the feedback without delay. In the case where the support of the feedback without delay is not included in the support of the feedback with delay, a local exponential stability result is proved if the weight of the delayed feedback is small enough.

Optimization
J.F. Scheid, V. Calesti (PhD Student) and I. Lucardesi study an optimal shape problem for an elastic structure immersed in a viscous incompressible fluid. They want to establish the existence of an optimal elastic domain associated with an energy-type functional for a Stokes-Elasticity system. We want to find an optimal reference domain (the domain before deformation) for the elasticity problem that minimizes an energy-type functional. This problem is concerned with 2D geometry and is an extension of the work of [113] for a 1D problem. The optimal domain is seeking in a class of admissible open sets defined with a diffeomorphism of a given domain. The main difficulty lies on the coupling between the Stokes problem written in a eulerian frame and the linear elasticity problem written in a lagrangian form. The shape derivative of the energy-type functional is also aimed to be determined in order to numerically obtain an optimal elastic domain. This work is in progress.

7.2. Direct and Inverse problems for heterogeneous systems

Participants: Rémi Buffe, Imene Djebour, David Dos Santos Ferreira, Ludovick Gagnon, Alexandre Munnier, Julien Lequeurre, Karim Ramdani, Takéo Takahashi, Jean-Claude Vivalda.

Direct problems
In [22], Imene Djebour and Takéo Takahashi consider a fluid–structure interaction system composed by a three-dimensional viscous incompressible fluid and an elastic plate located on the upper part of the fluid boundary. They use here Navier-slip boundary conditions instead of the standard no-slip boundary conditions. The main results are the local in time existence and uniqueness of strong solutions of the corresponding system and the global in time existence and uniqueness of strong solutions for small data and if one assumes the presence of frictions in the boundary conditions.

In [42], Mehdi Badra (University of Toulouse) and Takéo Takahashi analyze a bi-dimensional fluid-structure interaction system composed by a viscous incompressible fluid and a beam located at the boundary of the fluid domain. The main result is the existence and uniqueness of strong solutions for the corresponding coupled system. The proof is based on a the study of the linearized system and a fixed point procedure. In particular, they show that the linearized system can be written with a Gevrey class semigroup. The main novelty with respect to previous results is that they do not consider any approximation in the beam equation.

In [18], Muriel Boulakia (Sorbonne University), Sergio Guerrero (Sorbonne University) and Takéo Takahashi consider a system modeling the interaction between a viscous incompressible fluid and an elastic structure. The fluid motion is represented by the classical Navier–Stokes equations while the elastic displacement is described by the linearized elasticity equation. The elastic structure is immersed in the fluid and the whole system is confined into a general bounded smooth three-dimensional domain. The main result is the local in time existence and uniqueness of a strong solution of the corresponding system.

In [28], Debayan Maity (TIFR Bangalore), Jorge San Martin (University of Chile), Takéo Takahashi and Marius Tucsnak (University of Bordeaux) study the interaction of surface water waves with a floating solid constraint to move only in the vertical direction. They propose a new model for this interaction, taking into consideration the viscosity of the fluid. This is done supposing that the flow obeys a shallow water regime (modeled by the viscous Saint-Venant equations in one space dimension) and using a Hamiltonian formalism. Another contribution of this work is establishing the well-posedness of the obtained PDEs/ODEs system in function spaces similar to the standard ones for strong solutions of viscous shallow water equations. Their well-posedness results are local in time for any initial data and global in time if the initial data are close (in appropriate norms) to an equilibrium state. Moreover, they show that the linearization of the system around an equilibrium state can be described, at least for some initial data, by an integro-fractional differential equation related to the classical Cummins equation and which reduces to the Cummins equation when the

viscosity vanishes and the fluid is supposed to fill the whole space. Finally, they describe some numerical tests, performed on the original nonlinear system, which illustrate the return to equilibrium and the influence of the viscosity coefficient.

In [30], Benjamin Obando and Takéo Takahashi consider the motion of a rigid body in a viscoplastic material. This material is modeled by the 3D Bingham equations, and the Newton laws govern the displacement of the rigid body. The main result is the existence of a weak solution for the corresponding system. The weak formulation is an inequality (due to the plasticity of the fluid), and it involves a free boundary (due to the motion of the rigid body). They approximate it by regularizing the convex terms in the Bingham fluid and by using a penalty method to take into account the presence of the rigid body.

In [23], Alexandre Munnier and his co-authors consider the dynamics of several rigid bodies immersed in a perfect incompressible fluid. We show that this dynamics can be modelized by a second order ODE whose coefficients depend on the vorticity and the circulation of the fluid around the bodies. This formulation permits to point out the geodesic nature of the solutions, the added mass effect, the gyroscopic effects and the Kutta-Joukowski-type lift forces.

In [24], Julien Lequeurre and his co-authors study an unsteady nonlinear fluid–structure interaction problem. We consider a Newtonian incompressible two-dimensional flow described by the Navier-Stokes equations set in an unknown domain depending on the displacement of a structure, which itself satisfies a linear wave equation or a linear beam equation. The fluid and the structure systems are coupled via interface conditions prescribing the continuity of the velocities at the fluid–structure interface and the action-reaction principle. We prove existence of a unique local in time strong solution. In the case of the wave equation or a beam equation with inertia of rotation, this is, to our knowledge the first result of existence of strong solutions for which no viscosity is added. One key point, is to use the fluid dissipation to control, in appropriate function spaces, the structure velocity.

J.F. Scheid and M. Bouguezzi (PhD student) in collaboration with D. Hilhorst and Y. Miyamoto work on the convergence of the solution of the one-phase Stefan problem in one-space dimension to a self-similar profile. The evolutional self-similar profile is viewed as a stationary solution of a Stefan problem written in a self-similar coordinates system. The proof of the convergence relies on the construction of sub and super-solutions for which it must be proved that they both tend to the same function. It remains to show that this limiting function actually corresponds to the self-similar solution of the original Stefan problem. This work is in progress.

Rémi Buffe, Ludovick Gagnon *et al.* obtain the exponential decay of the solutions of coupled wave equations with a transmission condition at the interface and with a viscoelastic damping term. They prove that the exponential decay is obtained if the support of the viscoelastic term satisfies the uniform escaping geometry condition. They also deal with the case where the damping term touches the interface.

Inverse problems

In [43], the authors are interested in the homogenization of time-harmonic Maxwell's equations in a composite medium with periodically distributed small inclusions of a negative material. Here a negative material is a material modelled by negative permittivity and permeability. Due to the sign-changing coefficients in the equations, it is not straightforward to obtain uniform energy estimates to apply the usual homogenization techniques. The analysis is based on a precise study of two associated scalar problems: one involving the sign-changing permittivity with Dirichlet boundary conditions, another involving the sign-changing permeability with Neumann boundary conditions. For both problems, we obtain a criterion on the physical parameters ensuring uniform invertibility of the corresponding operators as the size of the inclusions tends to zero. Then we use the results obtained for the scalar problems to derive uniform energy estimates for Maxwell's system.

In [37], Jean-Claude Vivalda and his co-authors prove that the class of continuous-time systems who are strongly differentially observable after time sampling is everywhere dense in the set of pairs (f, h) where f is a (parametrized) vector field given on a compact manifold and h is an observation function.

In [47], using a partial boundary measurement, Jean-Claude Vivalda and his co-authors design an observer for a system that models a desalination device; this observer being used to make an output tracking trajectory.

Rémi Buffe, David Dos Santos Ferreira and Ludovick Gagnon obtain an estimate on the magnetic Laplacian with sharp dependance on the power of the zeroth and first order potential and close to sharp norm of these potentials. This estimate is related to the observability inequality for the wave equation and to the cost of the control.

7.3. Numerical analysis and simulation of heterogeneous systems

Participant: Xavier Antoine.

Acoustics

Artificial boundary conditions: while high-order absorbing boundary conditions (HABCs) are accurate for smooth fictitious boundaries, the precision of the solution drops in the presence of corners if no specific treatment is applied. In [29], the authors present and analyze two strategies to preserve the accuracy of Padé-type HABCs at corners: first by using compatibility relations (derived for right angle corners) and second by regularizing the boundary at the corner. Exhaustive numerical results for two- and three-dimensional problems are reported in the paper. They show that using the compatibility relations is optimal for domains with right angles. For the other cases, the error still remains acceptable, but depends on the choice of the corner treatment according to the angle.

Domain decomposition : in [49], Xavier Antoine and his co-authors develop the first application of the optimized Schwarz domain decomposition method to aeroacoustics. Highly accurate three-dimensional simulations for turbofans are conducted through a collaboration with Siemens (ongoing CIFRE Ph.D. Thesis of Philippe Marchner). In [26], the authors propose a new high precision IGA B-Spline approximation of the high frequency scattering Helmholtz problem, which minimizes the numerical pollution effects that affect standard Galerkin finite element approaches.

Underwater acoustics

New adiabatic pseudo-differential models as well as their numerical approximation are introduced in [53] for the simulation of the propagation of wave fields in underwater acoustics. In particular, the calculation of gallery modes is shown to be accurately obtained. This work is related to a new collaboration with P. Petrov from the V.I. Il'ichev Pacific Oceanological Institute, Vladivostok, Russia.

Quantum theory

With E. Lorin, Xavier Antoine proposes in [13] an optimization technique of the convergence rate of relaxation Schwarz domain decomposition methods for the Schrödinger equation. This analysis is based on the use of microlocal analysis tools. Convergence proofs are given in [11] for the real-time Schrödinger equation with optimized transmission conditions. We extend these results to the case of multiple subdomains in [13].

In [52], the authors analyze the convergence and stability in of a discretization scheme for the linear Schrödinger equation with artificial boundary conditions.

In [39], Xavier Antoine and his co-authors develop an implementation of the PML technique in the framework of Fourier pseudo-spectral approximation schemes for the fast rotating Gross-Pitaevskii equation. This is the first work related to the international Inria team BEC2HPC, associated with China (https://team.inria.fr/bec2hpc/).

In [12], Antoine and his co-authors develop new Fourier pseudo-spectral schemes including a PML for the dynamics of the Dirac equation. The implementation of the method leads to the possibility of simulating complex quantum situations. In [38], the authors extend the approximation to the curved static Dirac equation. The goal is to be able to better understand quantum phenomena related to the charge carriers in strained graphene, with potential long term applications for designing quantum computers. This is a collaboration with E. Lorin (Carleton University), F. Fillion-Gourdeau and S. Mac Lean from the Institute for Quantum Computing, University of Waterloo.

Fractional PDE

In [32], with J. Zhang and D. Li, Xavier Antoine is interested in the development and analysis of fast second-order schemes to simulate the nonlinear time fractional Schrödinger equation in unbounded domains.

The authors propose in [14] the construction of PML operators for a large class of space fractional PDEs in one- and two-dimensions. The specific case of the fractional laplacian is carefully considered.

Xavier Antoine and Emmanuel Lorin are interested in [40] in the problem of building fast and robust linear algebra algorithms based on the discretization of the Cauchy integral formula used to represent the power matrix. Applications related to stationary PDEs are presented, with possibly randomly perturbed potentials. Differential doubly preconditioned iterative schemes are investigated in details in [41] to evaluate the power, and more generally functions, of matrices.

Error estimates of a semi-implicit ALE scheme for the one-phase Stefan problem. J.F. Scheid, M. Bouguezzi (PhD student) and D. Hilhorst study the convergence with error estimates of an Arbitrary-Lagrangian-Eulerian (ALE) scheme for the classical one-phase Stefan problem. Despite Stefan problems as well as ALE techniques are well-known in the mathematical litterature for many decades, surprisingly there is no global result on convergence (with error estimates) for fully space-time discretized scheme based on ALE formulations. The main difficulty lies on the unbounded behavior of the exact (and approximate) free boundary. Stability results have already been obtained for a time-discretized scheme (and continuous in space) for the one-space dimension case.

Chaotic advection in a viscous fluid under an electromagnetic field. J.F. Scheid, J.P. Brancher and J. Fontchastagner study the chaotic behavior of trajectories of a dynamical system arising from a coupling system beetwen Stokes flow and an electromagnetic field. They consider an electrically conductive viscous fluid crossed by a uniform electric current. The fluid is subjected to a magnetic field induced by the presence of a set of magnets. The resulting electromagnetic force acts on the conductive fluid and generates a flow in the fluid. According to a specific arrangement of the magnets surrounding the fluid, vortices can be generated and the trajectories of the dynamical system associated to the stationary velocity field in the fluid may have chaotic behavior. The aim of this study is to numerically show the chaotic behavior of the flow for the proposed disposition of the magnets along the container of the fluid. The flow in the fluid is governed by the Stokes equations with the Laplace force induced by the electric current and the magnetic field. An article is in preparation.

TONUS Project-Team

7. New Results

7.1. Relaxation method for Guiding-Center equation

Participants: E. Franck, R. Helie, L. Navoret, P. Helluy.
In previous years, implicit kinetic relaxation methods have been developed to treat conservation laws without CFL and without a non-trivial matrix to reverse [6]-[4]. We have started to apply this method with a spectral discretization for transport equations such as the guiding-center equation (a non constant advection equation coupled with elliptic problem used in plasma physics). The scheme obtained has a very high order of convergence for an instability test case and is very simple to implement. We have also investigated the different kinetic relaxation representations. However, they suffer from inaccuracy at the boundaries. We have proposed a new approach in 1D [8] to analyse this behaviour and a new way to apply boundary conditions to ensure they are compatible both with the approximated system and its kinetic approximation. Extending this approach to higher dimensions is one of the objectives of the thesis of Romane Helie.

7.2. Relaxation method for transport in Tokamak

Participants: M. Boileau, P. Helluy, B. Bramas (Inria Camus).
To apply the relaxation method in a Tokamak context, we have developed a code called Chukrut (in Schnaps) that can handle kinetic relaxation models in Tokamak geometry [15]-[17]-[13]. In the poloidal direction the code uses an unstructured Discontinuous Galerkin solver which solves the transport equation (the main ingredient of the kinetic relaxation method) by using the scheduling graph linked to the upwind scheme. In the toroidal direction we use an exact solver on uniform grids (which will be replaced by a semi-Lagrangian solver). The algorithm is parallelised in the poloidal plane by a task-based OpenMP implementation and in the toroidal direction by MPI parallelism.

7.3. Relaxation method for Euler/MHD in low-Mach regime

Participants: E. Franck, L. Navoret, F. Bouchut (Marne la Vallée university).
Previously, we have proposed implicit relaxation methods for fluid models that allow us to reverse a simple system. However, previous methods [5] were not very effective in the multi-scale regimes of interest. We therefore proposed a semi-implicit scheme based on a dynamic splitting and a relaxation of fast waves only. The scheme was first applied to the Euler equations in low Mach regime. The scheme is stable and accurate regardless of the Mach number. We have successfully applied the method for the equilibria of the Shallow Water equations. Since last summer we have begun the extension for the 1D MHD with and without dispersive effects. The first results show that we obtain a similar method compared to the Euler case with acceptable stability conditions as for the Euler equations.

7.4. Reduced model for the Scrape-Off Layer

Participants: L. Navoret, M. Mehrenberger, P. Ghendrih (CEA Cadarache)
In this work, we consider a one-dimensional model for describing the two-species plasma dynamics in the scrape-off layer. This region is defined as the transition between the core of the plasma and the edge and is located around the first non-closed magnetic field line. The electron and ion distribution functions satisfy a Vlasov-Poisson system with source and absorption terms and a non-homogeneous equilibrium is expected to develop. A high-order semi-Lagrangian scheme has been implemented to correctly capture such a dynamics.

7.5. Recurrence phenomenon for finite element grid based Vlasov solver

Participants: L. Navoret, M. Mehrenberger, N. Pham

We have improved our previous (last year) result concerning the recurrence phenomenon by providing a complete proof of the asymptotic behaviour of the correlation function. Indeed, we prove that, in the fine grid limit, the correlation function of the density exactly concentrates at multiple times of the recurrence time. This thus fully confirms the fact the amplitude of the recurrence phenomenon is actually linked to the spectral accuracy of the velocity quadrature when computing the charge density at least for the linear transport equation.

7.6. Machine learning techniques for reduced model and stabilization

Participants: E. Franck, L. Navoret, V. Vigon (IRMA Strasbourg).
Just recently, we have begun to work on applications of machine learning techniques for the plasma simulation. This preliminary work is in the context of "Action exploratoire MALESI" and will really begin in 2020. The first point is the construction of a new closure for the fluid models using kinetic simulation as data. We have constructed 1D solvers for the Vlasov-Poisson equation with collisional operator and Compressible Navier-Stokes Poisson models. Comparing the models we observe that the classical Navier-Stokes closure is not sufficient when the Knudsen number is larger than 0.3-0.4. Currently we generate data using the Vlasov-Poisson code to train a neural-network for the closure. The second point is about the stabilization of the numerical method using CNN. We began a study to construct a Convulational Neural Network (CNN) to detect the Gibbs oscillations in the fluid simulations.

7.7. Asymptotic Preserving scheme for Vlasov-Maxwell to MHD

Participants: E. Franck, A. Crestetto (Nantes university), M. Badsi (Nantes university).
The MHD equation can be obtained by taking the limit of different small parameters of the bi-species Vlasov-Maxwell equations. Obtaining an "asymptotic preserving" scheme for the Vlasov equation (cost independent of the small parameters) is an important goal. Indeed, this type of scheme would allows us to construct coupling methods between MHD and Vlasov equations or to make simulations in various regimes to construct closures with data (see the previous point). During this year we have written a scheme able to treat the "quasi-neutral" and "mass-less" limits between the two-species Euler-Maxwell equations and the MHD model. The scheme is partially validated. We will finish the validation and add the collisional limit between Vlasov-Maxwell and Euler-Maxwell equations.

7.8. Optimal control for population dynamics

Participants: Y. Privat, L. Almeida (Sorbonne University), M. Duprez (Dauphine University) and N. Vauchelet (Paris 13 University).
Particular attention is being paid to the transmission of dengue fever, an arbovirus transmitted to humans by mosquitoes [3]-[2]. There is no vaccine to immunize a population. It has been observed that when a mosquito population was infected with the Wolbachia virus, they stopped transmitting the disease. In addition, the virus is transmitted from mother to child and is characterized by cytoplasmic incompatibility (no possible crosses between infected males and healthy females). On the other hand, infected mosquitoes have a reduced lifespan and fertility. Mathematically, this situation can be modelled (in a simplified way) using a controlled reaction-diffusion system. The control term represents the strategy of releasing (time-space) mosquitoes infected by Wolbachia. The practical questions that arise and that we wish to address are:

- how to carry out these releases to ensure the invasion?
- how to optimize the domain and form of releases?

Preliminary work has made it possible to determine a plausible temporal control strategy.

7.9. Observability for wave equation and high frequency behavior

Participants: Y. Privat, E. Humbert (Tours University) et E. Trélat (Sorbonne University).

We have determined the asymptotic in time of the observability constant in closed manifolds. In particular, we have proved that this limit can be represented as the minimum of two quantities: one purely spectral and another called the geometric quantity representing the limit of the average time spent by geodesics within the observation domain [19]-[19].

7.10. Developement of a Python library for tomography diagnostics

Participants: L. Mendoza

In the tofu code, a big component of both the direct and inverse solvers is the integration module. During the 2019 project it was developed and accelerated. Special attention was brought to memory optimization. Core functions for the inversions routines were developed and parallelized using OpenMP. The number of users and developers of the library has significantly increased in the last year (collaborators in CEA cadarache, ITER, CEA saclay, IPP Garching, etc.) so one of the main objectives was to better the continuous integration and documentation of the code: more unitary and simulation tests have been implemented, an online Web site with the documentation has been added, the library can be used on more platforms (windows, mac os x, and linux), and more fusion devices are now available (West, ITER, JET, etc.).

7.11. Finite volume methods for complex hyperbolic systems

Participants: P. Helluy, L. Quibel (EDF)

This year we have developed a Lattice Boltzmann scheme able to treat really complex and tabuled EOS (Equation Of State) for compressible multiphase flows (two and three phases). This new scheme have been implemented in the PyOpenCL Patapon. Additionally, in order to perform realistic simulations of such situations, we have also proposed a code based on a model that can handle both the thermodynamical disequilibrium between liquid and vapor and complex equations. This code is based on a relaxation scheme which is the best compromise between accuracy and stability.

7.12. The study of domain walls in micromagnetism

Participants: C. Courtès, R. Côte (IRMA)

A ferromagnetic material consists of a succession of isolated subdomains (known as the magnetic domains) in which the magnetic moments are aligned and point in the same direction. The interface separating two magnetic domains is called the domain wall and corresponds to a localized area where the direction of the magnetization suddenly changes. Mathematically, those domain walls correspond to the minimizers of the well-known micromagnatics energy. The magnetic behavior of ferromagnetic materials is due to the arrangement of the magnetic domains and to the dynamics of their domains walls. That dynamic is governed by the nonlinear Landau-Lifshitz-Gilbert equation. We study numerically and theoretically the stability and the interaction of two domain walls. Depending on the initial topological configuration, two domain walls may collide to give rise to a persistent profile or annihilate both, which results in aligning all magnetization vectors of the nanowire in the same direction.

7.13. Maxwell solvers

Participants: P. Helluy, M. Houillon

In collaboration with the AxesSim company, we continue the development of our CLAC software devoted to electromagnetic simulations in biological environment. We have implemented a new wire model. We have also run computations on the new CNRS supercomputer: Jean Zay. We now routinely launch simulations on 64 V100 GPUs in parallel for performing parameter studies of various antennas near to the human body (we can for instance vary the humidity level of the skin).

TOSCA Team

6. New Results

6.1. Probabilistic numerical methods, stochastic modeling and applications

Participants: Sofia Allende Contador, Alexis Anagnostakis, Mireille Bossy, Lorenzo Campana, Nicolas Champagnat, Quentin Cormier, Madalina Deaconu, Aurore Dupre, Coralie Fritsch, Vincent Hass, Pascal Helson, Christophe Henry, Ulysse Herbach, Igor Honore, Antoine Lejay, Rodolphe Loubaton, Radu Maftei, Kerlyns Martinez Rodriguez, Victor Martin Lac, Hector Olivero-Quinteros, Édouard Strickler, Denis Talay, Etienne Tanré, Denis Villemonais.

6.1.1. Published works and preprints

- H. AlRachid (Orléans University), M. Bossy, C. Ricci (University of Florence) and L. Szpruch (University of Edinburgh and The Alan Turing Institute, London) introduced several new particle representations for *ergodic* McKean-Vlasov SDEs. They construct new algorithms by leveraging recent progress in weak convergence analysis of interacting particle system. In [12] they present detailed analysis of errors and associated costs of various estimators, highlighting key differences between long-time simulations of linear (classical SDEs) versus non-linear (McKean-Vlasov SDEs) process.
- M. Di Iorio (Marine Energy Research and Innovation Center, Santiago, Chile), M. Bossy, C. Mokrani (Marine Energy Research and Innovation Center, Santiago, Chile), and A. Rousseau (LEMON team) obtained advances in stochastic Lagrangian approaches for the simulation of hydrokinetic turbines immersed in complex topography [42].
- M. Bossy, J.-F. Jabir (University of Edinburgh) and K. Martinez (University of Valparaiso) consider the problem of the approximation of the solution of a one-dimensional SDE with non-globally Lipschitz drift and diffusion coefficients behaving as x^α, with $\alpha > 1$ [44]. They propose an (semi-explicit) exponential-Euler scheme and study its convergence through its weak approximation error. To this aim, they analyze the $C^{1,4}$ regularity of the solution of the associated backward Kolmogorov PDE using its Feynman-Kac representation and the flow derivative of the involved processes. From this, under some suitable hypotheses on the parameters of the model ensuring the control of its positive moments, they recover a rate of weak convergence of order one for the proposed exponential Euler scheme. Numerical experiments are analyzed in order to complement their theoretical result.
- L. Campana et al. developed some Lagrangian stochastic model for anisotropic particles in turbulent flow [35]. Suspension of anisotropic particles can be found in various industrial applications. Microscopic ellipsoidal bodies suspended in a turbulent fluid flow rotate in response to the velocity gradient of the flow. Understanding their orientation is important since it can affect the optical or rheological properties of the suspension (e.g. polymeric fluids). The equations of motion for the orientation of microscopic ellipsoidal particles was obtained by Jeffery. But so far this description has been always investigated in the framework of direct numerical simulations (DNS) and experimental measurements. In this work, the orientation dynamics of rod-like tracer particles, i.e. long ellipsoidal particles (in the limit of infinite aspect-ratio) is studied. The size of the rod is assumed smaller than the Kolmogorov length scale but sufficiently large that its Brownian motion need not be considered. As a result, the local flow around a particle can be considered as inertia-free and Stokes flow solutions can be used to relate particle rotational dynamics to the local velocity gradient. The orientation of rod can be described as the normalised solution of the linear ordinary differential equation for the separation vector between two fluid tracers, under the action of the velocity gradient tensor. In this framework, the rod orientation is described by a Lagrangian stochastic model where cumulative velocity gradient fluctuations are represented by a white-noise tensor such that the incompressibility condition is preserved. A numerical scheme based on the decomposition into skew/symmetric part of the process dynamics is proposed.

- Together with M. Andrade-Restrepo (Univ. Paris Diderot) and R. Ferrière (Univ. Arizona and École Normale Supérieure), N. Champagnat studied deterministic and stochastic spatial eco-evolutionary dynamics along environmental gradients. This work focuses on numerical and analytical analysis of the clustering phenomenon in the population, and on the patterns of invasion fronts [13].
- Together with M. Benaïm (Univ, Neuchâtel), N. Champagnat and D. Villemonais studied stochastic algorithms to approximate quasi-stationary distributions of diffusion processes absorbed at the boundary of a bounded domain. They study a reinforced version of the diffusion, which is resampled according to its occupation measure when it reaches the boundary. They show that its occupation measure converges to the unique quasi-stationary distribution of the diffusion process [43].
- N. Champagnat, C. Fritsch and S. Billiard (Univ. Lille) studied models of food web adaptive evolution. They identified the biomass conversion efficiency as a key mechanism underlying food webs evolution and discussed the relevance of such models to study the evolution of food webs [51].
- N. Champagnat and J. Claisse (Univ. Paris-Dauphine) studied the ergodic and infinite horizon controls of discrete population dynamics with almost sure extinction in finite time. This can either correspond to control problems in favor of survival or of extinction, depending on the cost function. They have proved that these two problems are related to the quasi-stationary distribution of the processes controled by Markov controls [18].
- N. Champagnat and B. Henry (Univ. Lille 1) studied a probabilistic approach for the Hamilton-Jacobi limit of non-local reaction-diffusion models of adaptive dynamics when mutations are small. They used a Feynman-Kac interpretation of the partial differential equation and large deviation estimates to obtain a variational characterization of the limit. They also studied in detail the case of finite phenotype space with exponentially rare mutations, where they were able to obtain uniqueness of the limit [19].
- N. Champagnat and D. Villemonais solved a general conjecture on the Fleming-Viot particle systems approximating quasi-stationary distributions (QSD): in cases where several quasi-stationary distributions exist, it is expected that the stationary distributions of the Fleming-Viot processes approach a particular QSD, called minimal QSD. They proved that this holds true for general absorbed Markov processes with soft obstacles [20].
- N. Champagnat and D. Villemonais studied the geometric convergence of normalized unbounded semigroups. They proved in [47] that general criteria for this convergence can be easily deduced from their recent results on the theory of quasi-stationary distributions.
- N. Champagnat, S. Méléard (École Polytechnique) and V.C. Tran (Univ. Paris Est Marne-la-Vallée) studied evolutionary models of bacteria with horizontal transfer. They considered in [46] a scaling of parameters taking into account the influence of negligible but non-extinct populations, allowing them to study specific phenomena observed in these models (re-emergence of traits, cyclic evolutionary dynamics and evolutionary suicide).
- M. Bahlali (CEREA, France) , C. Henry and B. Carissimo (CEREA, France) clarify issues related to the expression of Lagrangian stochastic models used for atmospheric dispersion applications. They showed that accurate simulations are possible only if two aspects are properly addressed: the respect of the well-mixed criterion (related to the incorporation of the mean pressure-gradient term in the mean drift-term) and the consistency between Eulerian and Lagrangian turbulence models (regarding turbulence models, boundary and divergence-free conditions).
- A. Lejay and A. Brault have continued their work on rough flows, which provides an unified framework to deal with the theory of rough paths from the point of view of flows. In particular, they have studied consistency, stability and generic properties of rough differential equations [45].
- A. Lejay and P. Pigato have provided an estimator of a discontinuous drift coefficients [30], which follows their previous work on the oscillating Brownian motion and its application to financial models.

- A. Lejay and H. Mardones (U. la Serenan, Chile), have completed their work on the Monte Carlo simulation of the Navier-Stockes equations based on a new representation by Forward-Backward Stochastic Differential Equations [53].
- O. Faugeras, E. Soret and E. Tanré have obtained a Mean-Field description of thermodynamics limits of large population of neurons with random interactions. They have obtained the asymptotic behaviour for an asymmetric neuronal dynamics in a network of linear Hopfield neurons. They have a complete description of this limit with Gaussian processes. Furthermore, the limit object is not a Markov process [50].
- E. Tanré, P. Grazieschi (Univ. Warwick), M. Leocata (Univ. Pisa), C. Mascart (Univ. Côte d'Azur), J. Chevallier (Univ. of Grenoble) and F. Delarue (Univ. Côte d'Azur) have extended the previous work [9] to sparse networks of interacting neurons. They have obtained a precise description of the limit behavior of the mean field limit according to the probability of (random) interactions between two individual LIF neurons [24].
- P. Helson has studied the learning of an external signal by a neural network and the time to forget it when this network is submitted to noise. He has constructed an estimator of the initial signal thanks to the synaptic currents, which are Markov chains. The mathematical study of the Markov chains allow to obtain a lower bound on the number of external stimuli that the network can receive before the initial signal is forgotten [52].
- Q. Cormier and E. Tanré studied with Romain Veltz (team MATHNEURO) the long time behavior of a McKean-Vlasov SDE modeling a large assembly of neurons. A convergence to the unique (in this case) invariant measure is obtained assuming that the interactions between the neurons are weak enough. The key quantity in this model is the "firing rate": it gives the average number of jumps per unit of times of the solution of the SDE. They derive a non-linear Volterra equation satisfied by this rate. They used methods from integral equation to control finely the long time behavior of this firing rate [21].
- E. Tanré has worked with Nicolas Fournier (Sorbonne Université) and Romain Veltz (MATHNEURO Inria team) on a network of spiking networks with propagation of spikes along the dendrites. Consider a large number n of neurons randomly connected. When a neuron spikes at some rate depending on its electric potential, its membrane potential is set to a minimum value v_{min}, and this makes start, after a small delay, two fronts on the dendrites of all the neurons to which it is connected. Fronts move at constant speed. When two fronts (on the dendrite of the same neuron) collide, they annihilate. When a front hits the soma of a neuron, its potential is increased by a small value w_n. Between jumps, the potentials of the neurons are assumed to drift in $[v_{min}, \infty)$, according to some well-posed ODE. They prove the existence and uniqueness of a heuristically derived mean-field limit of the system when $n \to \infty$ [23].
- 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 [49].
- E. Tanré has worked with Alexandre Richard (Centrale-Supelec) and Soledad Torres (Universidad de Valparaíso, Chile) on a one-dimensional fractional SDE with reflection. They have proved the existence of the reflected SDE with a penalization scheme (suited to numerical approximation). Penalization also gives an algorithm to approach this solution [55].
- The Neutron Transport Equation (NTE) describes the flux of neutrons over time through an inhomogeneous fissile medium. A probabilistic solution of the NTE is considered in order to demonstrate a Perron-Frobenius type growth of the solution via its projection onto an associated leading eigenfunction. The associated eigenvalue, denoted k_{eff}, has the physical interpretation as being the ratio of neutrons produced (during fission events) to the number lost (due to absorption in the reactor or leakage at the boundary) per typical fission event. Together with A. M. G. Cox, E. L. Horton and A. E. Kyprianou (Univ. Bath), D. Villemonais developed the stochastic analysis of the NTE by giving a rigorous probabilistic interpretation of k_{eff} [48].

- In [34], D. Villemonais obtained a lower bound for the coarse Ricci curvature of continuous-time pure-jump Markov processes, with an emphasis on interacting particle systems. Applications to several models are provided, with a detailed study of the herd behavior of a simple model of interacting agents.
- In collaboration with C. Coron (Univ. Paris Sud) and S. Méléard (École Polytechnique), D. Villemonais studied in [22] the way alleles extinctions and fixations occur for a multiple allelic proportions model based on diffusion processes. It is proved in particular that alleles extinctions occur successively and that a 0-1 law holds for fixation and extinction: depending on the population dynamics near extinction, either fixation occurs before extinction, or the converse, almost surely.
- Mean telomere length in human leukocyte DNA samples reflects the different lengths of telomeres at the ends of the 23 chromosomes and in an admixture of cells. Together with S. Toupance (CHRU Nancy), D. Germain (Univ. Lorraine), A. Gégout-Petit (Univ. Lorraine and BIGS Inria team), E. Albuisson (CHRU Nancy) and A. Benetos (CHRU Nancy), D. Villemonais analysed telomere length distributions dynamics in adults individuals. It is proved in [33] that the shape of this distribution is stable over the lifetime of individuals.
- J. Bion-Nadal (Ecole Polytechnique) and D. Talay have pursued their work on their Wasserstein-type distance on the set of the probability distributions of strong solutions to stochastic differential equations. This new distance is defined by restricting the set of possible coupling measures and can be expressed in terms of the solution to a stochastic control problem, which allows one to deduce a priori estimates or to obtain numerical evaluations [15].

  A notable application concerns the following modeling issue: given an exact diffusion model, how to select a simplified diffusion model within a class of admissible models under the constraint that the probability distribution of the exact model is preserved as much as possible? The objective being to select a model minimizing the above distance to a target model, approximations of the optimal model have been established. The construction and analysis of an efficient stochastic algorithm are being in progress.
- D. Talay and M. Tomašević have continued to work on their 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. D. Talay and M. Tomašević are studying the well-posedness and the propagation of chaos of the particle system related to the two-dimensional parabolic-parabolic Keller-Segel system.
- V. Martin Lac, R. Maftei D. Talay and M. Tomašević have continued to work on theoretical and algorithmic questions related to the simulation of the Keller–Segel particle systems. The library DIAMSS has been developed.
- H. Olivero (Inria, now University of Valparaiso, Chile) and D. Talay have continued to work on their hypothesis test which helps to detect when the probability distribution of complex stochastic simulations has an heavy tail and thus possibly an infinite variance. This issue is notably important when simulating particle systems with complex and singular McKean-Vlasov interaction kernels whick make it extremely difficult to get a priori estimates on the probability laws of the mean-field limit, the related particle system, and their numerical approximations. In such situations the standard limit theorems do not lead to effective tests. In the simple case of independent and identically distributed sequences the procedure developed this year and its convergence analysis are based on deep tools coming from the statistics of semimartingales.
- I. Honoré and D. Talay have worked on statistical issues related to numerical approximations of invariant probability measures of ergodic diffusions. These approximations are based on the simulation of one single trajectory up to long time horizons. I. Honoré and D. Talay handle the critical situations where the asymptotic variance of the normalized error is infinite.
- V. Martin Lac, H. Olivero-Quinteros and D. Talay have worked on theoretical and algorithmic questions related to the simulation of large particle systems under singular interactions and to critical numerical issues related to the simulation of independent random variables with heavy tails. A preliminary version of a library has been developed.

- C. Graham (École Polytechnique) and D. Talay have ended the second volume of their series on Mathematical Foundation of Stochastic Simulation to be published by Springer.

6.1.2. Other works in progress

- K. Martinez, M. Bossy, C. Henry, R. Maftei and S. Sherkarforush work on a refined algorithm for macroscopic simulations of particle agglomeration using population balance equations (PBE). More precisely, their study is focused on identifying regions with non-homogeneous spatial distribution of particles. This is indeed a major drawback of PBE formulations which require a well-mixed condition to be satisfied. The developed algorithm identifies higher/lower density regions to treat them separately.
- S. Allende (CEMEF, France), J. Bec (CEMEF, France), M. Bossy, L. Campana, M. Ferrand (EDF, France), C. Henry and J.P. Minier (EDF, France) work together on a macroscopic model for the dynamics of small, flexible, inextensible fibers in a turbulent flow. Following the model developed at Inria, they perform numerical simulations of the orientation of such fibers in wall-bounded turbulent flows and compare it to microscopic simulations obtained with Direct Numerical Simulation (DNS). This work is performed under the POPART project.
- N. Champagnat, C. Fritsch and U. Herbach are working with A. Harlé (Institut de Cancérologie de Lorraine), J.-L. Merlin (ICL), E. Pencreac'h (CHRU Strasbourg), A. Gégout-Petit, P. Vallois, A. Muller-Gueudin (Inria BIGS team) and A. Kurtzmann (Univ. Lorraine) within an ITMO Cancer project on modeling and parametric estimation of dynamical models of circulating tumor DNA (ctDNA) of tumor cells, divided into several clonal populations. The goal of the project is to predict the emergence of a clonal population resistant to a targeted therapy in a patient's tumor, so that the therapy can be modulated more efficiently.
- N. Champagnat and R. Loubaton are working with P. Vallois (Univ. Lorraine and Inria BIGS team) and L. Vallat (CHRU Strasbourg) on the inference of dynamical gene networks from RNAseq and proteome data.
- N. Champagnat, E. Strickler and D. Villemonais are working on the characterization of convergence in Wasserstein distance of conditional distributions of absorbed Markov processes to a quasi-stationary distribution.
- N. Champagnat and V. Hass are studying evolutionary models of adaptive dynamics under an assumption of large population and small mutations. They expect to recover variants of the canonical equation of adaptive dynamics, which describes the long time evolution of the dominant phenotype in the population, under less stringent biological assumptions than in previous works.
- Q. Cormier, E. Tanré and Romain Veltz (team MATHNEURO) are working on the local stability of a stationary solution of some McKean-Vlasov equation. They also obtain spontaneous oscillation of the solution for critical values of the external currents or the interactions.
- M. Deaconu, A. Lejay and E. Mordecki (U. de la República, Uruguay) are studying an optimal stopping problem for the Snapping Out Brownian motion.
- M. Deaconu and A. Lejay are currently working on the simulation and the estimation of the fragmentation equation through its probabilistic representation.
- S. Allende (CEMEF, France) , C. Henry and J. Bec (CEMEF, France) work on the dynamics of small, flexible, inextensible fibers in a turbulent flow. They show that the fragmentation of fibers smaller than the smallest fluid scale in a turbulent flow occurs through tensile fracture (i.e. when the fiber is stretched along its main axis) or through flexural failure (i.e. when the fiber curvature is too high as it buckles under compressive load). Statistics of such events are provide together with measures of the rate of fragmentation and daughter size distributions, which are basic ingredients for macroscopic fragmentation models.

- C. Henry and M.L. Pedrotti (LOV, France) are working together on the topic of sedimentation of plastic that are populated by biological organisms (this is called biofouling). Biofouling modifies the density of plastic debris in the ocean and can lead to their sedimentation towards deeper regions. This work is done under the PLAISE project, which comprises measurements (by the LOV) and simulations (by C. Henry).
- C. Fritsch is working with A. Gégout-Petit (Univ. Lorraine and EPI BIGS), B. Marçais (INRA, Nancy) and M. Grosdidier (INRA, Avignon) on a statistical analysis of a Chalara Fraxinea model.
- C. Fritsch is working with Tanjona Ramiadantsoa (Univ. Wisconsin-Madison) on a model of extinction of orphaned plants.
- A. Lejay and M. Clausel (U. Lorraine) are studing the clustering method based on the use of the signature and the iterated integrals of time series. It is based on asymmetric spectral clustering [41].
- In collaboration with L. Lenotre (postdoc at IECL between Oct. 2018 and Sep. 2019), A. Gégout-Petit (Univ. Lorraine and Inria BIGS team) and O. Coudray (Master degree student), D. Villemonais conducted preliminary researches on branching models for the telomeres' length dynamics across generations.

VERIDIS Project-Team

7. New Results

7.1. Automated and Interactive Theorem Proving

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

7.1.1. Combination of Satisfiability Procedures

Joint work with Christophe Ringeissen (Inria Nancy – Grand Est, Pesto) and Paula Chocron (Insikt Intelligence, Spain).

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

In 2015, we defined a sound and complete combination procedure *à la* Nelson-Oppen for the theory of absolutely free data structures (including lists and trees) connected to another theory via bridging functions [59]. 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 [60] amenable to similar combinations: this class includes the theory of equality, the theory of absolutely free data structures, and all the theories in between.

In 2018 and 2019, we have been improving the framework and unified both results. This was published in the Journal of Automated Reasoning in 2019 [19].

7.1.2. Quantifier Handling in SMT

Joint work with Cezary Kaliszyk (Univ. of Innsbruck).

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

In 2019, we investigated machine learning techniques for predicting the usefulness of an instance in order to decrease the number of instances passed to the SMT solver. For this, we proposed a meaningful way to characterize the state of an SMT solver, to collect instantiation learning data, and to integrate a predictor in the core of a state-of-the-art SMT solver. This ultimately leads to more efficient SMT solving for quantified problems.

7.1.3. Higher-Order SMT

Joint work with Haniel Barbosa, Andrew Reynolds, Cesare Tinelli (Univ. of Iowa), and Clark Barrett (Stanford)

SMT solvers have throughout the years been able to cope with increasingly expressive formulas, from ground logics to full first-order logic (FOL). In contrast, the extension of SMT solvers to higher-order logic (HOL) was mostly unexplored. We proposed a pragmatic extension for SMT solvers to support HOL reasoning natively without compromising performance on FOL reasoning, thus leveraging the extensive research and implementation efforts dedicated to efficient SMT solving. We showed how to generalize data structures and the ground decision procedure to support partial applications and extensionality, as well as how to reconcile quantifier instantiation techniques with higher-order variables. We also discussed a separate approach for redesigning an SMT solver for higher-order logic from the ground up via new data structures and algorithms. We applied our pragmatic extension to the CVC4 SMT solver and discussed a redesign of the veriT SMT solver. Our evaluation showed that they are competitive with state-of-the-art HOL provers and often outperform the traditional encoding into FOL.

This result was published at CADE 2019 [27]. We are also currently investigating extending the CCFV algorithm to higher-order logic.

7.1.4. Proofs for SMT

We have previously developed a framework for processing formulas in automatic theorem provers, with generation of detailed proofs that can be checked by external tools, including skeptical proof assistants. 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. In 2019, the format of proof output was further improved, while also improving the reconstruction procedure in the proof assistant Isabelle/HOL. This allowed the tactic using SMT with proofs to be regularly suggested by Sledgehammer as the fastest method to automatically solve proof goals. This was the subject of a workshop publication [36].

7.1.5. Clause Learning from Simple Models

The goal of this research is to guide inferences in expressive logics via simple models. Intuitively, a model is simple if computations with respect to the model can be done in polynomial time. We have shown that for first-order logic, models built from ground literals are sufficient to guide resolution inferences between non-ground clauses [35]. We have also investigated the expressivity of model representation formalisms in general [41]. Model representation formalisms built on atoms with only linear variable occurrences have the finite model property. Hence, they cannot represent infinite models.

7.1.6. SPASS-SATT

We have further developed our CDCL(T) solver SPASS-SATT. It is the combination of our SAT solver SPASS-SAT with highly efficient theory solvers for linear arithmetic [31]. SPASS-SATT showed good performance at the SMT competition 2019 where it won the category on linear rational arithmetic and scored second on linear integer arithmetic. The winner of the linear integer arithmetic category was a portfolio solver including SPASS-SATT. Our main improvements are due to an advanced clause normal form translation, a close interaction between the theory solvers and the SAT solver SPASS-SAT, and and a new transformation turning unbounded integer problems into bounded integer problems.

7.1.7. Extension of a Highly Efficient Prover to λ-free Higher-Order Logic

Joint work with Simon Cruanes (Aesthetic Integration), Stephan Schulz (DHBW Stuttgart), and Petar Vukmirović (VU Amsterdam).

Superposition-based provers, such as E, SPASS, and Vampire, are among the most successful reasoning systems for first-order logic. They serve as backends in various frameworks, including software verifiers, automatic higher-order theorem provers, and one-click "hammers" in proof assistants. Decades of research have gone into refining calculi, devising efficient data structures and algorithms, and developing heuristics to guide proof search. This work has mostly focused on first-order logic with equality, with or without arithmetic.

To obtain better performance, we propose to start with a competitive first-order prover and extend it to full higher-order logic one feature at a time. Our goal is a *graceful* extension, in keeping with the zero-overhead principle: *What you don't use, you don't pay for.*

As a stepping stone towards full higher-order logic, we initially restricted our focus to a higher-order logic without λ-expressions. Compared with first-order logic, its distinguishing features are partial application and applied variables. Our vehicle is E, a prover developed primarily by Schulz. It is written in C and offers good performance. E regularly scores among the top systems at the CASC competition, and usually is the strongest open source prover in the relevant divisions. It also serves as a backend for competitive higher-order provers.

Our experiments show that the λ-free higher-order version of E is practically as fast as E on first-order problems and can also prove higher-order problems that do not require synthesizing λ-terms. As a next step, we plan to add support for λ-terms and higher-order unification. This work is described in a TACAS 2019 conference paper [42]; an extended version of this paper has been invited to a special issue of the *International Journal on Software Tools for Technology Transfer*.

7.1.8. Extension of the Superposition Calculus with λ-Abstractions

Joint work with Alexander Bentkamp (VU Amsterdam) and Petar Vukmirović (VU Amsterdam).

We designed a superposition calculus for a clausal fragment of extensional polymorphic higher-order logic that includes anonymous functions but excludes Booleans. The inference rules work on $\beta\eta$-equivalence classes of λ-terms and rely on higher-order unification to achieve refutational completeness.

We implemented the calculus in the Zipperposition prover. Our empirical evaluation includes benchmarks from the TPTP (Thousands of Problems for Theorem Provers) and interactive verification problems exported from Isabelle/HOL. The results appear promising and suggest that an optimized implementation inside a competitive prover such as E, SPASS, or Vampire would outperform existing higher-order automatic provers. This research was presented at the CADE 2019 conference [28].

7.1.9. Automated Reasoning over Biological Networks

[54] study toricity of steady state ideals of biological models. From a computational point of view, models identified as toric allow to employ tools from toric geometry for a complexity reduction step. From a scientific point of view, toric models are known to have scale invariant multistationarity in the space of linear conserved quantities. This can be interpreted as a dimension reduction of the multistationarity problem. We propose a generalization of the notion of toricity, compatible with our above remarks, in terms of the geometry of the variety instead of the syntactic shape of generators of the ideal. We consider 129 models from the BioModels repository [67], for which ODEbase [0] provides input data directly usable for symbolic computation. While the existing literature was mostly limited to the complex numbers, we use real quantifier elimination methods to treat also the real case, which is clearly the relevant domain from a scientific point of view. In practice, our real computations in Redlog [4] can compete with our complex ones. In theory we show that our real algorithms are in EXPTIME while Gröbner bases, which are typically used when working with ideal generators, are EXPSPACE-complete [68]. To our knowledge, this is the first time that such a comprehensive set of biomodels has been systematically processed using symbolic methods.

7.1.10. Towards an Improved Encoding of TLA+ Proof Obligations

We reconsider the encoding of proof obligations that arise in proofs about TLA$^+$ specifications in multi-sorted first-order logic, and specifically their translations to SMT solvers. Our previous work [69] relied on type inference for identifying expressions having atomic types such as integers but did not exploit more complex types, even if such types were constructed during type inference. A more pervasive use of types for translating set-theoretic expressions to the input language of SMT solvers appears promising in order to reduce the use of type injections and quantifiers and thus simplify the proof obligations passed to the solver, but it raises non-trivial soundness and completeness issues. Techniques of gradual typing designed for programming languages where type inference is not fully possible statically may be helpful in this context. A related problem is support for instantiation hints for quantified formulas given by the user. A first paper will be presented at JFLA 2020.

[0] http://odebase.cs.uni-bonn.de/

7.1.11. Formal Proofs of Tarjan's Algorithm

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

We consider Tarjan's classical algorithm for computing strongly connected components in a graph as a case study of intermediate complexity for comparing interactive proof assistants. Representing the algorithm as a functional program (rather than its more conventional imperative representation), we proved its correctness in three different proof assistants (Coq, Isabelle/HOL, and Why3). The proofs are based on essentially the same formulation of the algorithm and of its invariants, allowing us to compare differences due to idiosyncracies of the proof assistants, such as their ability to handle mutually recursive function definitions, proving termination beyond syntactic criteria, and their degree of automation. Our results were presented at ITP 2019 [33].

7.1.12. Implementation of an Efficient Validation of FOLID Cyclic Induction Reasoning

Checking the soundness of cyclic induction reasoning for first-order logic with inductive definitions (FOL_{ID}) is decidable but the standard checking method is based on an exponential complement operation for Büchi automata. We devised a polynomial method "semi-deciding" this problem; its most expensive steps are reminiscent of the comparisons with multiset path orderings. In practice, it has been integrated in the CYCLIST prover and successfully checked all the proofs included in its distribution. The work was presented at the CiSS2019 conference (Circularity in Syntax and Semantics) and the software is available at https://members.loria.fr/SStratulat/files/e-cyclist.zip.

7.2. Formal Methods for Developing and Analyzing Algorithms and Systems

Participants: Étienne André, Marie Duflot-Kremer, Yann Duplouy, Margaux Duroeulx, Igor Konnov, Dominique Méry, Stephan Merz, Nicolas Schnepf, Christoph Weidenbach.

7.2.1. Synthesis of Security Chains for Software Defined Networks

Joint work with Rémi Badonnel and Abdelkader Lahmadi (Inria Nancy – Grand Est, Resist).

The PhD thesis of Nicolas Schnepf focuses on applying techniques based on formal methods in the area of network communications, and in particular for the construction, verification, and optimization of chains of security functions in the setting of software-defined networks (SDN). The main objective is to prevent applications from disrupting the functioning of the network or services, for example by launching denial of service attacks, port scanning or similar activities.

We designed techniques for formally verifying security chains using SMT solving and symbolic model checking. Furthermore, we developed and prototypically implemented an approach for (i) learning a Markov chain characterizing the network behavior of an Android application based on its observed communications, (ii) inferring appropriate security functions from the structure of that Markov chain and thresholds set by the network operator, using techniques of logic programming, (iii) combining security functions for individual applications into larger security chains, and (iv) optimizing the deployment of security chains for a given SDN infrastructure using techniques of (linear or non-linear) optimization or optimizing SMT solvers. Two papers were presented at IM 2019 [39], [38], the PhD thesis [12] was defended in September 2019, and a journal paper is in preparation.

7.2.2. Satisfiability Techniques for Reliability Assessment

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

In the context of the PhD thesis of Margaux Durœulx, funded by the Lorraine University of Excellence program, we explore the applicability of satisfiability techniques for assessing the reliability of complex systems. In particular, we consider component-based systems modeled using fault trees that can be seen as a visual representation of the structure function indicating which combinations of component failures lead to system failures. We rely on SAT solvers to compute minimal tie sets, i.e., minimal sets of components whose functioning ensures that the overall system works. These tie sets are instrumental for a probabilistic reliability assessment. In 2019, we have extended this idea to dynamic fault trees where the order of component failures needs to be taken into account in order to determine the failure status of the overall system [34].

7.2.3. Statistical Model Checking of Distributed Programs

Yann Duplouy joined the HAC SPECIS project (cf. section 9.2) in December 2018 as a post-doctoral researcher with the objective of designing and implementing a statistical model checker within the SimGrid framework. So far he added to SimGrid the possibility to use stochastic profiles, introducing probabilities in the model of the network. He also developed a prototype tool that can be interfaced with the SimGrid simulators to perform statistical model checking on the actual programs simulated using the SimGrid framework. He now validates this prototype on concrete case studies, including the Bit Torrent protocol with probabilistic failures of the nodes.

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

Joint work with Nathalie Bertrand (Inria Rennes Bretagne – Atlantique, SUMO), Marijana Lazić (TU Munich) and Ilina Stoilkovska, Josef Widder, Florian Zuleger (TU Wien).

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

In earlier work, we introduced threshold automata for representing processes in such algorithms and showed that systems of threshold automata have bounded diameters that do not depend on the parameters such as n and t, provided that a single-step acceleration is allowed [62], [63], [64].

Our previous results apply to asynchronous algorithms. It is well-known that distributed consensus cannot be solved in purely asynchronous systems [61]. However, when an algorithm is provided with a random coin, consensus becomes solvable (e.g., the algorithm by Ben-Or, 1993). In [29], we introduced an approach to parameterized verification of randomized threshold-guarded distributed algorithms, which proceed in an unbounded number of rounds and toss a coin to break symmetries. This approach integrates two levels of reasoning: (1) proving safety and liveness of a single round system with ByMC by replacing randomization with non-determinism, (2) showing almost-sure termination of an algorithm by using the verification results for the non-deterministic system. To show soundness, we proved several theorems that reduce reasoning about multiple rounds to reasoning about a single round. We verified five prominent algorithms, including Ben-Or's randomized consensus and randomized one-step consensus (RS-BOSCO [71]). The verification of the latter algorithm required us to run experiments in Grid5000. This paper was presented at CONCUR 2019.

Another way of making consensus solvable is to impose synchrony on the executions of a distributed system. In [40] we introduced synchronous threshold automata, which execute in lock-step and count the number of processes in given local states. In general, we showed that even reachability of a parameterized set of global states in such a distributed system is undecidable. However, we proved that systems of automata with monotonic guards have bounded diameters, which allows us to use SMT-based bounded model checking as a complete parameterized verification technique. We introduced a procedure for computing the diameter of a counter system of synchronous threshold automata, applied it to the counter systems of 8 distributed algorithms from the literature, and found that their diameters are tiny (from 1 to 4). This makes our approach practically feasible, despite undecidability in general. This paper was presented at TACAS 2019. The paper was invited to the special issue of TACAS 2019, to appear in the *International Journal on Software Tools for Technology Transfer* in 2020.

7.2.5. Symbolic Model Checking of TLA+ Specifications

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

TLA$^+$ is a general language introduced by Leslie Lamport for specifying temporal behavior of computer systems [66]. The tool set for TLA$^+$ includes an explicit-state model checker TLC. As explicit state model

checkers do not scale to large verification problems, we started the project APALACHE [0] on developing a symbolic model checker for TLA$^+$ in 2016.

Following our results in 2018 [65], we have extended the symbolic model checker for TLA$^+$. In [22], we have defined the translation process from TLA$^+$ to SMT as a series of rewriting rules. Furthermore, we have proven soundness of this translation. Our experiments show that APALACHE runs faster than TLC when proving inductive invariants. APALACHE also implements bounded model checking, which has to be improved, in order to make it competitive with TLC. The paper [22] was presented at ACM OOPSLA 2019.

7.2.6. Incremental Development of Systems and Algorithms

Joint work with Rosemary Monahan (NUI Maynooth, Ireland) 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 2019 is the development of a distributed pattern [26] handling the dynamicity of the topology of networks.

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