RESEARCH CENTER
Grenoble - Rhône-Alpes

FIELD

Activity Report 2019

Section New Results

Edition: 2020-03-21

ARIC Project-Team

7. New Results

7.1. Efficient approximation methods

7.1.1. Exchange algorithm for evaluation and approximation error-optimized polynomials

Machine implementation of mathematical functions often relies on polynomial approximations. The particularity is that rounding errors occur both when representing the polynomial coefficients on a finite number of bits, and when evaluating it in finite precision. Hence, for finding the best polynomial (for a given fixed degree, norm and interval), one has to consider both types of errors: approximation and evaluation. While efficient algorithms were already developed for taking into account the approximation error, the evaluation part is usually a posteriori handled, in an ad-hoc manner. In [15], we formulate a semi-infinite linear optimization problem whose solution is the best polynomial with respect to the supremum norm of the sum of both errors. This problem is then solved with an iterative exchange algorithm, which can be seen as an extension of the well-known Remez algorithm. A discussion and comparison of the obtained results on different examples are finally presented.

7.1.2. On Moment Problems with Holonomic Functions

Many reconstruction algorithms from moments of algebraic data were developed in optimization, analysis or statistics. Lasserre and Putinar proposed an exact reconstruction algorithm for the algebraic support of the Lebesgue measure, or of measures with density equal to the exponential of a known polynomial. Their approach relies on linear recurrences for the moments, obtained using Stokes theorem. In [16], we extend this study to measures with holonomic densities and support with real algebraic boundary. In the framework of holonomic distributions (i.e. they satisfy a holonomic system of linear partial or ordinary differential equations with polynomial coefficients), an alternate method to creative telescoping is proposed for computing linear recurrences for the moments. When the coefficients of a polynomial vanishing on the support boundary are given as parameters, the obtained recurrences have the advantage of staying linear with respect to them. This property allows for an efficient reconstruction method. Given a finite number of numerically computed moments for a measure with holonomic density, and assuming a real algebraic boundary for the support, we propose an algorithm for solving the inverse problem of obtaining both the coefficients of a polynomial vanishing on the boundary and those of the polynomials involved in the holonomic operators which annihilate the density.

7.1.3. A certificate-based approach to formally verified approximations

In [17], we present a library to verify rigorous approximations of univariate functions on real numbers, with the Coq proof assistant. Based on interval arithmetic, this library also implements a technique of validation a posteriori based on the Banach fixed-point theorem. We illustrate this technique on the case of operations of division and square root. This library features a collection of abstract structures that organize the specification of rigorous approximations, and modularize the related proofs. Finally, we provide an implementation of verified Chebyshev approximations, and we discuss a few examples of computations.

7.2. Floating-point and validated numerics

7.2.1. Error analysis of some operations involved in the Cooley-Tukey Fast Fourier Transform

We are interested in [4] in obtaining error bounds for the classical Cooley-Tukey FFT algorithm in floating-point arithmetic, for the 2-norm as well as for the infinity norm. For that purpose we also give some results on the relative error of the complex multiplication by a root of unity, and on the largest value that can take the real or imaginary part of one term of the FFT of a vector x, assuming that all terms of x have real and imaginary parts less than some value b.

7.2.2. Algorithms for triple-word arithmetic

Triple-word arithmetic consists in representing high-precision numbers as the unevaluated sum of three floating-point numbers (with "nonoverlapping" constraints that are explicited in the paper). We introduce and analyze in [7] various algorithms for manipulating triple-word numbers: rounding a triple-word number to a floating-point number, adding, multiplying, dividing, and computing square-roots of triple-word numbers, etc. We compare our algorithms, implemented in the Campary library, with other solutions of comparable accuracy. It turns out that our new algorithms are significantly faster than what one would obtain by just using the usual floating-point expansion algorithms in the special case of expansions of length 3.

7.2.3. Accurate Complex Multiplication in Floating-Point Arithmetic

We deal in [24] with accurate complex multiplication in binary floating-point arithmetic, with an emphasis on the case where one of the operands in a "double-word" number. We provide an algorithm that returns a complex product with normwise relative error bound close to the best possible one, i.e., the rounding unit u.

7.2.4. Semi-automatic implementation of the complementary error function

The normal and complementary error functions are ubiquitous special functions for any mathematical library. They have a wide range of applications. Practical applications call for customized implementations that have strict accuracy requirements. Accurate numerical implementation of these functions is, however, non-trivial. In particular, the complementary error function erfc for large positive arguments heavily suffers from cancellation, which is largely due to its asymptotic behavior. We provide a semi-automatic code generator for the erfc function which is parameterized by the user-given bound on the relative error. Our solution, presented in [31], exploits the asymptotic expression of erfc and leverages the automatic code generator Metalibm that provides accurate polynomial approximations. A fine-grained a priori error analysis provides a libm developer with the required accuracy for each step of the evaluation. In critical parts, we exploit double-word arithmetic to achieve implementations that are fast, yet accurate up to 50 bits, even for large input arguments. We demonstrate that for high required accuracies the automatically generated code has performance comparable to that of the standard libm and for lower ones our code demonstrated roughly 25% speedup.

7.2.5. Posits: the good, the bad and the ugly

Many properties of the IEEE-754 floating-point number system are taken for granted in modern computers and are deeply embedded in compilers and low-level softare routines such as elementary functions or BLAS. In [32] we review such properties on the recently proposed Posit number system. Some are still true. Some are no longer true, but sensible work-arounds are possible, and even represent exciting challenge for the community. Some, in particular the loss of scale invariance for accuracy, are extremely dangerous if Posits are to replace floating point completely. This study helps framing where Posits are better than floating-point, where they are worse, and what tools are missing in the Posit landscape. For general-purpose computing, using Posits as a storage format only could be a way to reap their benefits without loosing those of classical floating-point. The hardware cost of this alternative is studied.

7.2.6. The relative accuracy of $(x + y) * (x - y)$

We consider in [8] the relative accuracy of evaluating $(x + y)(x - y)$ in IEEE floating-point arithmetic, when x and y are two floating-point numbers and rounding is to nearest. This expression can be used for example as an efficient cancellation-free alternative to $x^2 - y^2$ and is well known to have low relative error, namely, at most about $3u$ with u denoting the unit roundoff. In this paper we complement this traditional analysis with a finer-grained one, aimed at improving and assessing the quality of that bound. Specifically, we show that if the tie-breaking rule is *to away* then the bound $3u$ is asymptotically optimal. In contrast, if the tie-breaking rule is *to even*, we show that asymptotically optimal bounds are now $2.25u$ for base two and $2u$ for larger bases, such as base ten. In each case, asymptotic optimality is obtained by the explicit construction of a certificate, that is, some floating-point input (x, y) parametrized by u and such that the error of the associated result is equivalent to the error bound as $u \to 0$. We conclude with comments on how $(x + y)(x - y)$ compares with x^2 in the presence of floating-point arithmetic, in particular showing cases where the computed value of $(x + y)(x - y)$ exceeds that of x^2.

7.2.7. The MPFI Library: Towards IEEE 1788-2015 Compliance

The IEEE 1788-2015 has standardized interval arithmetic. However, few libraries for interval arithmetic are compliant with this standard. In the first part of [30], the main features of the IEEE 1788-2015 standard are detailed. These features were not present in the libraries developed prior to the elaboration of the standard. MPFI is such a library: it is a C library, based on MPFR, for arbitrary precision interval arithmetic. MPFI is not (yet) compliant with the IEEE 1788-2015 standard for interval arithmetic: the planned modifications are presented.

7.3. Lattices: algorithms and cryptology

7.3.1. Approx-SVP in ideal lattices with pre-processing

In [28], we describe an algorithm to solve the approximate Shortest Vector Problem for lattices corresponding to ideals of the ring of integers of an arbitrary number field K. This algorithm has a pre-processing phase, whose run-time is exponential in $\log|\Delta|$ with Δ the discriminant of K. Importantly, this pre-processing phase depends only on K. The pre-processing phase outputs an advice, whose bit-size is no more than the run-time of the query phase. Given this advice, the query phase of the algorithm takes as input any ideal I of the ring of integers, and outputs an element of I which is at most $\exp(\widetilde{O}((\log|\Delta|)^{\alpha+1}/n))$ times longer than a shortest non-zero element of I (with respect to the Euclidean norm of its canonical embedding). This query phase runs in time and space $\exp(\widetilde{O}((\log|\Delta|)^{\max(2/3,1-2\alpha)}))$ in the classical setting, and $\exp(\widetilde{O}((\log|\Delta|)^{1-2\alpha}))$ in the quantum setting. The parameter α can be chosen arbitrarily in $[0,1/2]$. Both correctness and cost analyses rely on heuristic assumptions, whose validity is consistent with experiments.

The algorithm builds upon the algorithms from Cramer al. [EUROCRYPT 2016] and Cramer et al. [EUROCRYPT 2017]. It relies on the framework from Buchmann [Séminaire de théorie des nombres 1990], which allows to merge them and to extend their applicability from prime-power cyclotomic fields to all number fields. The cost improvements are obtained by allowing precomputations that depend on the field only.

7.3.2. An LLL algorithm for module lattices

The LLL algorithm takes as input a basis of a Euclidean lattice, and, within a polynomial number of operations, it outputs another basis of the same lattice but consisting of rather short vectors. In [23], we provide a generalization to R-modules contained in K^n for arbitrary number fields K and dimension n, with R denoting the ring of integers of K. Concretely, we introduce an algorithm that efficiently finds short vectors in rank-n modules when given access to an oracle that finds short vectors in rank-2 modules, and an algorithm that efficiently finds short vectors in rank-2 modules given access to a Closest Vector Problem oracle for a lattice that depends only on K. The second algorithm relies on quantum computations and its analysis is heuristic.

7.3.3. The general sieve kernel and new records in lattice reduction

In [14], we propose the General Sieve Kernel (G6K), an abstract stateful machine supporting a wide variety of lattice reduction strategies based on sieving algorithms. Using the basic instruction set of this abstract stateful machine, we first give concise formulations of previous sieving strategies from the literature and then propose new ones. We then also give a light variant of BKZ exploiting the features of our abstract stateful machine. This encapsulates several recent suggestions (Ducas at Eurocrypt 2018; Laarhoven and Mariano at PQCrypto 2018) to move beyond treating sieving as a blackbox SVP oracle and to utilise strong lattice reduction as preprocessing for sieving. Furthermore, we propose new tricks to minimise the sieving computation required for a given reduction quality with mechanisms such as recycling vectors between sieves, on-the-fly lifting and flexible insertions akin to Deep LLL and recent variants of Random Sampling Reduction.

Moreover, we provide a highly optimised, multi-threaded and tweakable implementation of this machine which we make open-source. We then illustrate the performance of this implementation of our sieving strategies by applying G6K to various lattice challenges. In particular, our approach allows us to solve previously unsolved instances of the Darmstadt SVP (151, 153, 155) and LWE (e.g. (75, 0.005)) challenges. Our solution for the SVP-151 challenge was found 400 times faster than the time reported for the SVP-150

challenge, the previous record. For exact SVP, we observe a performance crossover between G6K and FPLLL's state of the art implementation of enumeration at dimension 70.

7.3.4. Statistical zeroizing attack: cryptanalysis of candidates of BP obfuscation over GGH15 multilinear map

In [19], we present a new cryptanalytic algorithm on obfuscations based on GGH15 multilinear map. Our algorithm, statistical zeroizing attack, directly distinguishes two distributions from obfuscation while it follows the zeroizing attack paradigm, that is, it uses evaluations of zeros of obfuscated programs.

Our attack breaks the recent indistinguishability obfuscation candidate suggested by Chen et al. (CRYPTO'18) for the optimal parameter settings. More precisely, we show that there are two functionally equivalent branching programs whose CVW obfuscations can be efficiently distinguished by computing the sample variance of evaluations.

This statistical attack gives a new perspective on the security of the indistinguishability obfuscations: we should consider the shape of the distributions of evaluation of obfuscation to ensure security.

In other words, while most of the previous (weak) security proofs have been studied with respect to algebraic attack model or ideal model, our attack shows that this algebraic security is not enough to achieve indistinguishability obfuscation. In particular, we show that the obfuscation scheme suggested by Bartusek et al. (TCC'18) does not achieve the desired security in a certain parameter regime, in which their algebraic security proof still holds.

The correctness of statistical zeroizing attacks holds under a mild assumption on the preimage sampling algorithm with a lattice trapdoor. We experimentally verify this assumption for implemented obfuscation by Halevi et al. (ACM CCS'17).

7.3.5. Cryptanalysis of the CLT13 multilinear map

The reference [6] is the journal version of the Eurocrypt'15 article with the same title and authors.

7.3.6. Multi-Client Functional Encryption for Linear Functions in the Standard Model from LWE

Multi-client functional encryption (MCFE) allows ℓ clients to encrypt ciphertexts $\mathbf{C}_{t,1}, \mathbf{C}_{t,2}, ..., \mathbf{C}_{t,\ell}$ under some label. Each client can encrypt his own data X_i for a label t using a private encryption key ek_i issued by a trusted authority in such a way that, as long as all $\mathbf{C}_{t,i}$ share the same label t, an evaluator endowed with a functional key dk_f can evaluate $f(X_1, X_2, ..., X_\ell)$ without learning anything else on the underlying plaintexts X_i. Functional decryption keys can be derived by the central authority using the master secret key. Under the Decision Diffie-Hellman assumption, Chotard *et al.* (Asiacrypt 2018) recently described an adaptively secure MCFE scheme for the evaluation of linear functions over the integers. They also gave a decentralized variant (DMCFE) of their scheme which does not rely on a centralized authority, but rather allows encryptors to issue functional secret keys in a distributed manner. While efficient, their constructions both rely on random oracles in their security analysis. In [27], we build a standard-model MCFE scheme for the same functionality and prove it fully secure under adaptive corruptions. Our proof relies on the Learning-With-Errors (LWE) assumption and does not require the random oracle model. We also provide a decentralized variant of our scheme, which we prove secure in the static corruption setting (but for adaptively chosen messages) under the LWE assumption.

7.3.7. Zero-Knowledge Elementary Databases with More Expressive Queries

Zero-knowledge elementary databases (ZK-EDBs) are cryptographic schemes that allow a prover to commit to a set D of key-value pairs so as to be able to prove statements such as "x belongs to the support of D and D(x) = y" or "x is not in the support of D". Importantly , proofs should leak no information beyond the proven statement and even the size of D should remain private. Chase et al. (Eurocrypt'05) showed that ZK-EDBs are implied by a special flavor of non-interactive commitment, called mercurial commitment, which enables efficient instantiations based on standard number theoretic assumptions. On the other hand,

the resulting ZK-EDBs are only known to support proofs for simple statements like (non-)membership and value assignments. In [25], we show that mercurial commitments actually enable significantly richer queries. We show that, modulo an additional security property met by all known efficient constructions, they actually enable range queries over keys and values-even for ranges of super-polynomial size-as well as membership/non-membership queries over the space of values. Beyond that, we exploit the range queries to realize richer queries such as k-nearest neighbors and revealing the k smallest or largest records within a given range. In addition, we provide a new realization of trapdoor mercurial commitment from standard lattice asssumptions, thus obtaining the most expressive quantum-safe ZK-EDB construction so far.

7.3.8. Lossy Algebraic Filters With Short Tags

Lossy algebraic filters (LAFs) are function families where each function is parametrized by a tag, which determines if the function is injective or lossy. While initially introduced by Hofheinz (Eurocrypt 2013) as a technical tool to build encryption schemes with key-dependent message chosen-ciphertext (KDM-CCA) security, they also find applications in the design of robustly reusable fuzzy extractors. So far, the only known LAF family requires tags comprised of $\Theta(n^2)$ group elements for functions with input space $\mathbb{Z}_p$, where p is the group order. In [26], we describe a new LAF family where the tag size is only linear in n and prove it secure under simple assumptions in asymmetric bilinear groups. Our construction can be used as a drop-in replacement in all applications of the initial LAF system. In particular, it can shorten the ciphertexts of Hofheinz's KDM-CCA-secure public-key encryption scheme by 19 group elements. It also allows substantial space improvements in a recent fuzzy extractor proposed by Wen and Liu (Asiacrypt 2018). As a second contribution , we show how to modify our scheme so as to prove it (almost) tightly secure, meaning that security reductions are not affected by a concrete security loss proportional to the number of adversarial queries.

7.3.9. Shorter Quadratic QA-NIZK Proofs

Despite recent advances in the area of pairing-friendly Non-Interactive Zero-Knowledge proofs, there have not been many efficiency improvements in constructing arguments of satisfiability of quadratic (and larger degree) equations since the publication of the Groth-Sahai proof system (J. of Cryptology 2012). In [20], we address the problem of aggregating such proofs using techniques derived from the interactive setting and recent constructions of SNARKs. For certain types of quadratic equations, this problem was investigated before by González et al. (Asiacrypt'15). Compared to their result, we reduce the proof size by approximately 50

7.3.10. Shorter Pairing-based Arguments under Standard Assumptions

The paper [22] constructs efficient non-interactive arguments for correct evaluation of arithmetic and Boolean circuits with proof size $O(d)$ group elements, where d is the multiplicative depth of the circuit, under falsifiable assumptions. This is achieved by combining techniques from SNARKs and QA-NIZK arguments of membership in linear spaces. The first construction is very efficient (the proof size is $\approx 4d$ group elements and the verification cost is $4d$ pairings and $O(n + n + d)$ exponentiations, where n is the size of the input and n of the output) but one type of attack can only be ruled out assuming the knowledge soundness of QA-NIZK arguments of membership in linear spaces. We give an alternative construction which replaces this assumption with a decisional assumption in bilinear groups at the cost of approximately doubling the proof size. The construction for Boolean circuits can be made zero-knowledge with Groth-Sahai proofs, resulting in a NIZK argument for circuit satisfiability based on falsifiable assumptions in bilinear groups of proof size $O(n + d)$. Our main technical tool is what we call an "argument of knowledge transfer". Given a commitment C_1 and an opening x, such an argument allows to prove that some other commitment C_2 opens to $f(x)$, for some function f, even if C_2 is not extractable. We construct very short, constant-size, pairing-based arguments of knowledge transfer with constant-time verification for any linear function and also for Hadamard products. These allow to transfer the knowledge of the input to lower levels of the circuit.

7.3.11. Shorter Ring Signatures from Standard Assumptions

Ring signatures, introduced by Rivest, Shamir and Tauman (ASIACRYPT 2001), allow to sign a message on behalf of a set of users while guaranteeing authenticity and anonymity. Groth and Kohlweiss (EUROCRYPT

2015) and Libert *et al.* (EUROCRYPT 2016) constructed schemes with signatures of size logarithmic in the number of users. An even shorter ring signature, of size independent from the number of users, was recently proposed by Malavolta and Schroeder (ASIACRYPT 2017). However, all these short signatures are obtained relying on strong and controversial assumptions. Namely, the former schemes are both proven secure in the random oracle model while the later requires non-falsifiable assumptions.

The most efficient construction under mild assumptions remains the construction of Chandran et al. (ICALP 2007) with a signature of size $\Theta(\sqrt{n})$, where n is the number of users, and security is based on the Diffie-Hellman assumption in bilinear groups (the SXDH assumption in asymmetric bilinear groups).

In [21], we construct an asymptotically shorter ring signature from the hardness of the Diffie-Hellman assumption in bilinear groups. Each signature comprises $\Theta(n^{1/3})$ group elements, signing a message requires computing $\Theta(n^{1/3})$ exponentiations, and verifying a signature requires $\Theta(n^{2/3})$ pairing operations.

7.3.12. Two-Party ECDSA from Hash Proof Systems and Efficient Instantiations

ECDSA is a widely adopted digital signature standard. Unfortunately, efficient distributed variants of this primitive are notoriously hard to achieve and known solutions often require expensive zero knowledge proofs to deal with malicious adversaries. For the two party case, Lindell (CRYPTO 2017) recently managed to get an efficient solution which, to achieve simulation-based security, relies on an interactive, non standard, assumption on Paillier's cryptosystem.

In this paper [18] we generalize Lindell's solution using hash proof systems. The main advantage of our generic method is that it results in a simulation-based security proof without resorting to non-standard interactive assumptions.

Moving to concrete constructions, we show how to instantiate our framework using class groups of imaginary quadratic fields. Our implementations show that the practical impact of dropping such interactive assumptions is minimal. Indeed, while for 128-bit security our scheme is marginally slower than Lindell's, for 256-bit security it turns out to be better both in key generation and signing time. Moreover, in terms of communication cost, our implementation significantly reduces both the number of rounds and the transmitted bits without exception.

7.3.13. Algebraic XOR-RKA-Secure Pseudorandom Functions from Post-Zeroizing Multilinear Maps

In [13], we construct the first pseudorandom functions that resist a strong class of attacks where an adversary is able to run the cryptosystem not only with the fixed secret key, but with related keys where bits of its choice of the original keys are flipped. This problem is motivated by practical attacks that have been performed against physical devices. Our construction guarantees that every output of our construction, for the original key or for tampered keys, are pseudorandom, i.e. are computationally hard to distinguish from truly random values. To achieve this, we rely on a recent tool introduced in cryptography and termed multilinear maps. While multilinear maps have been recently attacked by several techniques, we prove that our construction remains secure despite the numerous vulnerabilities of current constructions of multilinear maps.

7.3.14. Unifying Leakage Models on a Rényi Day

Most theoretical models in cryptography suppose that an attacker can only observe the input/output behavior of a cryptosystem and nothing more. Yet, in the real world, cryptosystems run on physical devices and auxiliary information leaks from these devices. This leakage can sometimes be used to attack the system, even though it is proven secure in theory. To circumvent these issues, cryptographers have introduces several new security models in an attempt to encompass the different forms of leakage. Some models are simple, such as the probing model, and simple compilers allow to transform a system into one secure in the probing model, while some more realistic problems such as the noisy-leakage model are very involved. In [29], we show that these models are actually equivalent, proving in particular that the simple compilers are sufficient to guarantee security in realistic environments.

7.4. Algebraic computing and high-performance kernels

7.4.1. Linear differential equations as a data-structure

A lot of information concerning solutions of linear differential equations can be computed directly from the equation. It is therefore natural to consider these equations as a data-structure, from which mathematical properties can be computed. A variety of algorithms has thus been designed in recent years that do not aim at "solving", but at computing with this representation. Many of these results are surveyed in [11].

7.4.2. Absolute root separation

The absolute separation of a polynomial is the minimum nonzero difference between the absolute values of its roots. In the case of polynomials with integer coefficients, it can be bounded from below in terms of the degree and the height (the maximum absolute value of the coefficients) of the polynomial. We improve the known bounds for this problem and related ones. Then we report on extensive experiments in low degrees, suggesting that the current bounds are still very pessimistic. [5]

7.4.3. Improving the complexity of block low-rank factorizations with fast matrix arithmetic

We consider in [9] the LU factorization of an $n \times n$ matrix represented as a block low-rank (BLR) matrix: most of its off-diagonal blocks are approximated by matrices of small rank r, which reduces the asymptotic complexity of computing the LU factorization down to $\mathcal{O}(n^2 r)$. Even though lower complexities can be achieved with hierarchical matrices, the BLR format allows for a very simple and efficient implementation. In this article, our aim is to further reduce the BLR complexity without losing its nonhierarchical nature by exploiting fast matrix arithmetic, that is, the ability to multiply two $n \times n$ full-rank matrices together for $\mathcal{O}(n^\omega)$ flops, where $\omega < 3$. We devise a new BLR factorization algorithm whose cost is $\mathcal{O}(n^{(\omega+1)/2} r^{(\omega-1)/2})$, which represents an asymptotic improvement compared with the standard BLR factorization as soon as $\omega < 3$. In particular, for Strassen's algorithm, $\omega \approx 2.81$ yields the cost $\mathcal{O}(n^{1.904} r^{0.904})$. Our numerical experiments are in good agreement with this analysis.

7.4.4. Fast computation of approximant bases in canonical form

In [10] we design fast algorithms for the computation of approximant bases in shifted Popov normal form. For K a commutative field, let F be a matrix in $\mathsf{K}[x]^{m \times n}$ (truncated power series) and $\vec{d}$ be a degree vector, the problem is to compute a basis $P \in \mathsf{K}[x]^{m \times m}$ of the $\mathsf{K}[x]$-module of the relations $p \in \mathsf{K}[x]^{1 \times m}$ such that $p(x) \cdot F(x) \equiv 0 \mod x^{\vec{d}}$. We obtain improved complexity bounds for handling arbitrary (possibly highly unbalanced) vectors $\vec{d}$. We also improve upon previously known algorithms for computing P in normalized shifted form for an arbitrary shift. Our approach combines a recent divide and conquer strategy which reduces the general case to the case where information on the output degree is available, and partial linearizations of the involved matrices.

CASH Project-Team

6. New Results

6.1. Dataflow-explicit futures

Participants: Ludovic Henrio, Matthieu Moy, Amaury Maillé.

A future is a place-holder for a value being computed, and we generally say that a future is resolved when the associated value is computed. In existing languages futures are either implicit, if there is no syntactic or typing distinction between futures and non-future values, or explicit when futures are typed by a parametric type and dedicated functions exist for manipulating futures. We defined a new form of future, named data-flow explicit futures [43], with specific typing rules that do not use classical parametric types. The new futures allow at the same time code reuse and the possibility for recursive functions to return futures like with implicit futures, and let the programmer declare which values are futures and where synchronisation occurs, like with explicit futures. We prove that the obtained programming model is as expressive as implicit futures but exhibits a different behaviour compared to explicit futures. The current status of this work is the following:

- With collaborators from University of Uppsala and University of Oslo we worked on the design of programming constructs mixing implicit and dataflow-explicit futures (DeF). This has been published in ECOOP 2019 [10].
- Amaury Maillé did his internship in the Cash team (advised by Matthieu Moy and Ludovic Henrio), he worked on an implementation of DeF in the Encore language. This raised a difficulty regarding the interaction of DeF with generic types that has been partially solved. Now we need to generalize our approach to completely solve the issue.

6.2. Distributed futures

Participant: Ludovic Henrio.

We proposed the definition of *distributed futures*, a construct that provides at the same time a data container similar to a distributed vector, and a single synchronization entity that behaves similarly to a standard future. This simple construct makes it easy to program a composition, in a task-parallel way, of several massively data-parallel tasks. This work will be presented in Sac 2020 (we are currently working on the final version of the paper). This work is realised in collaboration with Pierre Leca and Wijnand Suijlen (Huawei Technologies), and Françoise Baude (Université Côte d'Azur, CNRS, I3S).

6.3. Locally abstract globally concrete semantics

Participant: Ludovic Henrio.

This research direction aims at designing a new way to write semantics for concurrent languages. The objective is to design semantics in a compositional way, where each primitive has a local behavior, and to adopt a style much closer to verification frameworks so that the design of an automatic verifier for the language is easier. The local semantics is expressed in a symbolic and abstract way, a global semantics gathers the abstract local traces and concretizes them. We have a reliable basis for the semantics of a simple language (a concurrent while language) and for a complex one (ABS), but the exact semantics and the methodology for writing it is still under development. After 2 meetings in 2019, A journal article is still being written but the visit of Reiner Hähnle in the Cash team during two months (as invited professor) in Spring 2019 should allow us to make faster progress on the topic.

This is a joint with Reiner Hähnle (TU Darmstadt), Einar Broch Johnsen, Crystal Chang Din, Lizeth Tapia Tarifa (Univ Oslo), Ka I Pun (Univ Oslo and Univ of applied science).

6.4. Memory consistency for heterogeneous systems

Participant: Ludovic Henrio.

Together with Christoph Kessler (Linköping University), we worked on the formalization of the cache coherency mechanism used in the VectorPU library developed at Linköping University. Running a program on disjoint memory spaces requires to address memory consistency issues and to perform transfers so that the program always accesses the right data. Several approaches exist to ensure the consistency of the memory accessed, we are interested here in the verification of a declarative approach where each component of a computation is annotated with an access mode declaring which part of the memory is read or written by the component. The programming framework uses the component annotations to guarantee the validity of the memory accesses. This is the mechanism used in VectorPU, a C++ library for programming CPU-GPU heterogeneous systems and this article proves the correctness of the software cache-coherence mechanism used in the library. Beyond the scope of VectorPU, this article can be considered as a simple and effective formalisation of memory consistency mechanisms based on the explicit declaration of the effect of each component on each memory space. This year, we have the following new results:

- we extended the work to support the manipulation of overlapping array. This was accepted as an extended version of our conference paper (presented at 4PAD 2018). It will be published in the JLAMP journal in 2020 [3].

6.5. PNets: Parametrized networks of automata

Participant: Ludovic Henrio.

pNets (parameterised networks of synchronised automata) are semantic objects for defining the semantics of composition operators and parallel systems. We have used pNets for the behavioral specification and verification of distributed components, and proved that open pNets (i.e. pNets with holes) were a good formalism to reason on operators and parameterized systems. This year, we have the following new results:

- A weak bisimulation theory for open pNets. This work is realized with Eric Madelaine (Inria Sophia-Antipolis) and Rabéa Ameur Boulifa (Telecom ParisTech). A journal article has been written and will be submitted in January 2020.
- A translation from BIP model to open pNets has being formalized and encoded, this work is done in collaboration with Simon Bliudze (Inria Lille). More precisely, we extend the theory of architectures developed previously for the BIP framework with the elements necessary for handling data: definition and operations on data domains, syntax and semantics of composition operators involving data transfer. To verify that individual architectures do enforce their associated properties , we provide an encoding into open pNets, an intermediate model that supports SMT-based verification. This work has been published in Coordination 2019 [6].

These works are under progress and should be continued in 2020.

6.6. Decidability results on the verification of phaser programs

Participant: Ludovic Henrio.

Together with Ahmed Rezine and Zeinab Ganjei (Linköping University) we investigated the possibility to analyze programs with phasers (a construct for synchronizing processes that generalizes locks, barrier, and publish-subscribe patterns). They work with signal and wait messages from the processes (comparing the number of wait and signal received to synchronize the processes). We proved that in many conditions, if the number of phasers or processes cannot be bounded, or if the difference between the number of signal and the number of wait signal is unbounded, then many reachability problems are undecidable. We also proposed fragments where these problems become decidable, and proposed an analysis algorithm in these cases. The results have been published in TACAS 2019 [11].

6.7. A Survey on Verified Reconfiguration

Participant: Ludovic Henrio.

We are conducting a survey on the use of formal methods to ensure safety of reconfiguration of distributed system, that is to say the runtime adaptation of a deployed distributed software system. The survey article is written together with Hélène Coullon and Simon Robillard (IMT Atlantique, Inria, LS2N, UBL), and Frédéric Loulergue (Northern Arizona University). Hélène Coullon is the coordinator and we expect the article to be submitted in 2020.

6.8. A Survey on Parallelism and Determinacy

Participants: Ludovic Henrio, Laure Gonnord, Matthieu Moy, Christophe Alias.

We have started to investigate on the solutions that exist to ensure complete or partial determinacy in parallel programs. The objective of this work is to provide a survey based on the different kinds of solutions that exist to ensure determinism or at least limit data-races in concurrent execution of programs. The study will cover language-based, compilation-based and also runtime-based solutions. We started the bibliographic studies in 2019. The objective of this work is to write and submit a survey article in 2020.

This work, coordinated by Laure Gonnord and Ludovic Henrio, also involves contributors outside the CASH team. For the moment Gabriel Radanne (Inria Paris) and Lionel Morel (CEA).

6.9. Pipeline-aware Scheduling of Polyhedral Process Networks

Participants: Christophe Alias, Julien Rudeau.

The polyhedral model is a well known framework to develop accurate and optimal automatic parallelizers for high-performance computing kernels. It is progressively migrating to high-level synthesis through polyhedral process networks (PPN), a dataflow model of computation which serves as intermediate representation for high-level synthesis. Many locks must be overcome before having a fully working polyhedral HLS tool, both from a front-end (C $\rightarrow$ PPN) and back-end (PPN $\rightarrow$ FPGA) perspective. In this work [15], we propose a front-end scheduling algorithm which reorganizes the computation of processes to maximize the pipeline efficiency of the processes' arithmetic operators. We show that our approach improve significantly the overall latency as well as the pipeline efficiency.

6.10. A Compiler Algorithm to Guide Runtime Scheduling

Participants: Christophe Alias, Samuel Thibault, Laure Gonnord.

Task-level parallelism is usually exploited by a runtime scheduler, after tasks are mapped to processing units by a compiler. In this report, we propose a compilation-centric runtime scheduling strategy. We propose a complete compilation algorithm to split the tasks in three parts, whose properties are intended to help the scheduler to take the right decisions [16]. In particular, we show how the polyhedral model may provide a precious help to compute tricky scheduling and parallelism informations. Our compiler is available and may be tried online at http://foobar.ens-lyon.fr/kut.

This is a joint work with University of Bordeaux, which will be continued next year.

6.11. fkcc: the Farkas Calculator

Participant: Christophe Alias.

We propose a new domain-specific language and a tool, FKCC, to prototype program analyses and transformations exploiting the affine form of Farkas lemma. Our language is general enough to prototype in a few lines sophisticated termination and scheduling algorithms. The tool is freely available and may be tried online via a web interface. We believe that FKCC is the missing chain to accelerate the development of program analyses and transformations exploiting the affine form of Farkas lemma.

This work has been presented in the TAPAS'19 workshop [13] and will be presented at the IMPACT'20 workshop [13].

6.12. Standard-compliant Parallel SystemC simulation of Loosely-Timed Transaction Level Models

Participant: Matthieu Moy.

To face the growing complexity of System-on-Chips (SoCs) and their tight time-tomarket constraints, Virtual Prototyping (VP) tools based on SystemC/TLM must get faster while keeping accuracy. However, the Accellera SystemC reference implementation remains sequential and cannot leverage the multiple cores of modern workstations. In this paper, we present a new implementation of a parallel and standard-compliant SystemC kernel, reaching unprecedented performances. By coupling a parallel SystemC kernel and memory access monitoring, we are able to keep SystemC atomic thread evaluation while leveraging the available host cores. Evaluations show a ×19 speed-up compared to the Accellera SystemC kernel using 33 host cores reaching speeds above 2000 Million simulated Instructions Per Second (MIPS).

This work will be published at the ASP-DAC 2020 conference.

6.13. Response time analysis of dataflow applications on a many-core processor with shared-memory and network-on-chip

Participant: Matthieu Moy.

We consider hard real-time applications running on many-core processor containing several clusters of cores linked by a Network-on-Chip (NoC). Communications are done via shared memory within a cluster and through the NoC for inter-cluster communication. We adopt the time-triggered paradigm, which is well-suited for hard real-time applications, and we consider data-flow applications, where communications are explicit.

We extend the AER (Acquisition/Execution/Restitution) execution model to account for all delays and interferences linked to communications, including the interference between the NoC interface and the memory. Indeed, for NoC communications, data is first read from the initiator's local memory, then sent over the NoC, and finally written to the local memory of the target cluster. Read and write accesses to transfer data between local memories may interfere with shared-memory communication inside a cluster, and, as far as we know, previous work did not take these interferences into account.

Building on previous work on deterministic network calculus and shared memory interference analysis, our method computes a static, time-triggered schedule for an application mapped on several clusters. This schedule guarantees that deadlines are met, and therefore provides a safe upper bound to the global worst-case response time.

This work was published at RTNS 2019 [14].

6.14. Smart placement of dynamically allocated objects for heterogeneous memory

Participant: Matthieu Moy.

As part of a partnership with the CITI laboratory (Tristan Delizy's PhD, co-supervised with Guillaume Salagnac and Tanguy Risset), we worked on dynamic memory memory allocation for embedded systems with heterogeneous memory. Unlike cache-based systems, our target architecture exposes several memory banks with different performance characteristics directly to the software, without any hardware mechanism like a cache or an MMU for memory management. The software needs to chose which memory bank to use at allocation time, and cannot change this choice afterwards. We proposed a profiling-based placement policy that is shown to be near-optimal for several applications, and performs much better than naive placement policies especially for systems with a small fraction of fast memory.

This work documented as part of Tristan Delizy's Ph.D manuscript, and we plan to submit it for a journal publication in 2020.

6.15. Static Analysis Of Binary Code With Memory Indirections Using Polyhedra

Participant: Laure Gonnord.

Together with Clement Ballabriga, Julien Forget, Giuseppe Lipari, and Jordy Ruiz (University of Lille), we proposed in 2018 a new abstract domain for static analysis of binary code. Our motivation stems from the need to improve the precision of the estimation of the Worst-Case Execution Time (WCET) of safety-critical real-time code. WCET estimation requires computing information such as upper bounds on the number of loop iterations, unfeasible execution paths, etc. These estimations are usually performed on binary code, mainly to avoid making assumptions on how the compiler works. Our abstract domain, based on polyhedra and on two mapping functions that associate polyhedra variables with registers and memory, targets the precise computation of such information. We prove the correctness of the method, and demonstrate its effectiveness on benchmarks and examples from typical embedded code.

The results have been presented to VMCAI'19 on Model Checking and Abstract Interpretation [5] and has received the best paper award of the conference.

6.16. Polyhedral Value Analysis as Fast Abstract Interpretation

Participant: Laure Gonnord.

Together with Tobias Grosser, (ETH Zurich, Switzerland), Siddhart Bhat, (IIIT Hydrabad, India), Marcin Copik (ETH Zurich, Switzerland), Sven Verdoolaege (Polly Labs, Belgium) and Torsten Hoefler (ETH Zurich, Switzerland), we tried to bridge the gap between the well founded classical abstract interpretation techniques and their usage in production compilers.

We formulate the polyhedral value analysis (a classical algorithm in production compilers like LLVM, scalar evolution based on Presburger set as abstract interpretation), and rephrase a complete value and validity

In 2019, the formalisation has been rephrased in a simplier way and extented to deal with more llvm-related semantics (undefined behavior, poisoned values) and we started a collaboration with David Monniaux, Verimag, on this topic.

The paper is being rewritten and we are also writing a project on which we would extend our method to mode complex polyhedral transformations in a context of formally verified tools.

6.17. Decision results for solving Horn Clauses with arrays

Participants: Laure Gonnord, Julien Braine.

Many approaches exist for verifying programs operating on Boolean and integer values (e.g. abstract interpretation, counterexample-guided abstraction refinement using interpolants), but transposing them to array properties has been fraught with difficulties. In the context of the Phd of Julien Braine, we propose to work directly on horn clauses, because we think that it is a suitable intermediate representation for verifying programs.

Currently, two techniques strike out to infer very precise quantified invariants on arrays using Horn clauses: a quantifier instantiation method [1] and a cell abstraction method that can be rephrased on Horn clauses. However, the quantifier instantiation method is parametrized by an heuristic and finding a good heuristic is a major challenge, and the cell abstraction method uses an abstract interpretation to completely remove arrays and is limited to linear Horn clauses. We combine these two techniques. We provide an heuristic for the quantifier instantiation method of [29] by using the ideas from the cell abstraction method of [48] and discover a requirement such that, when met, the heuristic is complete, that is, there is no loss of information by using that heuristic. Furthermore, we prove that Horn clauses that come from program semantic translation verify the requirement and therefore, we have an optimal instantiation technique for program analysis.

This work is done in collaboration with David Monniaux (Verimag), coadvisor of the PhD of Julien Braine. A journal paper is currently being written for submission early 2020.

6.18. Scheduling Trees

Participants: Laure Gonnord, Paul Iannetta.

As a first step to schedule non polyhedral computation kernels, we investigated the tree datastructure. A large bibliography on tree algorithmics and complexity leds us to chose to work on balanced binary trees, for which we have designed algorithms to change their memory layout into adjacent arrays. We rephrased the classical algorithms (construction, search, destruction ...) in this setting, and implemented them in C.

The conclusion of this study is unfortunately negative : the locality gain in transforming trees into linear structures is not contrabalanced by a better cache usage, all our codes have been slowed down in the process. Our experiments are still in progress, but our hypothesis is that our trees are too sparse to be more clever that the *malloc* implementation.

A research paper will be published early 2020. This work is done in collaboration with Lionel Morel (CEA Grenoble), coadvisor of the PhD of Paul Iannetta.

6.19. Formalisation of the Polyhedral Model

Participants: Laure Gonnord, Paul Iannetta.

Last year, together with Lionel Morel (Insa/CEA) and Tomofumi Yuki (Inria, Rennes), we revisited the polyhedral model's key analysis, dependency analysis, published in a research report [44]. This year we pursued in this direction. We have now a better formalisation, and a better understanding of the expressivity and applicability.

We still have one step to study in order to be able to have a full semantic polyhedral model: properly formalise code scheduling and code generation within our semantic model.

This work is made in collaboration with Lionel Morel (CEA Grenoble) who coadvise Paul Iannetta.

6.20. Semantics diffs in LLVM

Participants: Laure Gonnord, Matthieu Moy.

Laure Gonnord and Matthieu Moy have coadvised a Master research Project ("TER") early in 2019 , whose objective was to study the LLVM LLVM compiler infrastructure with software engineering techniques in order to characterise how sequences of code analyses and transformations behave. The project has lead to a sequence of tools to evaluate experimentally how a sequence of passes influence performance.

Laure Gonnord and Matthieu Moy have, together with Sebastien Mosser, coadvised a second internship at UQAM for three months, between May and July 2019. During his internship, Sebastien Michelland has demonstrated that textual diffs are not sufficient to fully characterise the behaviours of code transformation inside compilers. He analysed llvm-diff, a tool of the distribution that makes an analysis at the intermediate representation level, and gives first hints to define a proper notion of semantic diff for this application.

For these interships two research reports have been produced.

This work was done in the context of an ongoing collaboration with Sebastien Mosser, previously in Nice, and now at UQAM. An Inria associate team was proposed for 2020-2023 on similar topics.

CONVECS Project-Team

7. New Results

7.1. New Formal Languages and their Implementations

7.1.1. LOTOS and LNT Specification Languages

Participants: Hubert Garavel, Frédéric Lang, Wendelin Serwe.

LNT [6] [31] is a next-generation formal description language for asynchronous concurrent systems. The design of LNT at CONVECS is the continuation of the efforts undertaken in the 80s to define sound languages for concurrency theory and, indeed, LNT is derived from the ISO standards LOTOS (1989) and E-LOTOS (2001). In a nutshell, LNT attempts to combine the best features of imperative programming languages, functional languages, and value-passing process calculi.

LNT is not a frozen language: its definition started in 2005, as part of an industrial project. Since 2010, LNT has been systematically used by CONVECS for numerous case studies (many of which being industrial applications — see § 7.5). LNT is also used as a back-end by other research teams who implement various languages by translation to LNT. It is taught in university courses, e.g., at University Grenoble Alpes and ENSIMAG, where it is positively accepted by students and industry engineers. Based on the feedback acquired by CONVECS, LNT is continuously improved.

In 2019, a new option `-depend` has been added to the LNT_DEPEND, LNT2LOTOS, and LNT.OPEN tools. LNT_DEPEND now supports the case where the user replaces the predefined LNT modules (e.g., BOOLEAN, NATURAL, etc.) with custom versions. LNT_DEPEND has been made faster and displays better error messages. The LOTOS code generated by LNT2LOTOS for parallel compositions could be semantically incorrect and has been fixed.

We continued working on the TRAIAN compiler for the LOTOS NT language (a predecessor of LNT), which is used for the construction of most CADP compilers and translators.

The version 2.x of TRAIAN that we have been developing for almost 20 years is increasingly difficult to maintain. It consists of a large collection of attribute grammars and is built using the FNC-2 compiler generation system, which is no longer supported. For this reason, TRAIAN 2.x only exists in a 32-bit version, and sometimes hits the 4 GB RAM limit when dealing with large compiler specifications, such as those of LNT2LOTOS or EVALUATOR 5.

For this reason, we undertook in 2018 a complete rewrite of TRAIAN (version 3.0) to get rid of FNC-2. Two main design decisions behind TRAIAN 3.0 are the following: (i) it supports (most of) the LOTOS NT language currently accepted by TRAIAN 2.x, but also extensions belonging to LNT, so as to allow a future migration from LOTOS NT to LNT; and (ii) TRAIAN 3.0 is currently written in LOTOS NT and compiled using TRAIAN 2.x, but should be ultimately capable of bootstrapping itself.

In 2019, we continued the development of TRAIAN 3.0, whose grammar and syntax analysis phase was already almost complete. We fully implemented several static program analysis phases, among which the following:

- binding analysis, which associates a declaration to every identifier occurring in the program (e.g., type, channel, variable, event, etc.)
- typing analysis (including resolution of function name overloading), which associates a type to every expression in the program
- type-productivity and type-finiteness analysis, which check respectively whether a type has at least one value and whether a type has a finite number of values

We also fully implemented the C function generation phase and started to implement the C type generation phase. To avoid problems when switching from TRAIAN 2.x to TRAIAN 3.0, TRAIAN 3.0 generates almost exactly the same code as TRAIAN 2.x. The principal differences concern the numbers used to uniquely identify symbols (variables and functions) in the generated C code, because these are often derived from the syntax tree.

TRAIAN 3.0 is checked regularly against a non-regression test suite consisting of 845 correct and 1545 incorrect programs.

In total, the functionalities that remain to be implemented in TRAIAN 3.0 represent less than 32% of the code of TRAIAN 2.x.

7.1.2. Nested-Unit Petri Nets

Participants: Pierre Bouvier, Hubert Garavel.

Nested-Unit Petri Nets (NUPNs) is a model of computation that can be seen as an upward-compatible extension of P/T nets, which are enriched with structural information on their concurrent and hierarchical structure. Such structural information can easily be produced when NUPNs are generated from higher-level specifications (e.g., process calculi) and allows logarithmic reductions in the number of bits required to represent reachable states, thus enabling verification tools to perform better. For this reason, NUPNs have been so far implemented in thirteen verification tools developed in four countries, and adopted by two international competitions (the Model Checking Contest and the Rigorous Examination of Reactive Systems challenge).

In 2019, a journal article [13] has been published, which formalizes the complete theory of NUPNs.

The development of software tools for NUPNs has steadily progressed. The file format for NUPNs has been enhanced and made more precise; the NUPN_INFO tool has been extended with two new options; the CAESAR.BDD tool as been extended with six new options and its capabilities and efficiency improved in many respects.

We also revisited the problem of decomposing a Petri net into a network of automata, a problem that has been around since the early 70s. We reformulated this problem as the transformation of an ordinary, one-safe Petri net into a unit-safe NUPN. We developed various transformation methods, all of which we implemented in a tool chain that combines NUPN tools with third-party software, such as SAT solvers, SMT solvers, and tools for graph colouring and finding maximal cliques. We performed an extensive evaluation of these methods on a collection of more than 12,000 nets from diverse sources, including nets whose marking graph is too large for being explored exhaustively.

7.1.3. Formal Modeling and Analysis of BPMN

Participant: Gwen Salaün.

A business process is a set of structured activities that provide a certain service or product. Business processes can be modeled using the BPMN (*Business Process Model and Notation*) standard, and several industrial platforms have been developed for supporting their design, modeling, and simulation.

In collaboration with Francisco Durán (University of Málaga, Spain) and Camilo Rocha (University of Cali, Colombia), we proposed an approach for the modeling and analysis of resource allocation for business processes. Our approach enables the automatic computation of measures for precisely identifying and optimizing the allocation of resources in business processes, including resource usage over time. The proposed analysis, especially suited to support decision-making strategies, is illustrated with a case study of a parcel ordering and delivery by a fleet of drones. This work comprises an encoding of a significant and expressive subset of BPMN in rewriting logic, an executable logic of concurrent change that can naturally deal with states and concurrent computations. The encoding is by itself a formal semantics and interpreter of the BPMN subset that captures all concurrent behavior and thus is used to simulate the concurrent evolution of any business process with a given number of resources and replicas. This work led to two publications, in an international conference [19] and an international journal [12].

7.2. Parallel and Distributed Verification

7.2.1. Debugging of Concurrent Systems using Counterexample Analysis

Participant: Gwen Salaün.

Model checking is an established technique for automatically verifying that a model satisfies a given temporal property. When the model violates the property, the model checker returns a counterexample, which is a sequence of actions leading to a state where the property is not satisfied. Understanding this counterexample for debugging the specification is a complicated task for several reasons: (i) the counterexample can contain hundreds of actions, (ii) the debugging task is mostly achieved manually, (iii) the counterexample does not explicitly highlight the source of the bug that is hidden in the model, (iv) the most relevant actions are not highlighted in the counterexample, and (v) the counterexample does not give a global view of the problem.

We proposed a novel approach to improve the usability of model checking by simplifying the comprehension of counterexamples. Our approach takes as input an LTS model and an (unsatisfied) temporal logic property, and operates in four steps. First, all counterexamples for the property are extracted from the model. Second, the model is analyzed to identify the actions that skip from correct to incorrect behaviours (intuitively, these are the most relevant actions from a debugging perspective). Third, using a panel of abstraction techniques, these actions are extracted from the counterexamples. Fourth, 3D visualization techniques are used for highlighting specific regions in the model, where a choice is possible between executing a correct behaviour or falling into an erroneous part of the model, according to the property under analysis. We developed a tool named CLEAR to fully automate our approach, and we applied it on real-world case studies from various application areas for evaluation purposes. This allowed us to identify several patterns corresponding to typical cases of errors (e.g., interleaving, iteration, causality, etc.).

These results led to two publications in international conferences [15] [16] and a publication to appear in an international journal [11].

7.2.2. Eliminating Data Races in Parallel Programs using Model Checking

Participants: Radu Mateescu, Wendelin Serwe.

Parallelization of existing sequential programs to increase their performance and exploit recent multi- and many-core architectures is a challenging but inevitable effort. One increasingly popular parallelization approach is based on OpenMP, which enables the designer to annotate a sequential program with constructs specifying the parallel execution of code blocks. These constructs are then interpreted by the OpenMP compiler and runtime, which assigns blocks to threads running on a parallel architecture. Although this scheme is very flexible and not (very) intrusive, it does not prevent the occurrence of synchronization errors (e.g., deadlocks) or data races on shared variables.

In the framework of the CAPHCA project (see § 9.2.1.1), in collaboration with Eric Jenn and Viet Anh Nguyen (IRT Saint-Exupéry, Toulouse), we proposed an iterative method to assist the OpenMP parallelization by using formal methods and verification. In each iteration, potential data races are identified by applying to the OpenMP program a lockset analysis, which computes the set of shared variables that potentially need to be protected by locks. To avoid the insertion of superfluous locks, an abstract, action-based formal model of the OpenMP program in LNT is extracted and analyzed using the EVALUATOR model checker of CADP. Spurious locks are detected by checking ACTL formulas expressing the absence of concurrent execution of shared variables accesses. This work led to an international publication [28].

7.3. Timed, Probabilistic, and Stochastic Extensions

7.3.1. On-the-fly Model Checking for Extended Regular Probabilistic Operators

Participants: Armen Inants, Radu Mateescu.

Specifying and verifying quantitative properties of concurrent systems requires expressive and user-friendly property languages combining temporal, data-handling, and quantitative aspects. To this aim, we undertook the quantitative analysis of concurrent systems modeled as PTSs (*Probabilistic Transition Systems*), whose actions contain channel names, data values, and probabilities. We proposed a new regular probabilistic operator that extends naturally the Until operators of PCTL (*Probabilistic Computation Tree Logic*) [41], by specifying the probability measure of a path characterized by a generalized regular formula involving arbitrary computations on data values. We devised an on-the-fly model checking method for this new operator, based on a combined local resolution of linear and Boolean equation systems.

In 2019, we continued this work as follows:

- The MCL v4 language was conservatively extended with the new probabilistic operator, leading to a new version MCL v5.
- A new version 5 of the EVALUATOR model checker that handles the MCL v5 language, was added to CADP. EVALUATOR 5 is backward compatible with EVALUATOR 4, to which it adds a new option "`-epsilon`" specifying the precision of floating-point computations. A new version 5 of the MCL_EXPAND tool, the front-end common to the EVALUATOR 3, 4, and 5 model checkers, was added to CADP. This version is upward compatible with the previous one (except for slight changes in some error messages), it corrects a bug and brings some optimizations in the C code generated. Two new manual pages "evaluator5" and "mcl5" have been added.
- For certain probabilistic formulas (e.g., expressing the step-bounded reachability of events), the on-the-fly model checking procedure can be optimized by taking advantage of the possible *query containments*, i.e., implications between instances of the formula with different data parameters. We studied query containment in DHMLR (*Data-based Hennessy-Milner Logic with Recursion*), a parameterized equational formalism used as intermediate language for model checking MCL formulas. Our method consists in detecting, by static analysis, the containment orders present in the DHMLR representation of an MCL formula, and using the information about parameterized Boolean variable implications to improve the convergence of the BES resolution algorithms. We implemented the method in a prototype extension of EVALUATOR 5 and of the CAESAR_SOLVE library for BES resolution, and applied it for verifying probabilistic and also functional properties (e.g., bounded inevitability). The experiments we carried out on self-stabilizing protocols and communication protocols over unreliable channels showed reductions of up to 50% in memory and up to 33% in execution time. This work led to a paper submitted to an international conference.

7.4. Component-Based Architectures for On-the-Fly Verification

7.4.1. Compositional Verification

Participants: Frédéric Lang, Radu Mateescu.

The CADP toolbox contains various tools dedicated to compositional verification, among which EXP.OPEN, BCG_MIN, BCG_CMP, and SVL play a central role. EXP.OPEN explores on the fly the graph corresponding to a network of communicating automata (represented as a set of BCG files). BCG_MIN and BCG_CMP respectively minimize and compare behavior graphs modulo strong or branching bisimulation and their stochastic extensions. SVL (*Script Verification Language*) is both a high-level language for expressing complex verification scenarios and a compiler dedicated to this language.

In 2019, in addition to small bug corrections, we updated SVL to support version 5 of EVALUATOR, and we corrected a semantic bug in the expansion of meta-operators of SVL.

In collaboration with Franco Mazzanti (ISTI-CNR, Pisa, Italy), we also used the compositional verification tools of CADP in the framework of the RERS'2019 challenge [0], which consisted in verifying 180 LTL properties and 180 CTL properties on large models of concurrent systems having up to 70 concurrent processes and 234 synchronization actions.

[0] http://rers-challenge.org/2019

We applied to these examples the *maximal hiding* technique [48], which consists in hiding in the model all actions that are not necessary to verify the property. We combined this technique with compositional minimization (using the smart reduction heuristic implemented in SVL) as follows:

- In a first attempt, we used the technique consisting in applying minimization modulo either strong bisimulation or divbranching (divergence-preserving branching) bisimulation, depending on the fragment of the modal μ-calculus to which the formula belongs, as proposed in [48]. This was more efficient than non-compositional verification on large models, but not sufficient to verify all RERS problems successfully.
- We then proposed a refinement of this approach, which consists in (1) partitioning the actions of the system to be verified into so-called strong and weak actions, depending on the formula, and (2) minimizing modulo divbranching bisimulation all processes and process compositions containing weak actions only. This is an improvement over the previous technique, since divbranching bisimulation can be used to minimize some processes of the system even though the formula does not belong to the fragment of the μ-calculus adequate with divbranching bisimulation (which corresponds to formulas with an empty set of strong actions). This new technique allowed us to verify a lot more problems successfully, but still letting a few of the largest RERS problems unresolved. We published a paper describing the approach in an international conference [23].
- At last, we designed a new bisimulation relation, named *sharp bisimulation*, parameterized by the strong actions of the system, and we implemented a prototype tool that reduces a behavior graph modulo this relation. Sharp bisimulation parameterized by a set S of strong actions is weaker than strong bisimulation, stronger than divbranching bisimulation, and adequate with formulas whose strong actions are included in S. Such a fine-tuning of the bisimulation relation by strong actions allowed us to verify all RERS problems successfully and to win the 2019 challenge. A paper describing the approach was accepted for publication in an international conference.

7.4.2. Other Component Developments

Participants: Hubert Garavel, Frédéric Lang, Philippe Ledent, Radu Mateescu, Wendelin Serwe.

In 2019, several components of CADP have been improved as follows:

- We enhanced the TESTOR tool by adding the possibility to interact with an SUT (*System Under Test*) using its standard input and output.
- We enhanced the XTL compiler with a function converting a transition label into a string (useful for handling the entire content of the label), and we also corrected three bugs.
- We enhanced MCL_EXPAND 5 with a better detection of nondeterminism in probabilistic formulas and a vacuity check for infinite looping operators, and we also corrected a semantic bug.
- We enhanced EVALUATOR 5 with more explanative messages about the assignment of probabilities to transitions, and we corrected two bugs in each of the tools EVALUATOR 4 and 5.
- The C code generated by CAESAR has been modified to suppress GCC 6.5 warnings.
- Several changes have been brought to CADP to enable its use on new platforms, including macOS 10.15 "Catalina" and the forthcoming Debian 10.0 Linux distribution. Various bugs specific to Linux and SunOS systems (Solaris or Illumos/OpenIndiana) have been fixed.

7.5. Real-Life Applications and Case Studies

7.5.1. Autonomous Resilience of Distributed IoT Applications in a Fog Environment

Participants: Umar Ozeer, Gwen Salaün.

Recent computing trends have been advocating for more distributed paradigms, namely Fog computing, which extends the capacities of the Cloud at the edge of the network, that is close to end devices and end users in the physical world. The Fog is a key enabler of the Internet of Things (IoT) applications as it resolves some of the needs that the Cloud fails to provide such as low network latencies, privacy, QoS, and geographical requirements. For this reason, the Fog has become increasingly popular and finds application in many fields such as smart homes and cities, agriculture, healthcare, transportation, etc.

The Fog, however, is unstable because it is constituted of billions of heterogeneous devices in a dynamic ecosystem. IoT devices may regularly fail because of bulk production and cheap design. Moreover, the Fog-IoT ecosystem is cyber-physical and thus devices are subjected to external physical world conditions, which increase the occurrence of failures. When failures occur in such an ecosystem, the resulting inconsistencies in the application affect the physical world by inducing hazardous and costly situations.

In the framework of the collaboration with Orange Labs (see § 8.1.1), we proposed an end-to-end autonomic failure management approach for IoT applications deployed in the Fog. The proposed approach recovers from failures in a cyber-physical consistent way. Cyber-physical consistency aims at maintaining a consistent behavior of the application with respect to the physical world, as well as avoiding dangerous and costly circumstances. The approach was validated using model checking techniques to verify important correctness properties. It was then implemented as a framework called F3ARIoT. This framework was evaluated on a smart home application. The results showed the feasibility of deploying F3ARIoT on real Fog-IoT applications as well as its good performances in regards to end user experience.

These results have been published in U. Ozeer's PhD thesis [10] and at an international conference [26]. Another paper was submitted to an international journal.

7.5.2. *Verified Composition and Deployment of IoT Applications*

Participants: Alejandro Martinez Rivero, Radu Mateescu, Ajay Muroor Nadumane, Gwen Salaün.

The Internet of Things (IoT) applications are built by interconnecting everyday objects over internet. As IoT is becoming popular among consumers, the challenge of making IoT applications easy to design and deploy is more relevant than ever. In 2019, we considered this challenge along two perspectives.

- In the framework of the collaboration with Nokia Bell Labs (see § 8.1.2), we focused on helping consumers to easily design IoT applications that are correct, and also support the deployment of these applications. The correctness of the applications is ensured through formal methods and verification techniques.

  Using W3C Web of Things (WoT) specification as the basis of our work, we extended the specification of objects in WoT with a behavioural model. This allows us to describe formally the composition of objects and thus, to verify their behavioural correctness. Typically, an IoT application is defined using Event-Condition-Action (ECA) rules of the type "IF event THEN action". Our work supports users to specify not only the ECA rules, but also the composition of rules using a simple, yet expressive language. This makes possible the construction of advanced compositions, which would have been hard or sometimes impossible to build using simple ECA rules. Finally, users are provided with an easy-to-deploy solution for these advanced compositions. All these steps were implemented and packaged in a tool named MozART, built on top of Mozilla WebThings platform. LNT is used as the formal specification language, and various tools of CADP are used for verifying the composition. Also, an execution engine based on Mozilla WebThings API was built to support the deployment of advanced compositions. The work has led to the preparation of two conference articles.
- Building IoT applications of added-value from a set of available devices with minimal human intervention is one of the main challenges facing the IoT. This is a difficult task that requires models for specifying objects, in addition to user-friendly and reliable composition techniques which in turn prevent the design of erroneous applications.

  In collaboration with Francisco Durán (University of Málaga, Spain), we tackled this problem by first describing IoT applications using abstract models obtained from existing models of concrete devices. Then, we proposed automated techniques for building compositions of devices using a repository of available devices, and an abstract goal of what the user expects from such compositions. Since the number of possible solutions can be quite high, we used both filtering and ranking techniques to provide the most relevant solutions to users. The provided solutions satisfy the given goal and may be analysed with respect to properties such as deadlock-freeness or unmatched send messages. Finally, the application can be deployed using existing execution engines. This work led to a publication in an international conference [20].

7.5.3. Autonomous Car

Participants: Philippe Ledent, Lina Marsso, Radu Mateescu, Wendelin Serwe.

Autonomous vehicles are complex cyber-physical systems that must satisfy critical correctness requirements to increase the safety of road traffic. The validation of autonomous driving is a challenging field because of the complexity of its key components (perception of the environment, scene interpretation, decision making and undertaking of actions) and the intertwinning of physical and software components. In 2019, we considered this challenge along two lines of work.

- From the embedded software perspective, autonomous cars can be considered as GALS systems, which integrate reactive synchronous components that interact asynchronously. The complexity induced by combining synchronous and asynchronous aspects makes GALS systems difficult to develop and debug.

  In the framework of the ARC6 collaboration (see § 9.1.1), we proposed a testing methodology for GALS systems that leverages conformance test generation for asynchronous systems to automatically derive realistic scenarios (inputs constraints and oracles), which are necessary ingredients for the unit testing of individual synchronous components, and are difficult and error-prone to design manually. The methodology consists of several steps (derivation of asynchronous test cases from a GALS model and a test purpose, projection of the complete test graph on a synchronous component, extraction and execution of test scenarios) and was illustrated on a simple, but relevant example inspired by autonomous cars. These results were published in L. Marsso's PhD thesis [9] and at an international conference [25].

- In collaboration with Christian Laugier, Anshul Paigwar, and Alessandro Renzaglia (CHROMA project-team), we proposed a new approach where formal verification is employed to validate systems with probabilistic predictions. We focused on the risk assessment generated by CMCDOT (*Conditional Monte Carlo Dense Occupancy Tracker*), a probabilistic perception system for autonomous cars. CMCDOT provides an environment representation through Bayesian probabilistic occupancy grids and estimates Time-to-Collision probabilities for every static and dynamic part of the grid in the near future. To validate the probabilistic collision risk estimation, we used the CARLA simulator to generate a large number of realistic intersection crossing scenarios with two vehicles. The set of scenarios is then validated using the XTL model checker, by defining appropriate KPIs (*Key Performance Indicators*) as temporal logic formulas and also performing a quantitative analysis. This work led to a publication at an international conference [24].

CORSE Project-Team

6. New Results

6.1. Compiler Optimizations and Analysis

Participants: Fabrice Rastello, Manuel Selva, Fabian Grüber, Diogo Sampaio [CORSE, Inria], Christophe Guillon [STMicroelectronics], P. Sadayappan [OSU, USA], Louis-Noël Pouchet [CSU, USA], Atanas Rountev [OSU, USA], Richard Veras [LSU, USA], Rui Li [UoU, USA], Aravind Sukumaran-Rajam [OSU, USA], Tse Meng Low [CMU, USA].

Our current efforts with regard to code optimization follows two directions. 1. The first consists in improving compiler optimization techniques by considering pattern specific applications such as those related to machine learning. Our first result presented at SC 2019 [10] focuses on tensor contractions. 2. The second consists in developing dynamic analysis based performance debugging tools. Our first results published at PPoPP 2019 [9] and TACO 2020 [7] shows a scalable approach that compresses an execution trace obtained from binary instrumentation and analyses it using a polyhedral compiler.

6.1.1. Analytical Cache Modeling and Tilesize Optimization for Tensor Contractions

Data movement between processor and memory hierarchy is a fundamental bottleneck that limits the performance of many applications on modern computer architectures. Tiling and loop permutation are key techniques for improving data locality. However, selecting effective tile-sizes and loop permutations is particularly challenging for tensor contractions due to the large number of loops. Even state-of-the-art compilers usually produce sub-optimal tile-sizes and loop permutations, as they rely on naïve cost models. In this work we provide an analytical model based approach to multilevel tile size optimization and permutation selection for tensor contractions. Our experimental results show that this approach achieves comparable or better performance than state-of-the-art frameworks and libraries for tensor contractions.

This work is the fruit of the collaboration 8.3.1.1 with OSU. It has been presented at ACM/IEEE International Conference for High Performance Computing, Networking, Storage, and Analysis, SC 2019 [10].

6.1.2. Profiling-based Polyhedral Optimization Feedback

This work addresses the problem of reconstructing a compact (static) representation of a binary execution, automatically detecting hot regions and enabling precise feedback about optimization opportunities potentially missed by the compiler. Our framework handles codes with irregular accesses, pointers with indirections, inter-procedural or recursive loop regions. By enabling binary execution analysis we are able to discover runtime properties (i.e., the ability to form a compact representation) as well as inter-procedural optimization opportunities that cannot be uncovered by standard static analyses. Our design choices were driven towards achieving portability, both in terms of targeted architecture, but also in terms of programming environment (e.g., being robust to arbitrary programming language, compiler, use of third-party binaries, etc.).

A compact and yet precise inter-procedural dynamic dependence graph (DDG) is first computed via: 1. a new instrumentation framework based on QEMU; 2. the use of a new concept of inter-procedural loop-nesting tree; 3. followed by new techniques we introduce for folding, clamping, and widening of the DDG to agglomerate dynamic dependence instances into polyhedra of integer points whenever possible. State-of-the-art polyhedral analysis and transformation systems we specifically modified to provide useful feedback to the user is then used. We extensively evaluate our tool on numerous benchmarks, demonstrating the pratical usefulness of our tool-chain.

This work is the fruit of the collaboration 8.3.1.1 with OSU and and the past collaboration Nano2017 with STMicroelectronics. The main contributions has been presented at the ACM conference on Principles and Practice of Parallel Programming, PPoPP 2019 [9]. The new techniques that allow to build the polyhedral representation from the instrumented execution in a scalable way lead to a separate publication in the ACM Transactions on Architecture and Code Optimization, TACO 2020 [7].

6.2. Extraction of Periodic Patterns of Scientific Applications to Identify DVFS Opportunities

Participants: Mathieu Stoffel, François Broquedis, Frederic Desprez, Abdelhafid Mazouz [Atos/Bull], Philippe Rols [Atos/Bull].

Mathieu Stoffel started his PhD in February 2018 on a CIFRE contract with Atos/Bull. The purpose of this work is to enhance the energy consumption of HPC applications on large-scale platforms. The first phase of the thesis project consists in an in-depth study of the evolution of the metrics characterizing the state of the supercomputer during the execution of a highly parallel application. Indeed, the utilization rates of the different components of the HPC system may demonstrate extreme variations during the execution of the aforementioned application. These variations are sometimes subject to repeat themselves on a regular basis during the application execution. We refer to this phenomena as application "phases". In this context, we developed a tool suite resorting to fine-grain profiling and periodicity analysis to identify optimization opportunities for both performance and power-efficiency. It leverages the fact that a large share of HPC parallel applications are constituted of a restrained set of compute kernels executed a huge number of times to extract periodic patterns representative of the aforementioned kernels. By doing so, our tool offers a simple and condensed proxy to analyze and predict the behavior of complex parallel applications. For instance, we were able to identify and extract periodic patterns for a panel of reference HPC applications such as NAMD and NEMO. Then, as an example of the many ways to exploit the aforementioned extracted periodic patterns, we evaluated the impact of the CPU frequency on the latter. As a result, we were able to identify DVFS opportunities we plan to exploit in a future work.

6.3. Runtime Monitoring, Verification, and Enforcement

Participants: Antoine El-Hokayem [Univ. Grenoble Alpes, Verimag], Yliès Falcone, Thierry Jéron [Inria Rennes], Ali Kassem, Hervé Marchand [Inria Rennes], Srinivas Pinisetty [IIT Bhubaneswar], Matthieu Renard [Foxi], Antoine Rollet [Université de Bordeaux], César Sànchez [IMDEA Madrid], Gerardo Schneider [University of Gothenborg].

Our contributions in the domain of runtime monitoring, verification, and enforcement are threefold. First, we contributed to the publication of general papers aimed to structure the community by publishing a tutorial on runtime enforcement of timed properties [16], a review of the first five years of the competition on runtime verification [15] and a survey of future challenges of runtime verification [6]. We also concluded some other previous work by realizing journal publications on the topics of decentralized runtime verification [3] and on runtime enforcement of timed properties [5]. We started a new activity on monitoring for security properties, and more particularly on the detection of fault-injection attacks [12].

6.3.1. On the Runtime Enforcement of Timed Properties

This work [16] is concerned with runtime enforcement which refers to the theories, techniques, and tools for enforcing correct behavior of systems at runtime. We are interested in such behaviors described by specifications that feature timing constraints formalized in what is generally referred to as timed properties. This tutorial presents a gentle introduction to runtime enforcement (of timed properties). First, we present a taxonomy of the main principles and concepts involved in runtime enforcement. Then, we give a brief overview of a line of research on theoretical runtime enforcement where timed properties are described by timed automata and feature uncontrollable events. Then, we mention some tools capable of runtime enforcement, and we present the TiPEX tool dedicated to timed properties. Finally, we present some open challenges and avenues for future work.

6.3.2. Detecting Fault Injection Attacks with Runtime Verification

This work [12] is concerned with fault injections which are increasingly used to attack/test secure applications. In this paper, we define formal models of runtime monitors that can detect fault injections that result in test inversion attacks and arbitrary jumps in the control flow. Runtime verification monitors offer several

advantages. The code implementing a monitor is small compared to the entire application code. Monitors have a formal semantics; and we prove that they effectively detect attacks. Each monitor is a module dedicated to detecting an attack and can be deployed as needed to secure the application. A monitor can run separately from the application or it can be weaved inside the application. Our monitors have been validated by detecting simulated attacks on a program that verifies a user PIN.

6.3.3. International Competition on Runtime Verification (CRV)

In this work [15], we review the first five years of the international Competition on Runtime Verification (CRV), which began in 2014. Runtime verification focuses on verifying system executions directly and is a useful lightweight technique to complement static verification techniques. The competition has gone through a number of changes since its introduction, which we highlight in this paper.

6.3.4. A Survey of Challenges for Runtime Verification from Advanced Application Domains (beyond software)

In this work [6], we survey the future challenges for runtime verification. Typically, the two main activities in runtime verification efforts are the process of creating monitors from specifications, and the algorithms for the evaluation of traces against the generated monitors. Other activities involve the instrumentation of the system to generate the trace and the communication between the system under analysis and the monitor. Most of the applications in runtime verification have been focused on the dynamic analysis of software, even though there are many more potential applications to other computational devices and target systems. In this paper we present a collection of challenges for runtime verification extracted from concrete application domains, focusing on the difficulties that must be overcome to tackle these specific challenges. The computational models that characterize these domains require to devise new techniques beyond the current state of the art in runtime verification.

6.3.5. On the Monitoring of Decentralized Specifications Semantics, Properties, Analysis, and Simulation

In this work [3], we define two complementary approaches to monitor decentralized systems. The first relies on those with a centralized specification, i.e, when the specification is written for the behavior of the entire system. To do so, our approach introduces a data-structure that i) keeps track of the execution of an automaton, ii) has predictable parameters and size, and iii) guarantees strong eventual consistency. The second approach defines decentralized specifications wherein multiple specifications are provided for separate parts of the system. We study two properties of decentralized specifications pertaining to monitorability and compatibility between specification and architecture. We also present a general algorithm for monitoring decentralized specifications. We map three existing algorithms to our approaches and provide a framework for analyzing their behavior. Furthermore, we introduce THEMIS, a framework for designing such decentralized algorithms and simulating their behavior. We show the usage of THEMIS to compare multiple algorithms and verify the trends predicted by the analysis by studying two scenarios: a synthetic benchmark and a real example.

6.3.6. Optimal Enforcement of (timed) Properties with Uncontrollable Events

This work deals with runtime enforcement of untimed and timed properties with uncontrollable events [5]. Runtime enforcement consists in defining and using mechanisms that modify the executions of a running system to ensure their correctness with respect to a desired property. We introduce a framework that takes as input any regular (timed) property described by a deterministic automaton over an alphabet of events, with some of these events being uncontrollable. An uncontrollable event cannot be delayed nor intercepted by an enforcement mechanism. Enforcement mechanisms should satisfy important properties, namely soundness, compliance and optimality – meaning that enforcement mechanisms should output as soon as possible correct executions that are as close as possible to the input execution. We define the conditions for a property to be enforceable with uncontrollable events. Moreover, we synthesise sound, compliant and optimal descriptions of runtime enforcement mechanisms at two levels of abstraction to facilitate their design and implementation.

6.4. Teaching of Algorithms, Programming, Debugging, and Automata

Participants: Florent Bouchez Tichadou, Yliès Falcone, Théo Barollet, Antoine Clavel, Thomas Hervé, Anthony Martinez, Beryl Piasentin, Steven Sengchanh.

This domain is a new axis of the Corse team. Our goal here is to combine our expertise in compilation and teaching to help teachers and learners in computer science fields such as programming, algorithms, data strucures, automata, or more generally computing litteracy. The most important project in this regard is the automated generation and recommendation of exercises using artificial intelligence, a thesis that started this year. Other projects focus on tools to help learning through visualization (data structures, debugger, automata) or gamification (AppoLab), and are the source of many internships that give younger students experience in a research team.

6.4.1. AI4HI: Artificial Intelligence for Human Intelligence

In an ideal educative world, each learner would have access to individual pedagogical help, tailored to its needs. For instance, a tutor who could rapidly react to the questions, and propose pedagogical contents that match the learner's kills, and who could identify and work on his or her weaknesses. However, the real world imposes constraints that make this individual pedagogical help hard to achieve.

The goal of the AI4HI project is to combine the new advances in artificial intelligence with the team's skills in compilation and teaching to aid teaching through the automated generation and recommendation of exercises to learners. In particular, we target the teaching of programming and debugging to novices. This system will propose exercises that match the learners' needs and hence improve the learning, progression, and self-confidence of learners.

This projet has received an "Action Exploratoire" funding from Inria and Théo Barollet started his PhD this September so is still in its early stages.

6.4.2. AppoLab

Classical teaching of algorithms and low-level data structures is often tedious and unappealing to students. AppoLab is an online platform to engage students in their learning by including gamification in Problem-Based Learning. In its core, it is a server with scripted "exercises". Students can communicate with the server manually, but ultimately they need to script the communication also from their side, since the server will gradually impose constraints on the problems such as timeouts or large input sizes.

6.4.3. Data Structures and Program Visualization at Runtime

Debuggers are powerful tools to observe a program behaviour and find bugs but they have a hard learning curve. They provide information on low level data but are not able to analyze higher level elements such as data structures. This work tries to provide a more intuitive representation of the program execution to ease debugging and algorithms understanding. We developed a prototype, Moly, a GDB extension that explores a program runtime memory and analyze its data structures. It provides an interface with an external visualizer, Lotos, through a formatted output. Work has also started to include a tutorial on how to use GDB and these extensions.

6.4.4. Aude

Aude is a pedagogical software for manipulating, learning, and teaching finite state automata and the automata theory. It is used by the students in the second year of the bachelor in computer science at Univ. Grenoble Alpes. It allows students to get acquainted and autonomously work on the concepts involved in the theory of regular languages and automata.

DATASPHERE Team

6. New Results

6.1. Political economy

We pursued our work on digital platforms and their impact on the structure of socio-economic systems, which results from the capacity to separate data or information from the actors of the physical world. In [5], we showed how the movement above ground of the intermediation activity transforms territories.

6.2. Anthropocene studies

We have investigated the possible similarities between biological systems and social systems facing shortage of resources, suggesting that the digital revolution might have something to do with the Anthropocene [4]. More comprehensive approaches that rely on digital systems to control society and nudge citizens to adapt their behavior have been developed in Asia. We analyse in particular the specificity of Asia in these transformations [6].

6.3. Network data analytics

In collaboration with the Chinese Academy of Sciences, we worked on packet processing algorithmic for high speed network measurements. In [1] a packet capture archive system is developed and described. a theoretical analysis of the TCAM updates delay that is the main shortcoming of TCAM usage in high speed packet processors is presented. Quality of service for network functions were considered in [3].

6.4. Geopolitics of BGP

We have investigated the logical layer of cyberspace through an analysis of the structure of connectivity and the Border Gateway Protocol (BGP). This protocol has been leveraged by countries to control the flow of information or for active strategic purposes. We focused on several countries and characterized their strategies by linking them to current architectures and understanding their resilience in times of crisis. We focus on the case of Iran and uncovered the deep changes that has happened in the past 5 years. This study was premonitory as we observed in november 2019 the full extend of the strategy with the large scale internet disruptions. This generates a lot of mediatic coverage.

PRIVATICS Project-Team

6. New Results

6.1. Differential Inference Testing

Participant: Claude Castelluccia.

In order to protect individuals' privacy, data have to be "well-sanitized" before sharing them, i.e. one has to remove any personal information before sharing data. However, it is not always clear when data shall be deemed well-sanitized. In [10], we argue that the evaluation of sanitized data should be based on whether the data allows the inference of sensitive information that is specific to an individual, instead of being centered around the concept of re-identification. We propose a framework to evaluate the effectiveness of different sanitization techniques on a given dataset by measuring how much an individual's record from the sanitized dataset influences the inference of his/her own sensitive attribute. Our intent is not to accurately predict any sensitive attribute but rather to measure the impact of a single record on the inference of sensitive information. We demonstrate our approach by sanitizing two real datasets in different privacy models and evaluate/compare each sanitized dataset in our framework.

6.2. Analyse des impacts de la reconnaissance faciale - Quelques éléments de méthode (in French)

Participants: Claude Castelluccia, Daniel Le Métayer.

Significant technical progress has been made in recent years in the field of image processing, in particular in facial recognition. The deployments and experiments of this type of systems are more and more numerous. However, opinions differ on their use, especially in public space. Noting the lack of consensus on a technology that can have a significant impact on society, many organizations have alerted public opinion and asked for a public debate on the subject. We believe that such a debate is indeed necessary. However, for it to be truly productive, it is necessary to be able to confront the arguments in a rigorous manner while avoiding, as far as possible, the preconceptions, and by distinguishing established facts from assumptions or opinions. The purpose of this document [14] is precisely to help put the terms of the debate on solid foundations. It is therefore not a question here of taking a position on facial recognition in general nor of providing an exhaustive review of its applications but of proposing elements of method, illustrated by a few examples. We first present a quick overview of the applications of facial recognition before detailing the reasons that make it a particularly sensitive subject, emphasizing in particular the risks linked to a possible generalization of its use. We then present an incremental, comparative and rigorous approach to analyze the impacts of a facial recognition system.

6.3. Towards a generic framework for black-box explanation methods

Participants: Daniel Le Métayer, Clément Hénin.

Explainability has generated increased interest during the last decade because the most accurate ML techniques often lead to opaque Algorithmic Decision Systems (ADS) and opacity is a major source of mistrust. Indeed, even if explanations are not a panacea, they can play a key role, not only to enhance trust in the system, but also to allow its users to better understand its outputs and therefore to make a better use of it. In addition, they are necessary to make it possible to challenge the decisions resulting from an ADS. Explanations can take different forms, they can target different types of users and different types of methods can be used to produce them. Our work on this topic [15] focuses on a category of methods, called "black-box", that do not make any assumption about the availability of the code of the ADS or its implementation techniques. Our first contribution is to bring to light a common structure for Black-box Explanation Methods and to define a generic framework allowing us to compare and classify different approaches. This framework consists of three components, called respectively

Sampling, Generation and Interaction. Beyond its interest as a systematic presentation of the state of the art, we believe that this framework can also provide new insights for the design of new explanation systems. For example, it may suggest new combinations of Sampling and Generation components or criteria to choose the most appropriate combination to produce a given type of explanation.

6.4. A generic information and consent framework for the IoT

Participants: Daniel Le Métayer, Mathieu Cunche, Victor Morel.

The development of the Internet of Things (IoT) raises specific privacy issues especially with respect to information and consent. People are generally unaware of the devices collecting data about them and do not know the organizations operating them. Solutions such as stickers or wall signs are not effective information means in most situations. As far as consent is concerned, individuals do not have simple means to express and communicate it to the entities collecting data. Furthermore, the devices used to collect data in IoT environments have scarce resources; some of them do not have any user interface, are battery-operated or operate passively. The Working Party 29 (now "European Data Protection Board") advocates the design of new consent mechanisms, such as "privacy proxies", on the devices themselves. Starting from their recommendations, we have defined general requirements that have to be met to ensure that information and consent are managed in a manner that is satisfactory both for data subjects and for data controllers. We have shown in [8] how these requirements can be implemented in different situations, in particular through declaration registers and beacons. Depending on the context and the types of devices involved, not all technical options are always possible. In order to provide guidance to IoT system designers, we have outlined the main choice factors in the design pace are illustrated the framework with several challenging case studies. We have also implemented a Proof of Concept prototype implementation of these techniques.

6.5. Analysis of privacy policies to enhance informed consent

Participant: Daniel Le Métayer.

A privacy policy language must meet a number of requirements to be able to express the valid consent of the data subject for the processing of their personal data. For example, under the GDPR, valid consent must be freely given, specific, informed and unambiguous. Therefore, the language must be endowed with a formal semantics in order to avoid any ambiguity about the meaning of a privacy policy. However, the mere existence of a semantics does not imply that DSs properly understand the meaning of a policy and its potential consequences. One way to enhance the understanding of the data subjects is to provide them information about the potential risks related to a privacy policy. This is in line with Recital 39 of the GDPR which stipulates that data subjects should be "made aware of the risks, rules, safeguards and rights in relation to the processing of personal data and how to exercise their rights in relation to such processing". To address this need, we have defined a language in [11], called PILOT, meeting these requirements and shown its benefits to define precise privacy policies and to highlight the associated privacy risks. In order to automatically answer questions related to privacy risks, we use the verification tool SPIN and the modeling language PROMELA. Risk properties are encoded in Linear Temporal Logic properties that can be automatically checked by SPIN.

6.6. Understanding algorithmic decision-making: Opportunities and challenges, Study for the European Parliament (STOA)

Participants: Claude Castelluccia, Daniel Le Métayer.

Algorithms are far from being a recent invention but they are increasingly involved in systems used to support decision making. Algorithmic Decision Systems (ADS) often rely on the analysis of large amounts of personal data to infer correlations or, more generally, to derive information deemed useful to make decisions. Humans may have a role of varying degree in the decision making and may even be completely out of the loop in entirely automated systems. In many situations, the impact of the decision on people can be significant: access to credit, employment, medical treatment, judicial sentences, etc. Entrusting ADS to make or to influence such decisions raises a variety of issues that differ in nature such as ethical, political, legal, technical, etc. and great care must be taken to analyse and address these issues. If they are neglected, the expected benefits of these systems may be offset by the variety of risks for individuals (discrimination, unfair practices, loss of autonomy, etc.), the economy (unfair practices, limited access to markets, etc.) and society as a whole (manipulation, threat to democracy, etc.).

We have written a report for the European Parliament reviewing the opportunities and risks related to the use of ADS. We present existing options to reduce these risks and explain their limitations. We sketch some recommendations to benefit from the tremendous possibilities of ADS while limiting the risks related to their use. Beyond providing an up-to-date and systematic review of the situation, the report gives a precise definition of a number of key terms and an analysis of their differences. This helps clarify the debate. The main focus of the report is the technical aspects of ADS. However, other legal, ethical and social dimensions are considered to broaden the discussion.

6.7. Saving Private Addresses: An Analysis of Privacy Issues in the Bluetooth-Low-Energy Advertising Mechanism

Participants: Mathieu Cunche, Guillaume Celiosa.

The Bluetooth Low Energy (BLE) protocol is being included in a growing number of connected objects such as fitness trackers and headphones. As part of the service discovery mechanism of BLE, devices announce themselves by broadcasting radio signals called advertisement packets that can be collected with off-the-shelf hardware and software. To avoid the risk of tracking based on those messages, BLE features an address randomization mechanism that substitutes the device address with random temporary pseudonyms,called Private addresses. We analyze the privacy issues associated with the advertising mechanism of BLE, leveraging a large dataset of advertisement packets collected in the wild. First, we identified in [7] that some implementations fail at following the BLE specifications on the maximum lifetime and the uniform distribution of random identifiers. Furthermore, we found that the payload of the advertisement packet can hamper the randomization mechanism by exposing counters and static identifiers. In particular, we discovered that advertising data of Apple and Microsoft proximity protocols can be used to defeat the address randomization scheme. Finally, we discuss how some elements of advertising data can be leveraged to identify the type of device, exposing the owner to inventory attacks

6.8. Fingerprinting Bluetooth-Low-Energy Devices Based on the Generic Attribute Profile

Participants: Mathieu Cunche, Guillaume Celiosa.

Bluetooth Low Energy (BLE) is a short range wireless technology included in many consumer devices such as smartphones, earphones and wristbands. As part of the Attribute (ATT) protocol, discoverable BLE devices expose a data structure called Generic Attribute (GATT) profile that describes supported features using concepts of services and characteristics. This profile can be accessed by any device in range and can expose users to privacy issues. We study how the GATT profile can be used to create a fingerprint that can be exploited to circumvent anti-tracking features of the BLE standard (i.e. MAC address randomization). Leveraging a dataset of more than 13000 profiles, we analyze the potential of this fingerprint and show that it can be used to uniquely identify a number of devices. We also shed light in [6] on several issues where GATT profiles can be mined to infer sensitive information that can impact privacy of users. Finally, we suggest solutions to mitigate those issues.

6.9. Privacy implications of switching ON a light bulb in the IoT world

Participants: Vincent Roca, Mathieu Thiery.

The number of connected devices is increasing every day, creating smart homes and shaping the era of the Internet of Things (IoT), and most of the time, end-users are unaware of their impacts on privacy. We analyze in [23] the ecosystem around a Philips Hue smart white bulb in order to assess the privacy risks associated to the use of different devices (smart speaker or button) and smartphone applications to control it. We show that using different techniques to switch ON or OFF this bulb has significant consequences regarding the actors involved (who mechanically gather information on the user's home) and the volume of data sent to the Internet (we measured differences up to a factor 100, depending on the control technique we used). Even when the user is at home, these data flows often leave the user's country, creating a situation that is neither privacy friendly (and the user is most of the time ignorant of the situation), nor sovereign (the user depends on foreign actors), nor sustainable (the extra energetic consumption is far from negligible). We therefore advocate a complete change of approach, that favors local communications whenever sufficient.

6.10. Security Analysis of Subject Access Request Procedures How to authenticate data subjects safely when they request for their data

Participants: Cédric Lauradoux, Coline Boniface.

With the GDPR in force in the EU since May 2018, companies and administrations need to be vigilant about the personal data they process. The new regulation defines rights for data subjects and obligations for data controllers but it is unclear how subjects and controllers interact concretely. In [4], we try to answer two critical questions: is it safe for a data subject to exercise the right of access of her own data? When does a data controller have enough information to authenticate a data subject? To answer these questions, we have analyzed recommendations of Data Protection Authorities and authentication practices implemented in popular websites and third-party tracking services. We observed that some data controllers use unsafe or doubtful procedures to authenticate data subjects. The most common flaw is the use of authentication based on a copy of the subject's national identity card transmitted over an insecure channel. We define how a data controller should react to a subject's request to determine the appropriate procedures to identify the subject and her data. We provide compliance guidelines on data access response procedures.

6.11. Plausible Deniability for Practical Privacy-Preserving Live Streaming

Participant: Antoine Boutet.

Video consumption is one of the most popular Internet activities worldwide. The emergence of sharing videos directly recorded with smartphones raises important privacy concerns. In this work we propose P3LS, the first practical privacy-preserving peer-to-peer live streaming system. To protect the privacy of its users, P3LS relies on k-anonymity when users subscribe to streams, and on plausible deniability for the dissemination of video streams. Specifically, plausible deniability during the dissemination phase ensures that an adversary is never able to distinguish a user's stream of interest from the fake streams from a statistical analysis (i.e., using an analysis of variance). We exhaustively evaluate P3LS and show that adversaries are not able to identify the real stream of a user with very high confidence. Moreover, P3LS consumes 30% less bandwidth than the standard k-anonymity approach where nodes fully contribute to the dissemination of k streams.

6.12. Protecting motion sensor data against sensitive inferences through an adversarial network approach

Participants: Antoine Boutet, Théo Jourdan.

With the widespread development of the quantified self movement, more and more motion sensor data are captured and transmitted through the intermediary of smartphones. However, granting to applications a direct access to sensor data expose users to many privacy risks, including in particular the possibility of inferring their activities and transportation mode to more sensitive inferences such as their demographic attributes or even mobility deficiency. In this work, we propose a privacy-preserving scheme to protect sensor data for activity recognition while at the same time preventing unwanted sensitive inferences on specific information. To achieve this objective, we leverage on the powerful framework of generative adversarial networks (GANs) to sanitize the sensor data. More precisely in our framework three neural networks are jointly trained, a generator that aim at sanitizing the data given at input as well two discriminators that try to infer respectively the sensitive attributes and the current activity of the user. By letting these neural networks compete against each other, the mechanism improves the protection while providing a good accuracy in terms of activity recognition and limiting sensitive inferences on specified attributes. Preliminary results demonstrate that the approach is promising in terms of achieving a good utility-privacy trade-off.

6.13. Inria white book on Cybersecurity: Current challenges and Inria's research directions

Participant: Vincent Roca.

This book provides an overview of research areas in cybersecurity, illustrated by contributions from Inria teams. The first step in cybersecurity is to identify threats and define a corresponding attacker model. Threats, including malware, physical damage or social engineering, can target the hardware, the network, the operating system, the applications, or the users themselves.

Then, detection and protection mechanisms must be designed to defend against these threats. One of the core mechanisms is cryptography, in order to ensure the confidentiality and integrity of data. These primitives must be the object of continuous cryptanalysis to ensure the highest level of security. However, secure cryptographic primitives alone are not sufficient for secure communications and services: cryptographic protocols, implementing richer interactions on top of the primitives, are needed. These protocols are distributed systems. Ensuring that they achieve their goals in the presence of an adversary requires the use of formal verification techniques, which have been extremely successful in this field.

Additional security services, such as authentication and access control, are needed to enforce a security policy. These security services, usually provided by the operating system or the network devices, can themselves be attacked and sometimes bypassed. Therefore, activities on the information system are monitored in order to detect any violation of the security policy. Finally, as attacks can spread extremely fast, the system must react automatically or at least reconfigure itself to avoid propagating attacks.

Privacy has also become an intrinsic part of cybersecurity. Privacy has its own properties, techniques, and methodology. Moreover, the study of privacy often requires to take legal, economical, and sociological aspects into account.

All these security mechanisms need to be carefully integrated in security-critical applications. These applications include traditional safety-critical applications that are becoming increasingly connected and therefore more vulnerable to security attacks, as well as new infrastructures running in the cloud or connected to a multitude of Things (IoT).

6.14. Inspect what your location history reveals about you - Raising user awareness on privacy threats associated with disclosing his location data

Participant: Antoine Boutet.

Location is one of the most extensively collected personal data on mobile by applications and third-party services. However, how the location of users is actually processed in practice by the actors of targeted advertising ecosystem remains unclear. Nonetheless, these providers have a strong incentive to create very detailed profile of users to better monetize the collected data. End users are usually not aware about the strength and wide range of inference that can be performed from their mobility traces. In this work, users interact with a web-based application to inspect their location history and to discover the inferential power of this kind of data. Moreover to better understand the possible countermeasures, users can apply a sanitization to protect their data and visualize the impact on both the mobility traces and the associated inferred information. The objective of this work is to raise the user awareness on the profiling capabilities and the privacy threats associated with disclosing his location data as well as how sanitization mechanisms can be efficient to mitigate these privacy risks. In addition, by collecting users feedbacks on the personal information revealed and the usage of a geosanitization mechanism, we hope that this work will also be useful to constitute a new and valuable dataset on users perceptions on these questions.

6.15. Pseudonymisation techniques and best practices

Participant: Cédric Lauradoux.

This ENISA report explores further the basic notions of pseudonymisation, as well as technical solutions that can support implementation in practice. Starting from a number of pseudonymisation scenarios, the report defines first the main actors that can be involved in the process of pseudonymisation along with their possible roles. It then analyses the different adversarial models and attacking techniques against pseudonymisation, such as brute force attack, dictionary search and guesswork. Moreover, it presents the main pseudonymisation techniques and policies available today.

SPADES Project-Team

6. New Results

6.1. Design and Programming Models

Participants: Pascal Fradet, Alain Girault, Gregor Goessler, Xavier Nicollin, Arash Shafiei, Jean-Bernard Stefani, Martin Vassor, Souha Ben Rayana.

6.1.1. Hypercells

The location graph framework we have introduced in [66] has evolved into the Hypercell framework presented in [18]. The Hypercell framework allows the definition of different component models for dynamic software architectures featuring both sharing and encapsulation. The basic behavioral theory of hypercells in the form of a contextual bisimulation has been developed and we are currently developing proofs of correctness for encapsulation policies based on this theory.

In collaboration with the Spirals team at Inria Lille – Nord Europe, and Orange, we have used hypercells as a pivot model for developing interpretations, formally defined with the Alloy specification language, of various languages and formalisms for the description of software configurations for cloud computing environments. Configuration languages considered include the TOSCA and OCCI standards, as well as the Open Stack Heat Orchestration Template (HOT), Docker Compose, and the Aeolus component model for cloud deployment. This work, developed as part of a bilateral contract with Orange, allowed the development of a verification tool for the correctness of HOT configurations, and helped uncover several flaws in the ETSI NFV standard.

6.1.2. Dynamicity in dataflow models

Recent dataflow programming environments support applications whose behavior is characterized by dynamic variations in resource requirements. The high expressive power of the underlying models (*e.g.*, Kahn Process Networks or the CAL actor language) makes it challenging to ensure predictable behavior. In particular, checking *liveness* (*i.e.*, no part of the system will deadlock) and *boundedness* (*i.e.*, the system can be executed in finite memory) is known to be hard or even undecidable for such models. This situation is troublesome for the design of high-quality embedded systems. In the past few years, we have proposed several parametric dataflow models of computation (MoCs) [49], [39], we have written a survey providing a comprehensive description of the existing parametric dataflow MoCs [42], and we have studied *symbolic* analyses of dataflow graphs [43]. More recently, we have proposed an original method to deal with lossy communication channels in dataflow graphs [48].

We are nowadays studying models allowing *dynamic reconfigurations* of the *topology* of the dataflow graphs. This is required by many modern streaming applications that have a strong need for reconfigurability, for instance to accommodate changes in the input data, the control objectives, or the environment.

We have proposed a new MoC called Reconfigurable Dataflow (RDF) [13]. RDF extends SDF with transformation rules that specify how the topology and actors of the graph may be reconfigured. Starting from an initial RDF graph and a set of *transformation rules*, an arbitrary number of new RDF graphs can be generated at runtime. Transformations can be seen as graph rewriting rules that match some sub-part of the dataflow graph and replace it by another one. Transformations can be applied an arbitrary number of times during execution and therefore can produce an arbitrary number of new graphs. The major feature and advantage of RDF is that it can be statically analyzed to guarantee that all possible graphs generated at runtime will be connected, consistent, and live. To the best of our knowledge, RDF is the only dataflow MoC allowing an arbitrary number of topological reconfigurations while remaining statically analyzable. It remains to complete the RDF implementation and to evaluate it on realistic case studies. Preliminary results indicate that dynamic reconfigurations can be implemented efficiently.

This is the research topic of Arash Shafiei's PhD, in collaboration with Orange Labs.

6.2. Certified Real-Time Programming

Participants: Pascal Fradet, Alain Girault, Gregor Goessler, Xavier Nicollin, Sophie Quinton, Xiaojie Guo, Maxime Lesourd.

6.2.1. Time predictable programming languages and architectures

Time predictability (PRET) is a topic that emerged in 2007 as a solution to the ever increasing unpredictability of today's embedded processors, which results from features such as multi-level caches or deep pipelines [46]. For many real-time systems, it is mandatory to compute a strict bound on the program's execution time. Yet, in general, computing a tight bound is extremely difficult [69]. The rationale of PRET is to simplify both the programming language and the execution platform to allow more precise execution times to be easily computed [35].

Within the CAPHCA project, we have proposed a new approach for predictable inter-core communication between tasks allocated on different cores. Our approach is based on the execution of synchronous programs written in the FOREC parallel programming language on PREcision Timed (hence deterministic) architectures [71], [72]. The originality resides in the time-triggered model of computation and communication that allows for a very precise control over the thread execution. Synchronization is done via configurable Time Division Multiple Access (TDMA) arbitrations (either physical or conceptual) where the optimal size and offset of the time slots are computed to reduce the inter-core synchronization costs. Results show that our model guarantees time-predictable inter-core communication, the absence of concurrent accesses (without relying on hardware mechanisms), and allows for optimized execution throughput [17]. This is a collaboration with Nicolas Hili and Eric Jenn, the postdoc of Nicolas Hili being funded by the CAPHCA project.

We have also proposed a *multi-rate* extension of FOREC [16]. Indeed, up to now FOREC programs were constrained to operate at a single rate, meaning that all the parallel threads had to share the same execution rate. While this simplified the semantics, it also represented a significant limitation.

Finally, we have extended the compiler of the PRET-C programming language [33], [34] in order to make it energy aware. PRET-C is a parallel programming language in the same sense as Esterel [44], meaning that the parallelism is "compiled away": the PRET-C compiler generates sequential code where the parallel threads from the source program are interleaved according to the synchronous semantics, and produces a classical Control Flow Graph (CFG). This CFG is then turned into a Timed Control Flow Graph (TCFG) by labeling each basic block with the number of clock cycles required to execute it on the chosen processor, based on its micro-architectural characteristics. From the TCFG, we use the method described in Section 6.2.5 to compute a Pareto front of non-dominated (worst-case execution time – WCET, worst-case energy consumption – WCEC) compromises.

6.2.2. Synthesis of switching controllers using approximately bisimilar multiscale abstractions

The use of discrete abstractions for continuous dynamics has become standard in hybrid systems design (see *e.g.*, [67] and the references therein). The main advantage of this approach is that it offers the possibility to leverage controller synthesis techniques developed in the areas of supervisory control of discrete-event systems [64]. The first attempts to compute discrete abstractions for hybrid systems were based on traditional systems behavioral relationships such as simulation or bisimulation, initially proposed for discrete systems most notably in the area of formal methods. These notions require inclusion or equivalence of observed behaviors which is often too restrictive when dealing with systems observed over metric spaces. For such systems, a more natural abstraction requirement is to ask for closeness of observed behaviors. This leads to the notions of approximate simulation and bisimulation introduced in [50]. These approaches are based on sampling of time and space where the sampling parameters must satisfy some relation in order to obtain abstractions of a prescribed precision. In particular, the smaller the time sampling parameter, the finer the lattice used for approximating the state-space; this may result in abstractions with a very large number of states when the sampling period is small. However, there are a number of applications where sampling has to be fast; though this is generally necessary only on a small part of the state-space.

In previous work we have proposed an approach using mode sequences as symbolic states for our abstractions [59]. By using mode sequences of variable length we are able to adapt the granularity of our abstraction to the dynamics of the system, so as to automatically trade off precision against controllability of the abstract states [12]. We have shown the effectiveness of the approach on examples inspired by road traffic regulation.

6.2.3. A Markov Decision Process approach for energy minimization policies

In the context of independent real-time sporadic jobs running on a single-core processor equipped with Dynamic Voltage and Frequency Scaling (DVFS), we have proposed a Markov Decision Process approach (MDP) to minimize the energy consumption while guaranteeing that each job meets its deadline. The idea is to leverage on the *statistical information* on the jobs' characteristics available at design time: release time, worst-case execution time (WCET), and relative deadline. This is the topic of Stephan Plassart's PhD, funded by the CASERM Persyval project. We have considered several cases depending on the amount of information available at design time:

Offline case: In the offline case, all the information is known and we have proposed the first linear complexity offline scheduling algorithm that minimizes the total energy consumption [15]: our complexity is $\mathcal{O}(n)$ where n is the number of jobs to be scheduled, while the previously best known algorithms were in $\mathcal{O}(n^2)$ and $\mathcal{O}(n \log n)$ [60].

Clairvoyant case: In the clairvoyant case, the characteristics of the jobs are only known statistically, and each job's WCET and relative deadline are only known at release time. We want to compute the *optimal* online scheduling speed policy that minimizes the *expected* energy consumption while guaranteeing that each job meets its deadline. This general constrained optimization problem can be modeled as an unconstrained MDP by choosing a proper state space that also encodes the constraints of the problem. In the finite horizon case we use a dynamic programming algorithm, while in the infinite horizon case we use a value iteration algorithm [25].

Non-clairvoyant case: In the non-clairvoyant case, the actual execution time (AET) of a job is only known only when this job completes its execution. This AET is of course assumed to be less than the WCET, which is known at the job's release time. Again, by building an MDP for the system with a well chosen state, we compute the *optimal* online scheduling speed policy that minimizes the *expected* energy consumption [26].

Learning case: In the learning case, the only information known for the jobs are a bound on the jobs' WCETs and a bound on their deadlines. We have proposed two *reinforcement learning* algorithms, one that learns the optimal value of the expected energy (Q-learning), and another one that learns the probability transition matrix of the system, from which we derive the optimal online speed policy.

This work led us to compare several existing speed policies with respect to their feasibility. Indeed, the policies (OA) [70], (AVR) [70], and (BKP) [37] all assume that the maximal speed S_{max} available on the processor is infinite, which is an unrealistic assumption. For these three policies and for our (MDP) policy, we have established necessary and sufficient conditions on S_{max} guaranteeing that no job will ever miss its deadline [27].

6.2.4. Formal proofs for schedulability analysis of real-time systems

We contribute to Prosa [31], a Coq library of reusable concepts and proofs for real-time systems analysis. A key scientific challenge is to achieve a modular structure of proofs, *e.g.*, for response time analysis. Our goal is to use this library for:

1. a better understanding of the role played by some assumptions in existing proofs;
2. a formal verification and comparison of different analysis techniques; and
3. the certification of results of existing (*e.g.*, industrial) analysis tools.

We have further developed CertiCAN, a tool produced using Coq for the formal certification of CAN analysis results [14]. Result certification is a process that is light-weight and flexible compared to tool certification, which makes it a practical choice for industrial purposes. The analysis underlying CertiCAN is based on a combined use of two well-known CAN analysis techniques [68]. Additional optimizations have been implemented (and proved correct) to make CertiCAN computationally efficient. Experiments demonstrate that CertiCAN is able to certify the results of RTaW-Pegase, an industrial CAN analysis tool, even for large systems.

In addition, we have started investigating how to connect Prosa with implementations and less abstract models. Specifically, we have used Prosa to provide a schedulability analysis proof for RT-CertiKOS, a single-core sequential real-time OS kernel verified in Coq [20]. A connection with a timed-automata based formalization of the CAN specification is also in progress. Our objective with this line of research is to understand and bridge the gap between the abstract models used for real-time systems analysis and actual real-time systems implementation.

Finally, we contributed to a major refactoring of the Prosa library to make it more easily extendable and usable.

6.2.5. Scheduling under multiple constraints and Pareto optimization

We have completed a major work on embedded systems subject to multiple non-functional constraints, by proposing the first of its kind multi-criteria scheduling heuristics for a DAG of tasks onto an homogeneous multi-core chip [9], [23]. Given an application modeled as a Directed Acyclic Graph (DAG) of tasks and a multicore architecture, we produce a set of non-dominated (in the Pareto sense) static schedules of this DAG onto this multicore. The criteria we address are the execution time, reliability, power consumption, and peak temperature. These criteria exhibit complex antagonistic relations, which make the problem challenging. For instance, improving the reliability requires adding some redundancy in the schedule, which penalizes the execution time. To produce Pareto fronts in this 4-dimension space, we transform three of the four criteria into constraints (the reliability, the power consumption, and the peak temperature), and we minimize the fourth one (the execution time of the schedule) under these three constraints. By varying the thresholds used for the three constraints, we are able to produce a Pareto front of non-dominated solutions. Each Pareto optimum is a static schedule of the DAG onto the multicore. We propose two algorithms to compute static schedules. The first is a ready list scheduling heuristic called ERPOT (Execution time, Reliability, POwer consumption and Temperature). ERPOT actively replicates the tasks to increase the reliability, uses Dynamic Voltage and Frequency Scaling to decrease the power consumption, and inserts cooling times to control the peak temperature. The second algorithm uses an Integer Linear Programming (ILP) program to compute an optimal schedule. However, because our multi-criteria scheduling problem is NP-complete, the ILP algorithm is limited to very small problem instances, namely DAGs of at most 8 tasks. Comparisons showed that the schedules produced by ERPOT are on average only 9% worse than the optimal schedules computed by the ILP program, and that ERPOT outperforms the PowerPerf-PET heuristic from the literature on average by 33%. This is a joint work with Athena Abdi and Hamid Zarandi from Amirkabir University in Tehran, Iran.

In a related line of work, we have considered the bi-criteria minimization problem in the (worst-case execution time – WCET, worst-case energy consumption – WCEC) space for real-time programs. To the best of our knowledge, this is the first contribution of this kind in the literature.

A real-time program is abstracted as a Timed Control Flow Graph (TCFG), where each basic block is labeled with the number of clock cycles required to execute it on the chosen processor at the nominal frequency. This timing information can be obtained, for instance, with WCET analysis tools. The target processor is equipped with dynamic voltage and frequency scaling (DVFS) and offers several (frequency f, voltage V) operating points. The goal is to compute a set of non-dominated points in the (WCET, WCEC) plane, non-dominated in the Pareto sense. Each such point is an assignment from the set of basic blocks of the TCFG to the set of available (f, V) pairs.

From the TCFG we extract the longest execution path, therefore deriving the WCET and the WCEC for a chosen fixed (f, V) pair. By construction, all the other execution paths are shorter, so this WCET and this WCEC hold for the whole program. This ensures that each single-frequency assignment is a non-dominated

point. Then, we study two frequencies assignments, still for the longest execution path. When the frequency switching costs in time and in energy are assumed to be negligible, we demonstrate that each two frequencies (say with f_i and f_j) assignment is a point in the segment between the single frequency assignment at f_i and the single frequency assignment at f_j. We also propose a linear time heuristic to assign a (f, V) pair to all the other blocks (*i.e.*, those not belonging to the longest path) such that all the other execution paths have a shorter WCET and a lesser WCEC. A key result is that we demonstrate that any two frequencies assignment where the two frequencies are not contiguous is dominated either by a single frequency assignment or by a two frequencies assignment with contiguous frequencies. A corollary is that the Pareto front is a continuous piece-wise affine function. Finally, we generalize these results to the case where the frequency switching costs are not negligible. This is the topic of Jia Jie Wang's postdoc.

We evaluate our method and heuristic on a set of hard real time benchmark programs and we show that they perform extremely well. Our DVFS assignment algorithm can also be used as a back-end for the compiler of the PRET-C programming language [33], [34] in order to make it energy aware, thanks to the ability of this compiler to generate TCFGs (see Section 6.2.1).

6.3. Fault Management and Causal Analysis

Participants: Gregor Goessler, Jean-Bernard Stefani, Sihem Cherrared, Thomas Mari, Martin Vassor.

6.3.1. Fault Ascription in Concurrent Systems

Fault ascription is a precise form of fault diagnosis that relies on counterfactual analysis for pinpointing the causes of system failures. Research on counterfactual causality has been marked, until today, by a succession of definitions of causation that are informally validated against human intuition on mostly simple examples. This approach suffers from its dependence on the tiny number and incompleteness of examples in the literature, and from the lack of objective correctness criteria [52].

We have defined in [28] a set of expected properties for counterfactual analysis, and presented a refined analysis that conforms to our requirements. As an early study of the behavior of our analysis under abstraction we have established its monotony under refinement.

6.3.2. Causal Explanations in Discrete Event Systems

Model-Based Diagnosis of discrete event systems (DES) usually aims at detecting failures and isolating faulty event occurrences based on a behavioural model of the system and an observable execution log. The strength of a diagnostic process is to determine *what* happened that is consistent with the observations. In order to go a step further and explain *why* the observed outcome occurred, we borrow techniques from causal analysis.

In [21] we have presented two constructions of explanations that are able to extract the relevant part of a property violation that can be understood by a human operator. Both support partial observability of events. The first construction is based on minimal sub-sequences of the traces of the log that entail a violation of the property. The second approach is based on a construction of layers similar to [56], in which the explanation is constructed from the choices that definitely move the system closer to the violation of the property. Both approaches are complementary: while subsequence-based explanations are well suited to "condense" the execution trace in sequential portions of the model but are prone to keep non-pertinent parts such as initialisation sequences in the explanation, effective choice explanations highlight the "fateful" choices in an execution, as well as alternative events that would have helped avoid the outcome. Effective choice explanations are therefore able to explain failures stemming from non-deterministic choices, such as concurrency bugs.

6.3.3. Fault Management in Virtualized Networks

From a more applied point of view we have been investigating, in the context of Sihem Cherrared's PhD thesis, approaches for fault explanation and localization in virtualized networks. In essence, Network Function Virtualization (NFV), widely adopted by the industry and the standardization bodies, is about running network functions as software workloads on commodity hardware to optimize deployment costs and simplify the life-cycle management of network functions. However, it introduces new fault management challenges including dynamic topology and multi-tenant fault isolation.

In [29] we have proposed a model-based root cause analysis framework called SAKURA. In order to overcome the lack of accurate previous knowledge, SAKURA features a self-modeling algorithm that models the dependencies within and between layers of virtual networks, including auto-recovery and elasticity aspects. Model-based diagnosis is performed using constraint solving on the previous and acquired knowledge. As an illustration we have applied SAKURA to the virtual IpMultimedia Subsystem (vIMS).

Finally, in our survey on fault management in network virtualization environments [11] we have addressed the impact of virtualization on fault management, proposed a new classification of the recent fault management research achievements in network virtualization environments, and compared their major contributions and shortcomings.

ELAN Project-Team

7. New Results

7.1. Static simulation of thin elastic ribbons

Participants: Raphaël Charrondière, Florence Bertails-Descoubes, Victor Romero.

In collaboration with Sébastien Neukirch (Sorbonne Université, Institut Jean le Rond d'Alembert), we have proposed a robust and efficient numerical model to compute stable equilibrium configurations of thin elastic ribbons featuring arbitrarily curved natural shapes. Our spatial discretization scheme relies on elements characterized by a linear normal curvature and a quadratic geodesic torsion with respect to arc length. Such a high-order discretization allows for a great diversity of kinematic representations, while guaranteeing the ribbon to remain perfectly inextensible. Stable equilibria are calculated by minimizing gravitational and elastic energies of the ribbon, under a developability constraint. This work is currently under review in a journal of Mechanics. Some preliminary results have already been communicated about in two French congresses, one in Mechanics [6] and one in Computer Graphics [7] (best paper award).

7.2. Video-based measurement of the friction coefficient between cloth and a substrate

Participants: Haroon Rasheed, Victor Romero, Florence Bertails-Descoubes.

In collaboration with Arnaud Lazarus (Sorbonne Université, Institut Jean le Rond d'Alembert), Jean-Sébastien Franco and Stefanie Wuhrer (Inria, Morphéo team), we have investigated a first non-invasive measurement network for estimating cloth friction at contact with a substrate. Our network was trained on data exclusively generated by the solver ARGUS co-developed by the ELAN team, which we have carefully validated against real experiments under controlled conditions. We have shown promising friction measurement results on multiple real cloth samples contacting various kinds of substrates, by comparing our estimations based on a simple video acquisition protocol against standard measurements. This work has been submitted in late 2019 for publication in a Computer Vision conference, and some preliminary results have been communicated about in a mechanical congress [4].

7.3. Willmore flow simulation with diffusion-redistanciation numerical schemes

Participant: Thibaut Metivet.

In collaboration with Arnaud Sengers (Université Claude Bernard), Emmanuel Maitre (Laboratoire Jean Kuntzmann, Grenoble INP) and Mourad Ismail (Laboratoire Interdisciplinaire de Physique, UGA), we have proposed original diffusion-redistanciation numerical schemes to compute the static shapes of elastic membranes with bending stiffness under constant area constraints. This numerical method relies on an implicit representation of the surface which is used as an initial condition for diffusion-like equations. This allows to circumvent the usual difficulties pertaining to the high geometrical order and non-linearities of the bending energy and to benefit from the robustness of discretised diffusion operators. We have implemented the schemes within the finite element library Feel++ and studied the numerical convergence properties in 2D and 3D. We have also validated our method using comparative benchmarks computed with standard approaches. This work has led to the PhD defense of Arnaud Sengers [35] and a publication is under preparation.

MISTIS Project-Team

7. New Results

7.1. Mixture models

7.1.1. Mini-batch learning of exponential family finite mixture models

Participant: Florence Forbes.

Joint work with: Hien Nguyen, La Trobe University Melbourne Australia and Geoffrey J. McLachlan, University of Queensland, Brisbane, Australia.

Mini-batch algorithms have become increasingly popular due to the requirement for solving optimization problems, based on large-scale data sets. Using an existing online expectation-maximization (EM) algorithm framework, we demonstrate [28] how mini-batch (MB) algorithms may be constructed, and propose a scheme for the stochastic stabilization of the constructed mini-batch algorithms. Theoretical results regarding the convergence of the mini-batch EM algorithms are presented. We then demonstrate how the mini-batch framework may be applied to conduct maximum likelihood (ML) estimation of mixtures of exponential family distributions, with emphasis on ML estimation for mixtures of normal distributions. Via a simulation study, we demonstrate that the mini-batch algorithm for mixtures of normal distributions can outperform the standard EM algorithm. Further evidence of the performance of the mini-batch framework is provided via an application to the famous MNIST data set.

7.1.2. Component elimination strategies to fit mixtures of multiple scale distributions

Participants: Florence Forbes, Alexis Arnaud.

We address the issue of selecting automatically the number of components in mixture models with non-Gaussian components. As amore efficient alternative to the traditional comparison of several model scores in a range, we consider procedures based on a single run of the inference scheme. Starting from an over-fitting mixture in a Bayesian set-ting, we investigate two strategies to eliminate superfluous components.We implement these strategies for mixtures of multiple scale distributions which exhibit a variety of shapes not necessarily elliptical while remaining analytical and tractable in multiple dimensions. A Bayesian formulation and a tractable inference procedure based on variational approximation are proposed. Preliminary results on simulated and real data show promising performance in terms of model selection and computational time. This work has been presented at RSSDS 2019 - Research School on Statistics and Data Science in Melbourne, Australia [33].

7.1.3. Approximate Bayesian Inversion for high dimensional problems

Participants: Florence Forbes, Benoit Kugler.

Joint work with: Sylvain Douté from Institut de Planétologie et d'Astrophysique de Grenoble (IPAG).

The overall objective is to develop a statistical learning technique capable of solving complex inverse problems in setting with specific constraints. More specifically, the challenges are 1) the large number of observations to be inverted, 2) their large dimension, 3) the need to provide predictions for correlated parameters and 4) the need to provide a quality index (eg. uncertainty).

In the context of Bayesian inversion, one can use a regression approach, such as in the so-called Gaussian Locally Linear Mapping (GLLiM) [7], to obtain an approximation of the posterior distribution. In some cases, exploiting this approximate distribution remains challenging, for example because of its multi-modality. In this work, we investigate the possible use of Importance Sampling to build on the standard GLLiM approach by improving the approximation induced by the method and to better handle the potential existence of multiple solutions. We may also consider our approach as a way to provide an informed proposal distribution as requested by Importance Sampling techniques. We experiment our approach on simulated and real data in the context of a photometric model inversion in planetology. Preliminary results have been presented at StatLearn 2019 [76]

7.1.4. MR fingerprinting parameter estimation via inverse regression

Participants: Florence Forbes, Fabien Boux, Julyan Arbel.

Joint work with: Emmanuel Barbier from Grenoble Institute of Neuroscience.

Magnetic resonance imaging (MRI) can map a wide range of tissue properties but is often limited to observe a single parameter at a time. In order to overcome this problem, Ma et al. introduced magnetic resonance fingerprinting (MRF), a procedure based on a dictionary of simulated couples of signals and parameters. Acquired signals called fingerprints are then matched to the closest signal in the dictionary in order to estimate parameters. This requires an exhaustive search in the dictionary, which even for moderately sized problems, becomes costly and possibly intractable . We propose an alternative approach to estimate more parameters at a time. Instead of an exhaustive search for every signal, we use the dictionary to learn the functional relationship between signals and parameters. . A dictionary-based learning (DBL) method was investigated to bypass inherent MRF limitations in high dimension: reconstruction time and memory requirement. The DBL method is a 3-step procedure: (1) a quasi-random sampling strategy to produce the dictionary, (2) a statistical inverse regression model to learn from the dictionary a probabilistic mapping between MR fingerprints and parameters, and (3) this mapping to provide both parameter estimates and their confidence levels. On synthetic data, experiments show that the quasi-random sampling outperforms the grid when designing the dictionary for inverse regression. Dictionaries up to 100 times smaller than usually employed in MRF yield more accurate parameter estimates with a 500 time gain. Estimates are supplied with a confidence index, well correlated with the estimation bias. On microvascular MRI data, results showed that dictionary-based methods (MRF and DBL) yield more accurate estimates than the conventional, closed-form equation, method. On MRI signals from tumor bearing rats, the DBL method shows very little sensitivity to the dictionary size in contrast to the MRF method. The proposed method efficiently reduces the number of required simulations to produce the dictionary, speeds up parameter estimation, and improve estimates accuracy. The DBL method also introduces a confidence index for each parameter estimate. Preliminary results have been presented at the third *Congrés National d'Imagerie du Vivant* (CNIV 2019) [53] and at the fourth *Congrés de la Société Française de Résonance Magnétique en Biologie et Médecine* (SFRMBM 2019) [54].

7.1.5. Characterization of daily glycemic variability in subjects with type 1 diabetes using a mixture of metrics

Participants: Florence Forbes, Fei Zheng.

Joint work with: Stéphane Bonnet from CEA Leti and Pierre-Yves Benhamou, Manon Jalbert from CHU Grenoble Alpes.

Glycemic variability is an important component of glycemic control for patients with type 1 diabetes. Glycemic variability (GV) must be taken into account in the efficacy of treatment of type 1 diabetes because it determines the quality of glycemic control, the risk of complication of the patient's disease. In a first study [24], our goal was to describe GV scores in patients with pancreatic islet transplantation (PIT) type 1 diabetes in the TRIMECO trial, and change of thresholds, for each index. predictive of success of PIT.

In a second study, we address the issue of choosing an appropriate measure of GV. Many metrics have been proposed to account for this variability but none is unanimous among physicians. The inadequacy of existing measurements lies in the fact that they view the variability from different aspects, so that no consensus has been reached among physicians as to which metrics to use in practice. Moreover, although glycemic variability, from one day to another, can show very different patterns, few metrics have been dedicated to daily evaluations. In this work [50], [30], a reference (stable-glycemia) statistical model is built based on a combination of daily computed canonical glycemic control metrics including variability. The metrics are computed for subjects from the TRIMECO islet transplantation trial , selected when their β-score (composite score for grading success) is greater than 6 after a transplantation. Then, for any new daily glycemia recording, its likelihood with respect to this reference model provides a multi-metric score of daily glycemic variability severity. In addition, determining the likelihood value that best separates the daily glycemia with a zero β-score from that greater than 6, we propose an objective decision rule to classify daily glycemia into "stable" or "unstable". The proposed characterization framework integrates multiple standard metrics and provides a comprehensive daily glycemic variability index, based on which, long term variability evaluations and investigations on the implicit link between variability and β-score can be carried out. Evaluation, in a daily glycemic variability classification task, shows that the proposed method is highly concordant to the experience of diabetologists. A multivariate statistical model is therefore proposed to characterize the daily glycemic variability of subjects with type 1 diabetes. The model has the advantage to provide a single variability score that gathers the information power of a number of canonical scores, too partial to be used individually. A reliable decision rule to classify daily variability measurements into stable or unstable is also provided.

7.1.6. Dirichlet process mixtures under affine transformations of the data

Participant: Julyan Arbel.

Joint work with: Riccardo Corradin and Bernardo Nipoti from Milano Bicocca, Italy.

Location-scale Dirichlet process mixtures of Gaussians (DPM-G) have proved extremely useful in dealing with density estimation and clustering problems in a wide range of domains. Motivated by an astronomical application, in this work we address the robustness of DPM-G models to affine transformations of the data, a natural requirement for any sensible statistical method for density estimation. In [63], we first devise a coherent prior specification of the model which makes posterior inference invariant with respect to affine transformation of the data. Second, we formalize the notion of asymptotic robustness under data transformation and show that mild assumptions on the true data generating process are sufficient to ensure that DPM-G models feature such a property. As a by-product, we derive weaker assumptions than those provided in the literature for ensuring posterior consistency of Dirichlet process mixtures, which could reveal of independent interest. Our investigation is supported by an extensive simulation study and illustrated by the analysis of an astronomical dataset consisting of physical measurements of stars in the field of the globular cluster NGC 2419.

7.1.7. Approximate Bayesian computation via the energy statistic

Participants: Julyan Arbel, Florence Forbes, Hongliang Lu.

Joint work with: Hien Nguyen, La Trobe University Melbourne Australia.

Approximate Bayesian computation (ABC) has become an essential part of the Bayesian toolbox for addressing problems in which the likelihood is prohibitively expensive or entirely unknown, making it intractable. ABC defines a quasi-posterior by comparing observed data with simulated data, traditionally based on some summary statistics, the elicitation of which is regarded as a key difficulty. In recent years, a number of data discrepancy measures bypassing the construction of summary statistics have been proposed, including the Kullback-Leibler divergence, the Wasserstein distance and maximum mean discrepancies. In this work [79], we propose a novel importance-sampling (IS) ABC algorithm relying on the so-called two-sample energy statistic. We establish a new asymptotic result for the case where both the observed sample size and the simulated data sample size increase to infinity, which highlights to what extent the data discrepancy measure impacts the asymptotic pseudo-posterior. The result holds in the broad setting of IS-ABC methodologies, thus generalizing previous results that have been established only for rejection ABC algorithms. Furthermore, we

propose a consistent V-statistic estimator of the energy statistic, under which we show that the large sample result holds. Our proposed energy statistic based ABC algorithm is demonstrated on a variety of models, including a Gaussian mixture, a moving-average model of order two, a bivariate beta and a multivariate g-and-k distribution. We find that our proposed method compares well with alternative discrepancy measures.

7.1.8. Industrial applications of mixture modeling

Participant: Julyan Arbel.

Joint work with: Kerrie Mengersen and Earl Duncan from QUT, School of Mathematical Sciences, Brisbane, Australia, and Clair Alston-Knox, Griffith University Brisbane, Australia, and Nicole White, Institute for Health and Biomedical Innovation, Brisbane, Australia.

In [61], we illustrate the wide diversity of applications of mixture models to problems in industry, and the potential advantages of these approaches, through a series of case studies. The first of these focuses on the iconic and pervasive need for process monitoring, and reviews a range of mixture approaches that have been proposed to tackle complex multimodal and dynamic or online processes. The second study reports on mixture approaches to resource allocation, applied here in a spatial health context but which are applicable more generally. The next study provides a more detailed description of a multivariate Gaussian mixture approach to a biosecurity risk assessment problem, using big data in the form of satellite imagery. This is followed by a final study that again provides a detailed description of a mixture model, this time using a nonparametric formulation, for assessing an industrial impact, notably the influence of a toxic spill on soil biodiversity.

7.2. Semi and non-parametric methods

7.2.1. Deep learning models to study the early stages of Parkinson's Disease

Participants: Florence Forbes, Veronica Munoz Ramirez, Virgilio Kmetzsch Rosa E Silva.

Joint work with: Michel Dojat from Grenoble Institute of Neuroscience.

Current physio-pathological data suggest that Parkinson's Disease (PD) symptoms are related to important alterations in subcortical brain structures. However, structural changes in these small structures remain difficult to detect for neuro-radiologists, in particular, at the early stages of the disease (*de novo* PD patients) [58], [43], [59]. The absence of a reliable ground truth at the voxel level prevents the application of traditional supervised deep learning techniques. In this work, we consider instead an anomaly detection approach and show that auto-encoders (AE) could provide an efficient anomaly scoring to discriminate *de novo* PD patients using quantitative Magnetic Resonance Imaging (MRI) data.

7.2.2. Estimation of extreme risk measures

Participants: Stephane Girard, Antoine Usseglio Carleve.

Joint work with: A. Daouia (Univ. Toulouse), L. Gardes (Univ. Strasbourg) and G. Stupfler (Univ. Nottingham, UK).

One of the most popular risk measures is the Value-at-Risk (VaR) introduced in the 1990's. In statistical terms, the VaR at level $\alpha \in (0,1)$ corresponds to the upper α-quantile of the loss distribution. The Value-at-Risk however suffers from several weaknesses. First, it provides us only with a pointwise information: VaR(α) does not take into consideration what the loss will be beyond this quantile. Second, random loss variables with light-tailed distributions or heavy-tailed distributions may have the same Value-at-Risk. Finally, Value-at-Risk is not a coherent risk measure since it is not subadditive in general. A first coherent alternative risk measure is the Conditional Tail Expectation (CTE), also known as Tail-Value-at-Risk, Tail Conditional Expectation or Expected Shortfall in case of a continuous loss distribution. The CTE is defined as the expected loss given that the loss lies above the upper α-quantile of the loss distribution. This risk measure thus takes into account the whole information contained in the upper tail of the distribution.

However, the asymptotic normality of the empirical CTE estimator requires that the underlying distribution possess a finite variance; this can be a strong restriction in heavy-tailed models which constitute the favoured class of models in actuarial and financial applications. One possible solution in very heavy-tailed models where this assumption fails could be to use the more robust Median Shortfall, but this quantity is actually just a quantile, which therefore only gives information about the frequency of a tail event and not about its typical magnitude. In [23], we construct a synthetic class of tail L_p-medians, which encompasses the Median Shortfall (for $p = 1$) and Conditional Tail Expectation (for $p = 2$). We show that, for $1 < p < 2$, a tail L_p-median always takes into account both the frequency and magnitude of tail events, and its empirical estimator is, within the range of the data, asymptotically normal under a condition weaker than a finite variance. We extrapolate this estimator, along with another technique, to proper extreme levels using the heavy-tailed framework. The estimators are showcased on a simulation study and on a set of real fire insurance data showing evidence of a very heavy right tail.

A possible coherent alternative risk measure is based on expectiles [6]. Compared to quantiles, the family of expectiles is based on squared rather than absolute error loss minimization. The flexibility and virtues of these least squares analogues of quantiles are now well established in actuarial science, econometrics and statistical finance. have recently received a lot of attention, especially in actuarial and financial risk man-agement. Their estimation, however, typically requires to consider non-explicit asymmetric leasts-quares estimates rather than the traditional order statistics used for quantile estimation. This makes the study of the tail expectile process a lot harder than that of the standard tail quantile process. Under the challenging model of heavy-tailed distributions, we derive joint weighted Gaussian approximations of the tail empirical expectile and quantile processes. We then use this powerful result to introduce and study new estimators of extreme expectiles and the standard quantile-based expected shortfall, as well as a novel expectile-based form of expected shortfall [22].

Both quantiles and expectiles were embedded in the more general class of L_p-quantiles [21] as the minimizers of a generic asymmetric convex loss function. It has been proved very recently that the only L_p-quantiles that are coherent risk measures are the expectiles. In [75], we work in a context of heavy tails, which is especially relevant to actuarial science, finance, econometrics and natural sciences, and we construct an estimator of the tail index of the underlying distribution based on extreme L_p-quantiles. We establish the asymptotic normality of such an estimator and in doing so, we extend very recent results on extreme expectile and L_p-quantile estimation. We provide a discussion of the choice of p in practice, as well as a methodology for reducing the bias of our estimator. Its finite-sample performance is evaluated on simulated data and on a set of real hydrological data. This work is submitted for publication.

7.2.3. Conditional extremal events

Participants: Stephane Girard, Antoine Usseglio Carleve.

Joint work with: G. Stupfler (Univ. Nottingham, UK), A. Ahmad, E. Deme and A. Diop (Université Gaston Berger, Sénégal).

The goal of the PhD thesis of Aboubacrene Ag Ahmad is to contribute to the development of theoretical and algorithmic models to tackle conditional extreme value analysis, *ie* the situation where some covariate information X is recorded simultaneously with a quantity of interest Y. In such a case, extreme quantiles and expectiles are functions of the covariate. In [13], we consider a location-scale model for conditional heavy-tailed distributions when the covariate is deterministic. First, nonparametric estimators of the location and scale functions are introduced. Second, an estimator of the conditional extreme-value index is derived. The asymptotic properties of the estimators are established under mild assumptions and their finite sample properties are illustrated both on simulated and real data.

As explained in Paragraph 7.2.2 , expectiles have recently started to be considered as serious candidates to become standard tools in actuarial and financial risk management. However, expectiles and their sample versions do not benefit from a simple explicit form, making their analysis significantly harder than that of quantiles and order statistics. This difficulty is compounded when one wishes to integrate auxiliary information about the phenomenon of interest through a finite-dimensional covariate, in which case the problem becomes the estimation of conditional expectiles. In [74], we exploit the fact that the expectiles of a distribution F are

in fact the quantiles of another distribution E explicitly linked to F, in order to construct nonparametric kernel estimators of extreme conditional expectiles. We analyze the asymptotic properties of our estimators in the context of conditional heavy-tailed distributions. Applications to simulated data and real insurance data are provided. The results are submitted for publication.

7.2.4. Estimation of the variability in the distribution tail

Participant: Stephane Girard.

Joint work with: L. Gardes (Univ. Strasbourg).

We propose a new measure of variability in the tail of a distribution by applying a Box-Cox transformation of parameter $p \geq 0$ to the tail-Gini functional. It is shown that the so-called Box-Cox Tail Gini Variability measure is a valid variability measure whose condition of existence may be as weak as necessary thanks to the tuning parameter p. The tail behaviour of the measure is investigated under a general extreme-value condition on the distribution tail. We then show how to estimate the Box-Cox Tail Gini Variability measure within the range of the data. These methods provide us with basic estimators that are then extrapolated using the extreme-value assumption to estimate the variability in the very far tails. The finite sample behavior of the estimators is illustrated both on simulated and real data. This work is submitted for publication [72].

7.2.5. Extrapolation limits associated with extreme-value methods

Participant: Stephane Girard.

Joint work with: L. Gardes (Univ. Strasbourg) and A. Dutfoy (EDF R&D).

The PhD thesis of Clément Albert (co-funded by EDF) is dedicated to the study of the sensitivity of extreme-value methods to small changes in the data and to their extrapolation ability. Two directions are explored:

(i) In [15], we investigate the asymptotic behavior of the (relative) extrapolation error associated with some estimators of extreme quantiles based on extreme-value theory. It is shown that the extrapolation error can be interpreted as the remainder of a first order Taylor expansion. Necessary and sufficient conditions are then provided such that this error tends to zero as the sample size increases. Interestingly, in case of the so-called Exponential Tail estimator, these conditions lead to a subdivision of Gumbel maximum domain of attraction into three subsets. In constrast, the extrapolation error associated with Weissman estimator has a common behavior over the whole Fréchet maximum domain of attraction. First order equivalents of the extrapolation error are then derived and their accuracy is illustrated numerically.

(ii) In [14], We propose a new estimator for extreme quantiles under the log-generalized Weibull-tail model, introduced by Cees de Valk. This model relies on a new regular variation condition which, in some situations, permits to extrapolate further into the tails than the classical assumption in extreme-value theory. The asymptotic normality of the estimator is established and its finite sample properties are illustrated both on simulated and real datasets.

7.2.6. Bayesian inference for copulas

Participants: Julyan Arbel, Marta Crispino, Stephane Girard.

We study in [16] a broad class of asymmetric copulas known as Liebscher copulas and defined as a combination of multiple—usually symmetric—copulas. The main thrust of this work is to provide new theoretical properties including exact tail dependence expressions and stability properties. A subclass of Liebscher copulas obtained by combining Fréchet copulas is studied in more details. We establish further dependence properties for copulas of this class and show that they are characterized by an arbitrary number of singular components. Furthermore, we introduce a novel iterative construction for general Liebscher copulas which *de facto* insures uniform margins, thus relaxing a constraint of Liebscher's original construction. Besides, we show that this iterative construction proves useful for inference by developing an Approximate Bayesian computation sampling scheme. This inferential procedure is demonstrated on simulated data.

7.2.7. Approximations of Bayesian nonparametric models

Participant: Julyan Arbel.

Joint work with: Stefano Favaro and Pierpaolo De Blasi from Collegio Carlo Alberto, Turin, Italy, Igor Prunster from Bocconi University, Milan, Italy, Caroline Lawless from Université Paris-Dauphine, France, Olivier Marchal from Université Jean Monnet.

For a long time, the Dirichlet process has been the gold standard discrete random measure in Bayesian nonparametrics. The Pitman–Yor process provides a simple and mathematically tractable generalization, allowing for a very flexible control of the clustering behaviour. Two commonly used representations of the Pitman–Yor process are the stick-breaking process and the Chinese restaurant process. The former is a constructive representation of the process which turns out very handy for practical implementation, while the latter describes the partition distribution induced. Obtaining one from the other is usually done indirectly with use of measure theory. In contrast, we propose in [25] an elementary proof of Pitman–Yor's Chinese Restaurant process from its stick-breaking representation.

In [17], we consider approximations to the popular Pitman-Yor process obtained by truncating the stick-breaking representation. The truncation is determined by a random stopping rule that achieves an almost sure control on the approximation error in total variation distance. We derive the asymptotic distribution of the random truncation point as the approximation error goes to zero in terms of a polynomially tilted positive stable random variable. The practical usefulness and effectiveness of this theoretical result is demonstrated by devising a sampling algorithm to approximate functionals of the-version of the Pitman–Yor process.

In [18], we approximate predictive probabilities of Gibbs-type random probability measures, or Gibbs-type priors, which are arguably the most "natural" generalization of the celebrated Dirichlet prior. Among them the Pitman–Yor process certainly stands out for the mathematical tractability and interpretability of its predictive probabilities, which made it the natural candidate in several applications. Given a sample of size n, in this paper we show that the predictive probabilities of any Gibbs-type prior admit a large n approximation, with an error term vanishing as $o(1/n)$, which maintains the same desirable features as the predictive probabilities of the Pitman–Yor process.

In [18], we prove a monotonicity property of the Hurwitz zeta function which, in turn, translates into a chain of inequalities for polygamma functions of different orders. We provide a probabilistic interpretation of our result by exploiting a connection between Hurwitz zeta function and the cumulants of the exponential-beta distribution.

7.2.8. Concentration inequalities

Participant: Julyan Arbel.

Joint work with: Olivier Marchal from Université Jean Monnet and Hien Nguyen from La Trobe University Melbourne Australia.

In [19], we investigate the sub-Gaussian property for almost surely bounded random variables. If sub-Gaussianity per se is de facto ensured by the bounded support of said random variables, then exciting research avenues remain open. Among these questions is how to characterize the optimal sub-Gaussian proxy variance? Another question is how to characterize strict sub-Gaussianity, defined by a proxy variance equal to the (standard) variance? We address the questions in proposing conditions based on the study of functions variations. A particular focus is given to the relationship between strict sub-Gaussianity and symmetry of the distribution. In particular, we demonstrate that symmetry is neither sufficient nor necessary for strict sub-Gaussianity. In contrast, simple necessary conditions on the one hand, and simple sufficient conditions on the other hand, for strict sub-Gaussianity are provided. These results are illustrated via various applications to a number of bounded random variables, including Bernoulli, beta, binomial, uniform, Kumaraswamy, and triangular distributions.

7.2.9. Extraction and data analysis toward "industry of the future"

Participants: Florence Forbes, Hongliang Lu, Fatima Fofana.

Joint work with: J. F. Cuccaro and J. C Trochet from Vi-Technology company.

The overall idea of this project with Vi-Technology is to work towards manufacturing processes where machines communicate automatically so as to optimize the process performance as a whole. Starting from the assumption that transmitted information is essentially of statistical nature, the role of MISTIS in this context was to identify what statistical methods might be useful for the printed circuits boards assembly industry. A first step was to extract and analyze data from two inspection machines in an industrial process making electronic cards. After a first extraction in the SQL database, the goal was to enlighten the statistical links between these machines. Preliminary experiments and results on the Solder Paste Inspection (SPI) step, at the beginning of the line, helped identifying potentially relevant variables and measurements (eg related to stencil offsets) to identify future defects and discriminate between them. More generally, we had access to two databases at both ends (SPI and Component Inspection) of the assembly process. The goal was to improve our understanding of interactions in the assembly process, find out correlations between defects and physical measures, generate proactive alarms so as to detect departures from normality.

7.2.10. Tracking and analysis of large population of dynamic single molecules

Participant: Florence Forbes.

Joint work with: Virginie Stoppin-Mellet from Grenoble Institute of Neuroscience, Vincent Brault from Laboratoire Jean Kuntzmann, Emilie Lebarbier from Nanterre University and Guy Bendao from AgroParisTech.

In the last decade, the number of studies using single molecule approaches has increased significantly. Thanks to technological progress and in particular with the development of TIRFM (Total Internal Reflection Fluorescence Microscopy), biologists can now observe single molecules at work. However, real time single molecule approaches remain mastered by a limited number of labs, and challenging obstacles have to be overcome before it becomes more broadly accessible. One important issue is the efficient detection and tracking of individual molecules in noisy images (low signal-to-noise ratio, SNR). Considering for example a TIRFM movie where single molecules stochastically appear and disappear at random positions, the low SNR implies that each individual molecule has to be detected at sub-pixel resolution over its local background and that this operation has to be repeated on each frame of the movie, thus requiring considerable amount of calculations. Procedures to detect single molecules are available, but they are mostly applicable to immobile molecules, are not statistically robust, and they often require an image processing that alters the quantitative signal information. In particular the intensity of a signal might be modified so that it becomes difficult to know the number of molecules associated with a specific signal. Crucial information such as the stoichiometry of the molecular complexes are then lost. Another challenging issue concerns data processing. Molecule tracking generate traces of time-dependent intensity fluctuations for each molecule. But single traces contain limited amount of information, and thus a very large number of traces must be analysed to extract general rules. In this context, the first aim of the present project was to provide a general procedure to track in real time transient interactions of a large number of biological molecules observed with TIRF microscopy and to generate traces of time-dependent intensity fluctuations. The second aim was to define a robust statistical approach to detect discrete events in a noisy time-dependent signal and extract parameters that describe the kinetics of these events. For this task we gathered expertise from biology (Grenoble Institute of Neuroscience) and statistics (Inria Mistis, LJK and AgroParisTech) in the context of a multidisciplinary project funded by the Grenoble data institute for 2 years.

7.3. Graphical and Markov models

7.3.1. Structure learning via Hadamard product of correlation and partial correlation matrices

Participants: Sophie Achard, Karina Ashurbekova, Florence Forbes.

Structure learning is an active topic nowadays in different application areas, i.e. genetics, neuroscience. Classical conditional independences or marginal independences may not be sufficient to express complex relationships. This work [39] is introducing a new structure learning procedure where an edge in the graph corresponds to a non zero value of both correlation and partial correlation. Based on this new paradigm, we define an estimator and derive its theoretical properties. The asymptotic convergence of the proposed graph estimator and its rate are derived. Illustrations on a synthetic example and application to brain connectivity are displayed.

7.3.2. Optimal shrinkage for robust covariance matrix estimators in a small sample size setting

Participants: Sophie Achard, Karina Ashurbekova, Florence Forbes, Antoine Usseglio Carleve.

When estimating covariance matrices, traditional sample covariance-based estimators are straightforward but suffer from two main issues: 1) a lack of robustness, which occurs as soon as the samples do not come from a Gaussian distribution or are contaminated with outliers and 2) a lack of data when the number of parameters to estimate is too large compared to the number of available observations, which occurs as soon as the covariance matrix dimension is greater than the sample size. The first issue can be handled by assuming samples are drawn from a heavy-tailed distribution, at the cost of more complex derivations, while the second issue can be addressed by shrinkage with the difficulty of choosing the appropriate level of regularization. In this work [66] we offer both a tractable and optimal framework based on shrinked likelihood-based M-estimators. First, a closed-form expression is provided for a regularized covariance matrix estimator with an optimal shrinkage coefficient for any sample distribution in the elliptical family. Then, a complete inference procedure is proposed which can also handle both unknown mean and tail parameter, in contrast to most existing methods that focus on the covariance matrix parameter requiring pre-set values for the others. An illustration on synthetic and real data is provided in the case of the t-distribution with unknown mean and degrees-of-freedom parameters.

7.3.3. Robust penalized inference for Gaussian Scale Mixtures

Participants: Sophie Achard, Karina Ashurbekova, Florence Forbes.

The literature on sparse precision matrix estimation is rapidly growing. Many strong methods are valid only for Gaussian variables. One of the most commonly used approaches in this case is glasso which aims to minimize the negative L1-penalized log-likelihood function. In practice, data may deviate from normality in various ways, outliers and heavy tails frequently occur that can severely degrade the Gaussian models performance. A natural solution is to turn to heavier tailed distributions that remain tractable. For this purpose, we propose [51] a penalized version of the EM algorithm for Gaussian Scale Mixtures.

7.3.4. Non parametric Bayesian priors for graph structured data

Participants: Florence Forbes, Julyan Arbel, Hongliang Lu.

We consider the issue of determining the structure of clustered data, both in terms of finding the appropriate number of clusters and of modelling the right dependence structure between the observations. Bayesian nonparametric (BNP) models, which do not impose an upper limit on the number of clusters, are appropriate to avoid the required guess on the number of clusters but have been mainly developed for independent data. In contrast, Markov random fields (MRF) have been extensively used to model dependencies in a tractable manner but usually reduce to finite cluster numbers when clustering tasks are addressed. Our main contribution is to propose a general scheme to design tractable BNP-MRF priors that combine both features: no commitment to an arbitrary number of clusters and a dependence modelling. A key ingredient in this construction is the availability of a stick-breaking representation which has the threefold advantage to allowing us to extend standard discrete MRFs to infinite state space, to design a tractable estimation algorithm using variational approximation and to derive theoretical properties on the predictive distribution and the number of clusters of the proposed model. This approach is illustrated on a challenging natural image segmentation task for which it shows good performance with respect to the literature. This work [77] will be presented as a poster at BayesComp2020 in Gainesville, Florida, USA, [78].

7.3.5. Bayesian nonparametric models for hidden Markov random fields on count variables and application to disease mapping

Participants: Julyan Arbel, Fatoumata Dama, Jean-Baptiste Durand, Florence Forbes.

Hidden Markov random fields (HMRFs) have been widely used in image segmentation and more generally, for clustering of data indexed by graphs. Dependent hidden variables (states) represent the cluster identities and determine their interpretations. Dependencies between state variables are induced by the notion of neighborhood in the graph. A difficult and crucial problem in HMRFs is the identification of the number of possible states K. Recently, selection methods based on Bayesian non parametric priors (Dirichlet processes) have been developed. They do not assume that K is bounded a priori, thus allowing its adaptive selection with respect to the quantity of available data and avoiding costly systematic estimation and comparison of models with different fixed values for K. Our previous work [77] has focused on Bayesian nonparametric priors for HMRFs and continuous, Gaussian observations. In this work, we consider extensions to discrete observed data typically issued from counts. We define and implement Bayesian nonparametric models for HMRFs with Poisson distributed observations. As an illustration, we propose a new disease mapping model for epidemiology. The inference is done by Variational Bayesian Expectation Maximization (VBEM). Results on synthetic data sets suggest that our model is able to recover the true number of risk levels (clusters) and to provide a good estimation of the true risk level partition. Application on real data then also shows satisfying results.

As a perspective, Bayesian nonparametric models for hidden Markov random fields could be extended to non-Poissonian models (particularly to account for zero-inflated and over-/under-dispersed cases of application) and to regression models.

7.3.6. Hidden Markov models for the analysis of eye movements

Participants: Jean-Baptiste Durand, Brice Olivier, Sophie Achard.

This research theme is supported by a LabEx PERSYVAL-Lab project-team grant.
Joint work with: Anne Guérin-Dugué (GIPSA-lab) and Benoit Lemaire (Laboratoire de Psychologie et Neurocognition)

In the last years, GIPSA-lab has developed computational models of information search in web-like materials, using data from both eye-tracking and electroencephalograms (EEGs). These data were obtained from experiments, in which subjects had to decide whether a text was related or not to a target topic presented to them beforehand. In such tasks, reading process and decision making are closely related. Statistical analysis of such data aims at deciphering underlying dependency structures in these processes. Hidden Markov models (HMMs) have been used on eye-movement series to infer phases in the reading process that can be interpreted as strategies or steps in the cognitive processes leading to decision. In HMMs, each phase is associated with a state of the Markov chain. The states are observed indirectly though eye-movements. Our approach was inspired by Simola *et al.* (2008) [86], but we used hidden semi-Markov models for better characterization of phase length distributions (Olivier *et al.*, 2017) [85]. The estimated HMM highlighted contrasted reading strategies, with both individual and document-related variability. New results were obtained in the standalone analysis of the eye-movements. A comparison between the effects of three types of texts was performed, considering texts either closely related, moderately related or unrelated to the target topic.

Then, using the restored state values, statistical characteristics of EEGs were compared according to strategies, brain wave frequencies and EEG channels (i.e., location on scalp). Differences in variance and correlations related to strategy changes were highlighted. Dependency graphs interpreted as maps of functional brain connectivity were estimated for each strategy and frequency and their changes were interpreted.

These results were published in Brice Olivier's PhD manuscript [12]. Although the approach was sufficient to highlight significant discrimination of strategies, it suffered from somewhat overlapping eye-movement characteristics over strategies. As a result, high uncertainty in the phase changes arose, which could induce underestimation of EEG and eye movement abilities to discriminate strategies.

This is why we developed integrated models coupling EEG and eye movements within one single HMM for better identification of strategies. Here, the coupling incorporated some delay between transitions in both EEG and eye-movement state sequences, since EEG patterns associated to cognitive processes occur lately with respect to eye-movement state switches. Moreover, EEGs and scanpaths were recorded with different time resolutions, so that some resampling scheme had to be added into the model, for the sake of synchronizing both processes. An associated EM algorithm for maximum likelihood parameter estimation was derived.

Our goal for this coming year is to implement and validate our coupled model for jointly analyzing eye-movements and EEGs in order to improve the discrimination of reading strategies.

7.3.7. Comparison of initialization strategies in the EM algorithm for hidden Semi-Markov processes

Participants: Jean-Baptiste Durand, Brice Olivier.

This research theme is supported by a LabEx PERSYVAL-Lab project-team grant.
Joint work with: Anne Guérin-Dugué (GIPSA-lab)

In Subsection 7.3.6 , hidden semi-Markov models (HSMMs) were used to infer reading strategies from eye-movement and EEG time series. Model parameters were estimated by the EM algorithm. Its principle is to build a sequence of parameters with increasing likelihood values, starting from a starting point. The impact of this starting point has not been investigated in the case of HSMMs; this is why we aimed at developing and assessing an initialization method based on the available sequence lengths [48]. This consists in randomly choosing a number of transitions and then, uniformly-distributed transition times given the number of transitions. These transition times break the sequences into segements and assign uniformly-distributed states to each segment with the constraint that two consecutive states should be different.

The method was compared to other initialization strategies and was shown to be efficient on several data sets with multiple categorical sequences.

7.3.8. Lossy compression of tree structures

Participant: Jean-Baptiste Durand.

Joint work with: Christophe Godin and Romain Azaïs (Inria Mosaic)

The class of self-nested trees presents remarkable compression properties because of the systematic repetition of subtrees in their structure. The aim of our work is to achieve compression of any unordered tree by finding the nearest self-nested tree. Solving this optmization problem without more assumptions is conjectured to be an NP-complete or NP-hard problem. In [40], we firstly provided a better combinatorial characterization of this specific family of trees. In particular, we showed from both theoretical and practical viewpoints that complex queries can be quickly answered in self-nested trees compared to general trees. We also presented an approximation algorithm of a tree by a self-nested one that can be used in fast prediction of edit distance between two trees.

Our goal for this coming year is to apply this approach to quantify the degree of self-nestedness of several plant species and extend first results obtained on rice panicles stating that near self-nestedness is a fairly general pattern in plants.

7.3.9. Bayesian neural networks

Participants: Julyan Arbel, Mariia Vladimirova.

Joint work with: Pablo Mesejo from University of Granada, Spain, Jakob Verbeek from Inria Grenoble Rhône-Alpes, France.

We investigate in [45] deep Bayesian neural networks with Gaussian priors on the weights and ReLU-like nonlinearities, shedding light on novel sparsity-inducing mechanisms at the level of the units of the network, both pre- and post-nonlinearities. The main thrust of the paper is to establish that the units prior distribution becomes increasingly heavy-tailed with depth. We show that first layer units are Gaussian, second layer units are sub-Exponential, and we introduce sub-Weibull distributions to characterize the deeper layers units. Bayesian neural networks with Gaussian priors are well known to induce the weight decay penalty on the weights. In contrast, our result indicates a more elaborate regularisation scheme at the level of the units. This result provides new theoretical insight on deep Bayesian neural networks, underpinning their natural shrinkage properties and practical potential.

NANO-D Team

6. New Results

6.1. Reconstructing molecular shapes from SAXS data

We are working on a novel method to reconstruct the three-dimensional shape of a molecule from low-resolution experimental data. The structural data we are currently focusing on is obtained through small-angle X-ray scattering (SAXS) experiments, but we plan to also consider small-angle neutron scattering (SANS) data. Our *ab initio* reconstruction method is inspired by iterative phase-retrieval algorithms that can produce an image for an object when only the amplitudes of its Fourier transform are known and the phases are unknown. In our context, the X-ray scattering amplitudes associated with a molecule are the Fourier transform of its electron density. The novelty of our approach resides in the use of spherical harmonics expansions, which will allow performing the whole reconstruction process in Fourier space—contrary to existing methods that iterate between Fourier space and real space—for an improved computational efficiency. Our method is being implemented within the software PEPSI (polynomial expansions of protein structures and interactions).

6.2. Docking of cyclic molecules

In 2018 we participated in a docking Grand Challenge 4, which was organized by the Drug Design Data Resource (D3R) group. The goal was to predict correct poses and affinities of ligands binding the beta secretase 1 (BACE) receptor. Most of the ligands were macrocyclic. The challenge answers and results were fully released in February 2019. Upon that we started analyzing our protein-ligand docking strategy and wrote a corresponding paper [5].

6.3. Convex-PL and entropy

During the 2019 we continued developing our knowledge-based scoring function for protein-ligand interactions called Convex-PL [44]. We introduced a new descriptor characterizing side-chain entropy, tried two ways of computing the solvent-accessible surface area (SASA) descriptors with either using buried SASA, or SASA of those atoms that are contributing to the protein-ligand interaction according to the scoring function design (i.e which are within the cutoff distance of an interaction potential). Using these descriptors alongwith Convex-PL score and ligand side chain entropy descriptor, we trained a ridge regression model to predict binding affinities. This modification of Convex-PL was assessed on the CASF Benchmark 2016 released in the end of 2018, and on a subset of the DUD benchmark [42]. We are currently preparing the corresponding manuscript for submission.

6.4. Orientational potential for small molecules

With Pablo Chacon and Karina Dos Santos Machado we have developed a novel statistical protein-ligand scoring function called KORP-PL. KORP-PL is based on the coarse grained backbone-only representation of protein that is very common in protein structure modeling. It is based on ideas implemented by Pablo Chacon in the preceding scoring function for protein quality assessment called KORP [49], where each amino acid residue was characterized with a 3D oriented frame. We kept this representation for protein residues and computed the distances and angles between the oriented frames and points that describe ligand atoms. Using this data, we then derived the scoring function statistically in the same way, as it was done for protein-protein interactions in KORP. We also ran an optimization procedure to compute weights for each residue-atom interaction that would minimize the difference between predicted and experimental binding constants. We have assessed KORP-PL on several benchmarks [47], [67], one of which was manually derived from the user-submitted data of the D3R Challenges 2 [35], 3 [36], and 4. On all of them it performed exceptionally in pose prediction despite being a coarse-grained scoring function. We are currently preparing the corresponding manuscript for submission.

6.5. Analysis of a deep-learning architecture for fold recognition

Deep learning has recently demonstrated outstanding capabilities in classical pattern recognition problems. It has also obtained a tremendous success in the very recent protein structure prediction tasks. This work studies recurrent structural patterns in protein structures recognized by a deep neural network. We demonstrated that neural networks can automatically learn a vast amount of chemo-structural features with only a very little amount of human supervision. For example, our architecture correctly learns atomic, amino acid, and also higher-level molecular descriptors. The network architecture and the results are available at https://team.inria.fr/nano-d/software/Ornate/.

6.6. Controlled-advancement rigid-body optimization of nanosystems

In this study, we proposed a novel optimization algorithm, with application to the refinement of molecular complexes. Particularly, we considered optimization problem as the calculation of quasi-static trajectories of rigid bodies influenced by the inverse-inertia-weighted energy gradient and introduce the concept of advancement region that guarantees displacement of a molecule strictly within a relevant region of conformational space. The advancement region helps to avoid typical energy minimization pitfalls, thus, the algorithm is suitable to work with arbitrary energy functions and arbitrary types of molecular complexes without necessary tuning of its hyper-parameters. Our method, called controlled-advancement rigid-body optimization of nanosystems (Carbon), is particularly useful for the large-scale molecular refinement, as for example, the putative binding candidates obtained with protein–protein docking pipelines. Implementation of Carbon with user-friendly interface is available in the SAMSON platform for molecular modeling at https://www.samson-connect.net. The method was published in [10].

6.7. SAXS- and SANS-assisted modeling of proteins

We collaborated on data-assisted modeling of a KRAB-domain associated protein 1. Our work sheds light on its overall organization and combins solution scattering diffraction data, integrative modeling and single-molecule experiments [3].

We also participated in a combination of coarse-grained molecular dynamics simulations with previously measured small-angle scattering data to study the conformation of three-domain protein TIA-1 in solution. More precisely, we contributed with a specifically developed version of the Pepsi-SANS code. Our results suggest a general strategy for studying the conformation of multi-domain proteins in solution that combines coarse-grained simulations with small-angle X-ray scattering data that are generally most easy to obtain [13].

6.8. Predicting protein functional motions

Large macromolecules, including proteins and their complexes, very often adopt multiple conformations. Some of them can be seen experimentally, for example with X-ray crystallography or cryo-electron microscopy. This structural heterogeneity is not occasional and is frequently linked with specific biological function. Thus, the accurate description of macromolecular conformational transitions is crucial for understanding fundamental mechanisms of life's machinery.

We report on a real-time method to predict such transitions by extrapolating from instantaneous eigen-motions, computed using the normal mode analysis, to a series of twists. We demonstrate the applicability of our approach to the prediction of a wide range of motions, including large collective opening-closing transitions and conformational changes induced by partner binding. We also highlight particularly difficult cases of very small transitions between crystal and solution structures. Our method guaranties preservation of the protein structure during the transition and allows to access conformations that are unreachable with classical normal mode analysis. We provide practical solutions to describe localized motions with a few low-frequency modes and to relax some geometrical constraints along the predicted transitions. This work opens the way to the systematic description of protein motions, whatever their degree of collectivity. Our method is available as a part of the NOn-Linear rigid Block (NOLB) package at https://team.inria.fr/nano-d/software/nolb-normal-modes/ [12].

6.9. Protein structure prediction experiments

We participated in the CAPRI Round 46, the third joint CASP-CAPRI protein assembly prediction challenge. The Round comprised a total of 20 targets including 14 homo-oligomers and 6 heterocomplexes. Eight of the homo-oligomer targets and one heterodimer comprised proteins that could be readily modeled using templates from the Protein Data Bank, often available for the full assembly. The remaining 11 targets comprised 5 homodimers, 3 heterodimers, and two higher-order assemblies. These were more difficult to model, as their prediction mainly involved “ab-initio” docking of subunit models derived from distantly related templates [7].

NECS Team

7. New Results

7.1. Network systems: modeling, analysis, and estimation

7.1.1. Network reduction towards a scale-free structure preserving physical properties

Participants: N. Martin, P. Frasca, C. Canudas-de-Wit [Contact person].

In the context of the ERC project, we are addressing a problem of graph reduction, where a given arbitrary weighted graph is reduced to a (smaller) scale-free graph while preserving a consistency with the initial graph and some physical properties. This problem can be formulated as a minimization problem. We give specifications to this general problem to treat a particular case: to this end we define a metric to measure the scale-freeness of a graph and another metric to measure the similarity between two graphs with different dimensions, based on a notion of spectral centrality. Moreover, through the reduction we also preserve a property of mass conservation (essentially, Kirchoff's first law). We study the optimization problem and, based on the gained insights, we derive an algorithm allowing to find an approximate solution. Finally, we have simulated the algorithm both on synthetic networks and on real-world examples of traffic networks that represent the city of Grenoble. These results are presented in [22] and in [48].

7.1.2. Boundary Control for Output Regulation in Scale-FreePositive Networks

Participants: D. Nikitin, C. Canudas-de-Wit [Contact person], P. Frasca.

This work addresses the problem of controlling aggregate quantities in large networks. More precisely, we deal with the problem of controlling a scalar output of a large-scale positive scale-free network to a constant reference value. We design an output-feedback controller such that no information about state vector or system matrices is needed. This controller can have arbitrary positivegains, and only one sufficient sign condition on system matricesshould be satisfied. This controller can be used to regulate the average state in a large-scale network with control applied to boundary nodes of the domain [51].

7.1.3. A functional approach to target controllability of networks

Participants: C. Commault, J. Van Der Woude [TU Delft], P. Frasca [Contact person].

In the control of networks, it is natural to consider the problem of controlling a limited number of *target nodes* of a network. Equivalently, we can see this problem as controlling the target variables of a structured system, where the state variables of the system are associated to the nodes of the network. We deal with this problem from a different point of view as compared to most recent literature. Indeed, instead of considering controllability in the Kalman sense, that is, as the ability to drive the target states to a desired value, we consider the stronger requirement of driving the target variables as time functions. The latter notion is called functional target controllability. We think that restricting the controllability requirement to a limited set of important variables justifies using a more accurate notion of controllability for these variables. Remarkably, the notion of functional controllability allows formulating very simple graphical conditions for target controllability in the spirit of the structural approach to controllability. The functional approach enables us, moreover, to determine the smallest set of steering nodes that need to be actuated to ensure target controllability, where these steering nodes are constrained to belong to a given set. We show that such a smallest set can be found in polynomial time. We are also able to classify the possible actuated variables in terms of their importance with respect to the functional target controllability problem. This research is reported in [16].

7.1.4. Cyber-Physical Systems: a control-theoretic approach to privacy and security

Participants: F. Garin [Contact person], A. Kibangou, S. Gracy [KTH Stockholm], S.m. Fosson [Politecnico di Torino].

Cyber-physical systems are composed of many simple components (agents) with interconnections giving rise to a global complex behaviour. One line or research on security of cyber-physical systems models an attack as an unknown input being maliciously injected in the system. We study linear network systems, and we aim at characterizing input and state observability (ISO), namely the conditions under which both the whole network state and the unknown input can be reconstructed from some measured local states. We complement the classical algebraic characterizations with novel structural results, which depend only on the graph of interactions (equivalently, on the zero pattern of the system matrices). More precisely, we obtain two kinds of results: structural results, true for almost all interaction weights, and strongly structural results, true for all non-zero interaction weights. Our results in 2019 concern structural and strongly structural ISO for time-varying systems [19], strongly structural ISO for time-invariant systems [46]. Moreover in [44] we study delay-L left-invertibility, where the input reconstruction is allowed to take L time steps instead of requiring immediate reconstruction in a single step. We obtain preliminary results for structural delay-L left-invertibility, which include a full characterization for the case where the input is scalar, and for the cases where L is one and two, while the general case remains an open problem. When the conditions for ISO are satisfied, one can run well-known algorithms in the same vein as a Kalman filter, in order to reconstruct the state and the unknown input from noisy measurements. In [43], we consider cases where the system is not ISO, and we exploit compressive sensing techniques in order to obtain nevertheless a unique reconstruction of the input, under the assumption that the input is highly sparse (e.g., when only one or few states are under attack, albeit the attack position is unknown).

7.1.5. Collaborative monitoring of network structural robustness

Participants: A. Kibangou [Contact person], T.m.d. Tran [Univ. of Danang].

Interacting systems can be naturally viewed as networks modelled by graphs, whose vertices represent the components of the system while edges stand for the interactions between these components. The efficiency of a network of a network can be evaluated through its functional robustness and structural robustness. The former usually stands for robustness against noise while the latter is related to the network performance despite changes in network topology (node or edge failure). Structural robustness has been an important topic in various domains: in distribution networks (e.g. power or water distribution networks), breakdowns can prevent service to customers; in communication networks, equipment failures may disrupt the network and block users from communicating; in contact networks, removing nodes (persons) by means of vaccination can prevent epidemic propagation. In [31] we have considered the critical threshold of a network and the effective graph resistance (Kirchhoff index) of a sub-graph characterizing the interconnection of sub-networks, that are partitioned from the given network as robustness metric. In which, the critical threshold depends only on the two first moments of the degree distribution while the Kirchhoff index can be computed with Laplacian eigenvalues. Therefore, we show how to estimate jointly the Laplacian eigenvalues and the two first moments of the degree distribution in a distributed way.

7.1.6. Estimation of the average state in large scale networks

Participants: A. Kibangou [Contact person], C. Canudas-de-Wit, U. Niazi, D. Deplano [Univ. Cagliari].

State estimation for monitoring large-scale systems requires tremendous amounts of computational and sensing resources, which is impractical in most applications. However, knowledge of some aggregated quantity of the state suffices in several applications. Processes over physical networks such as traffic, epidemic spread, and thermal control are examples of large-scale systems. Due to the diffusive nature of these systems, the average state is usually sufficient for monitoring purposes. For instance, estimating the average traffic density in some sector of a traffic network helps to monitor the congestion effectively. In the event of an epidemic, estimating the average proportion of infected people over several towns, which are interconnected through people commuting for work or other purposes, helps to devise the preventive measures for controlling the epidemic spread. For the temperature regulation of a building, the thermistors can only be placed either on the walls or the roof, therefore, estimating the average temperature of the interior of a large corridor is crucial. Other examples include the averaging systems such as opinion networks and wireless sensor networks where the average state is of paramount importance. In [40] we address observability and detectability of the average

state of a network system when few gateway nodes are available. To reduce the complexity of the problem, the system is transformed to a lower dimensional state space by aggregation. The notions of average observability and average detectability are then defined, and the respective necessary and sufficient conditions are provided. In [25] we provide a computationally tractable necessary and sufficient condition for the existence of an average state observer for large-scale linear time-invariant (LTI) systems. Two design procedures, each with its own significance, are proposed. When the necessary and sufficient condition is not satisfied, a methodology is devised to obtain an optimal asymptotic estimate of the average state. In particular, the estimation problem is addressed by aggregating the unmeasured states of the original system and obtaining a projected system of reduced dimension. This approach reduces the complexity of the estimation task and yields an observer of dimension one. Moreover, it turns out that the dimension of the system also does not affect the upper bound on the estimation error.

7.1.7. Structure-based Clustering Algorithm for Model Reduction of Large-scale Network Systems

Participants: C. Canudas-de-Wit [Contact person], U. Niazi, J. Scherpen [Univ. Groningen], X. Cheng [Univ. Groningen].

In [41], A model reduction technique is presented that identifies and aggregates clusters in a large-scale network system and yields a reduced model with tractable dimension. The network clustering problem is translated to a graph reduction problem, which is formulated as a minimization of distance from lumpability. The problem is a non-convex, mixed-integer optimization problem and only depends on the graph structure of the system. We provide a heuristic algorithm to identify clusters that are not only suboptimal but are also connected, that is, each cluster forms a connected induced subgraph in the network system.

7.2. Control of multi-agent systems and opinion dynamics

7.2.1. Robust average consensus over unreliable networks

Participants: F. Acciani [Univ. Twente], P. Frasca [Contact person], G. Heijenk [Univ. Twente], A. Stoorvogel [Univ. Twente].

Packet loss is a serious issue in wireless consensus networks, as even few failures might prevent a network to converge to the desired consensus value. In the last four years, we have devised some possible ways to compensate for the errors caused by packet collisions, by modifying the updating weights. Since these modifications may result in a reduced convergence speed, a gain parameter is used to increase the convergence speed, and an analysis of the stability of the network is performed, leading to a criterion to choose such gain to guarantee network stability. For the implementation of the compensation method, we propose a new communication algorithm, which uses both synchronous and asynchronous mechanisms to achieve average consensus and to deal with uncertainty in packet delivery. The paper [11] provides a complete account of our results.

7.2.2. Message-passing computation of harmonic influence in social networks

Participants: W. S. Rossi [Univ. Groningen], P. Frasca [Contact person].

In the study of networks, identifying the most important nodes is of capital importance. The concept of Harmonic Influence has been recently proposed as a metric for the importance of nodes in a social network. This metric evaluates the ability for one node to sway the opinions of the other nodes in the network, under the assumption of a linear diffusion of opinions in the network. A distributed message passing algorithm for its computation has been proposed by Vassio et al., 2014, but its convergence guarantees were limited to trees and regular graphs. In [29], we prove that the algorithm converges on general graphs.

7.2.3. Hybrid models of opinion dynamics

Participants: P. Frasca [Contact person], S. Tarbouriech [LAAS CNRS], L. Zaccarian [LAAS CNRS].

Hybrid dynamical systems are a promising framework to model social interactions. In this research line, we are beginning to use tools from the theory of hybrid systems to study opinion dynamics on networks with opinion-dependent connectivity. According to the hybrid framework, our dynamics are represented by the combination of continuous flow dynamics and discrete jump dynamics. The flow embodies the attractive forces between the agents and is defined by an ordinary differential equation whose right-hand side is a Laplacian, whereas the jumps describe the activation or deactivation of the pairwise interactions between agents. We first reformulate the classical Hegselmann–Krause model in this framework and then define a novel interaction model, which has the property of being scale-invariant. We study the stability and convergence properties of both models by a Lyapunov analysis, showing convergence and clusterization of opinions [18].

7.2.4. Stability of Metabolic Networks

Participants: F. Garin [Contact person], B. Piccoli [Rutgers Univ. Camden], N. Merrill [Rutgers Univ. Camden], Z. An [Rutgers Univ. Camden], S. Mc Quade [Rutgers Univ. Camden].

Quantitative Systems Pharmacology (QSP) aims to gain more information about a potential drug treatment on a human patient before the more expensive stages of development begin. QSP models allow us to perform insilico experiments on a simulated metabolic system that predicts the response ofperturbing a flux. The methodology named LIFE (Linear-in-Flux Expressions) was developed with the purpose of simulating and analyzing large metabolic systems. These systems can be associated to directed graphs: the edges represent the reaction rates (fluxes), and the vertices represent quantities of chemical compounds (metabolites). In [23], we study LIFE systems, addressing two main problems: 1. for fixed metabolite levels, find all fluxes for which the metabolite levels are an equilibrium, and 2. for fixed fluxes, find all metabolite levels which are equilibria for the system. We show how stability analysis from the fields of network flows, compartmental systems, control theory and Markov chains apply to LIFE systems.

7.3. Transportation networks and vehicular systems

7.3.1. Heterogeneity in synchronization: an adaptive control approach, with applications to vehicle platooning

Participants: S. Baldi [Univ. Delft], P. Frasca [Contact person].

Heterogeneity is a substantial obstacle to achieve synchronisation of interconnected systems (that is, in control). In order to overcome heterogeneity, advanced control techniques are needed, such as the use of "internal models" or of adaptive techniques. In a series of papers motivated by multi-vehicle platooning and coordinated autonomous driving, we have explored the application of adaptive control techniques. Our results cover both the cases of state-feedback [12] and of output-feedback [14], under the assumption that the topology of the interconnections has no circuits. Further investigation on relaxing this restrictive assumption is in progress. We also showed that agents need no leader to synchronise, even in presence of heterogeneity [13].

7.3.2. Stability of vehicle platoons with AVs

Participants: V. Giammarino [Univ. Delft], M. Lv [Univ. Delft], P. Frasca [Contact person], M.l. Delle Monache, S. Baldi [TU Delft].

A key notion to understand the impact of Autonomous Vehicles on traffic is the notion of *stability* of the vehicle collective motion. In this line of research, we have sought criteria to determine when stop-and-go waves form in platoons of human-driven vehicles, and when they can be dissipated by the presence of an autonomous vehicle. Our analysis takes the start from the observation that the standard notion of string/ring stability definition, which requires uniformity with respect to the number of vehicles in the platoon, is too demanding for a mixed traffic scenario. The setting under consideration is the following: the vehicles run along a ring road and the human-driven vehicles obey a combined follow-the-leader and optimal velocity model, while the autonomous vehicle obeys an appropriately designed model. The criteria are tested on a linearized version of the resulting platoon dynamics and simulation tests using nonlinear model are carried out [45].

7.3.3. Control and estimation using autonomous vehicles

Participants: R. Stern [Vanderbilt University], S. Cui [Temple University], M.l. Delle Monache [Contact person], T. Liard, Y. Chen [Vanderbilt University], R. Bhadani [University of Arizona], M. Bunting [University of Arizona], M. Churchill [UIUC], N. Hamilton [Vanderbilt University], R. Haulcy [Yale University], H. Pohlmann [Temple University], F. Wu [UC Berkeley], B. Piccoli [Rutgers University], B. Seibold [Temple University], J. Sprinkle [University of Arizona], D.b. Work [Vanderbilt University].

It is anticipated that in the near future, the penetration rate of vehicles with some autonomous capabilities will increase on roadways. In [30], we analyze the potential reduction of vehicular emissions caused by the whole traffic stream, when a small number of autonomous vehicles are designed to stabilize the traffic flow and dampen stop-and-go waves. To demonstrate this, vehicle velocity and acceleration data are collected form a series of field experiments that use a single autonomous-capable vehicle to dampen traffic waves on a circular ring road with 20 to 21 human-piloted vehicles. From the experimental data, vehicle emissions (hydrocarbons, carbon monoxide, carbon dioxide, and nitrogen oxides) are estimated using the MOVES emissions models. We find that vehicle emissions of the entire fleet may be reduced by between 15% (for carbon dioxide) and 73% (for nitrogen oxides) when stop-and-go waves are reduced or eleiminated by the dampening action of the autonomous vehicle in the flow of human drivers. This is possible if a small fraction (5%) of vehicles are autonomous and designed to actively dampen traffic waves. In [57], we look at the problem of traffic control in which an autonomous vehicle is used to regulate human piloted traffic to dissipate stop and go traffic waves. We investigate the controllability of well-known microscopic traffic flow models: i) the Bando model (also known as the optimal velocity model), ii) the follow-the-leader model and iii) a combined optimal velocity – follow the leader model. Based on the controllability results, we propose three control strategies for an autonomous vehicles to stabilize the human piloted traffic. After, we simulate the control effects on the microscopic models of human drivers in numerical experiments to quantify the potential benefits of the controllers. Based on the simulations, finally we conduct a field experiment with 22 human drivers and a fully autonomous-capable vehicle, to assess the feasibility of autonomous vehicle based traffic control on real human piloted traffic. We show that both in simulation and in the field test an autonomous vehicle is able to dampen waves generated by 22 cars, and that as a consequence, the total fuel consumption of all vehicles is reduced by up to 20%. In [17], we consider a partial differential equation – ordinary differential equation system to describe the dynamics of traffic with autonomous vehicles. In the model the bulk flow is represented by a scalar conservation law, while each autonomous vehicle is described by a car following model. The autonomous vehicles act as tracer vehicles in the flow and collect measurements along their trajectory to estimate the bulk flow. The main result is to prove theoretically and show numerically how to reconstruct the correct traffic density using only the measurements from the autonomous vehicles.

7.3.4. Two-dimensional traffic flow models

Participants: S. Mollier, M.l. Delle Monache, C. Canudas-de-Wit [Contact person], B. Seibold [Temple University].

In [24], we introduce a new traffic flow model for a dense urban area. We consider a two-dimensional conservation law in which the velocity magnitude is given by the fundamental diagram and the velocity direction is constructed following the network geometry. The model is validated using synthetic data from Aimsun and a reconstruction technique to recover the 2D density from the data of individual vehicles is proposed. In [50], [49], we introduce a two dimensional and multi-layer traffic model with a new planning and decision making method in large scale traffic networks for predicting how traffic evolves in special events, emergencies and changes in the city mobility demands. The proposed method is based on a 2-D aggregated traffic model for large scale traffic networks which describes traffic evolution as a fluid in two space dimensions extended with additional state density variables, each one associated to a particular layer describing vehicles evolving in different directions. The model is a 2D-PDE described by a system of conservation laws. For this specific case, the resulting PDE is not anymore hyperbolic as typically the LWR model but results in a hybrid hyperbolic-elliptic PDE depending on the density level. In this case, usual numerical schemes may be not valid and often lead to oscillation in the solution. Thus, we consider a high order numerical scheme to improve the

numerical solution. Finally, the model is used to predict how the typical traffic evolution will be impacted in particular scenarios like special events or changes in demands.

7.3.5. High-fidelity vehicle trajectory data

Participants: F. Wu [UC Berkeley], R. Stern [Vanderbilt University], S. Cui [Temple University], M.l. Delle Monache [Contact person], R. Bhadani [University of Arizona], M. Bunting [University of Arizona], M. Churchill [UIUC], N. Hamilton [Vanderbilt University], R. Haulcy [Yale University], B. Piccoli [Rutgers University], J. Sprinkle [University of Arizona], D.b. Work [Vanderbilt University], B. Seibold [Temple University].

High fidelity-vehicle trajectory data is becoming increasingly important in traffic modeling, especially to capture dynamic features such as stop-and-go waves. In [34], we present data collected in a series of eight experiments on a circular track with human drivers. The data contains smooth flowing and stop-and-go traffic conditions. The vehicle trajectories are collected using a panoramic 360-degree camera, and fuel rate data is recorded via an on-board diagnostics scanner installed in each vehicle. The video data from the 360-degree camera is processed with an offline unsupervised algorithm to extract vehicle trajectories from experimental data. The trajectories are highly accurate, with a mean positional bias of less than 0.01 m and a standard deviation of 0.11 m. The velocities are also validated to be highly accurate with a bias of 0.02 m/s and standard deviation of 0.09 m/s.

7.3.6. Robust tracking control design for fluid traffic dynamics

Participants: L. Tumash, C. Canudas-de-Wit [contact person], M.l. Delle Monache.

In [53] we analyze the boundary control of the traffic system described by the LWR model with a triangular fundamental diagram and a space-dependent indomain unknown disturbance, which can be described as an inhomogeneous transport equation. The controller design strategy aims first at stabilizing the deviation from the desired time-dependent trajectory and then at minimizing the deviation in the sense of two possible space-norms.

7.3.7. Urban traffic control

Participants: C. Canudas-de-Wit [Contact person], F. Garin, P. Grandinetti.

In [20] we study near-optimal operation of traffic lights in an urban area, e.g., a town or a neighborhood. The goal is on-line optimization of traffic lights schedule in real time, so as to take into account variable traffic demands, with the objective of obtaining a better use of the road infrastructure. More precisely, we aim at maximizing total travel distance within the network, together with balancing densities across the network. The complexity of optimization over a large area is addressed both in the formulation of the optimization problem, with a suitable choice of the traffic model, and in a distributed solution, which not only parallelizes computations, but also respects the geometry of the town, i.e., it is suitable for an implementation in a smart infrastructure where each intersection can compute its optimal traffic lights by local computations combined with exchanges of information with neighbor intersections.

7.3.8. Modeling and control strategies for improving environmental sustainability of road transportation

Participants: B. Othman, G. de Nunzio [IFP Energies nouvelles], D. Di Domenico [IFP Energies nouvelles], C. Canudas-de-Wit [Contact person].

As road transportation energy use and environmental impact are globally rising at an alarming pace, authorities seek in research and technological advancement innovative solutions to increase road traffic sustainability. The unclear and partial correlation between road congestion and environmental impact is promoting new research directions in traffic management. We review the existing modeling approaches to accurately represent traffic behavior and the associated energy consumption and pollutant emissions [26]. The review then covers the transportation problems and control strategies that address directly environmental performance criteria, especially in urban networks. A discussion on the advantages of the different methods and on the future outlook for the eco-traffic management completes the proposed survey.

7.3.9. Data analysis for smart multi-modal transportation planning

Participants: A. Kibangou [Contact person], T. Moyo [Univ. of Johannesburg], W. Musakwa [Univ. of Johannesburg].

Modern cities have managed to balance the relationship between supply and demand of services through clear planning strategies which advocate smart solutions to the ever increasing demand for public tranportation services. The end goal is not to prohibit citizens to use their private cars, but to create an enabling smart system at a suitable scale which would lead to citizens not needing to own or drive a car. Having an efficiently and effectively run public transportation system is a crucial and indispensable factor for any developing city region. However as the provision of public transportation is a multifaceted process, with intertwining elements such as culture, politics,finance and shareholder interests, smart means of monitoring and mitigating the challenges faced in the provision of public transportation need to be developed continuously. The Gauteng city region is likewise faced with this challenge. With this region being the economic hub of South Africa, this has greatly affected the operation of the Gautrain system and the BRT systems within the region, as more and more people require a fast and reliable transportation means to move in and out the metropolitan cities. The study relied on a questionnaire-based survey that was administered to 60 respondents. The questionnaire had both closed and open-ended questions which were administered online through Google forms so as to obtain a good response rate from commuters who reside within the study area. The questions centred on identifying factors influencing the commuter's travelling patterns. Gautrain Management Agency reports and literature were also utilised to supplement infomation gleaned from the questionnaire. Besides the questionnaire, secondary data was collected from Twitter (tweets) concerning the Gaubus and Gautrain (between the period of August to November 2018). Posts from 380 users were analysed. This data was used to spatially identify POI of Gaubus users and also to identify the spatial relationship between land use activities, Gaubus routes, Gaubus stops, Gautrain stations and Guatrain routes. A neighboorhood analysis was run using a focal statistics based tool to map the spatial distribution of commuters of the Gaubus [56].

7.3.10. Location of turning ratio and flow sensors for flow reconstruction in large traffic networks

Participants: M. Rodriguez-Vega, C. Canudas-de-Wit [Contact person], H. Fourati.

We examine the problem of minimizing the number of sensors needed to completely recover the vehicular flow in a steady state traffic network [28]. We consider two possible sensor technologies: one that allows the measurement of turning ratios at a given intersection and the other that directly measures the flow in a road. We formulate an optimization problem that finds the optimal location of both types of sensors, such that a minimum number is required. To solve this problem, we propose a method that relies on the structure of the underlying graph, which has a quasi-linear computational complexity, resulting in less computing time when compared to other works in the literature. We evaluate our results using dynamical traffic simulations in synthetic networks.

7.4. Multisensor data fusion for navigation

7.4.1. Heterogeneity and uncertainty in distributed estimation from relative measurements

Participants: C. Ravazzi [Politecnico Torino], N.k. Chan [Univ. Groningen], P. Frasca [Contact person].

This work, presented in [27], has studied the problem of estimation from relative measurements in a graph, in which a vector indexed over the nodes has to be reconstructed from pairwise measurements of differences between its components associated to nodes connected by an edge. In order to model heterogeneity and uncertainty of the measurements, we assume them to be affected by additive noise distributed according to a Gaussian mixture. In this original setup, we formulate the problem of computing the Maximum-Likelihood (ML) estimates and we design two novel algorithms, based on Least Squares regression and Expectation-Maximization (EM). The first algorithm (LSEM) is centralized and performs the estimation from relative measurements, the soft classification of the measurements, and the estimation of the noise parameters. The second algorithm (Distributed LS-EM) is distributed and performs estimation and soft classification of the

measurements, but requires the knowledge of the noise parameters. We provide rigorous proofs of convergence for both algorithms and we present numerical experiments to evaluate their performance and compare it with solutions from the literature. The experiments show the robustness of the proposed methods against different kinds of noise and, for the Distributed LS-EM, against errors in the knowledge of noise parameters.

7.4.2. Cooperative localization and navigation: Theory, research, and practice

Participants: C. Gao [Naval Aviation University, China], G. Zhao [Naval Aviation University, China], H. Fourati [Contact person].

The idea of the book [58] comes as a response to the immense interest and strong activities in the field of cooperative localization and navigation during the past few years, both in theoretical and practical aspects. This book is targeted toward researchers, academics, engineers, and graduate students working in the field of sensor fusion, filtering, and signal processing for localization and navigation. This book, entitled Cooperative Localization and Navigation: Theory, Research and Practice, captures the latest results and techniques for cooperative navigation drawn from a broad array of disciplines. It is intended to provide the reader with a generic and comprehensive view of contemporary state estimation methodologies for localization and navigation, as well as the most recent researches and novel advances on cooperative localization and navigation task exploring the design of algorithms and architectures, benefits, and challenging aspects, as well as a potential broad array of disciplines, including wireless communication, in-door localization, robotics, and emergency rescue. These issues arise from the imperfection and diversity of cooperative sources, the contention and collision of communication channels, the selection and fusion of cooperative data, and the nature of the application environment. The issues that make cooperative-based navigational state estimation a challenging task, and which will be discussed through the different chapters of the book, are related to (1) the nature and model of sensors and cooperative sources (e.g., range-based sensor, angle-based sensor, inertialsensor, and vision sensor); (2) the communication medium and cooperative strategies; (3) the theoretical developments of state estimation and data fusion; and (4) the applicable platforms.

7.4.3. Data fusion from multi-inertial and magnetic sensors

- **Attitude estimation from multi-sensor observations**
  Participants: J. Wu [Hong Kong University of Science and Technology], Z. Zhou [University of Electronic Science and Technology of China], H. Fourati [Contact person], R. Li [University of Electronic Science and Technology of China], M. Liu [Hong Kong University of Science and Technology], A. Kibangou, A. Makni.

  Focusing on generalized sensor combinations, we deal with attitude estimation problem using a linear complementary filter [36]. The quaternion observation model is obtained via a gradient descent algorithm (GDA). An additive measurement model is then established according to derived results. The filter is named as the generalized complementary filter (GCF) where the observation model is simplified to its limit as a linear one that is quite different from previous-reported brute-force computation results. Moreover, we prove that representative derivative-based optimization algorithms are essentially equivalent to each other. Derivations are given to establish the state model based on the quaternion kinematic equation. The proposed algorithm is validated under several experimental conditions involving free-living environment, harsh external field disturbances and aerial flight test aided by robotic vision. Using the specially designed experimental devices, data acquisition and algorithm computations are performed to give comparisons on accuracy, robustness, time-consumption and etc. with representative methods. The results show that not only the proposed filter can give fast, accurate and stable estimates in terms of various sensor combinations, but it also produces robust attitude estimation in the presence of harsh situations e.g. irregular magnetic distortion. In other recent work, related to the attitude estimation, we add some corrections to update that version [35]. In [21], we propose the design of an attitude estimation algorithm for a rigid body subject to accelerated maneuvers. Unlike the current literature where the process model is usually driven by triaxial gyroscope measurements, we investigate a new formulation of the state-space model where the process model is given by triaxial accelerometer measurements. The observation

model is given by triaxial gyroscope and magnetometer measurements. The proposed model is written as a descriptor system and takes the external acceleration sensed by the accelerometer into account. Based on this model, a Quaternion Descriptor Filter (QDF) is developed and its performance is evaluated through simulations and experimental tests in pedestrian navigation.

- **Convexity analysis of optimization framework of attitude determination**
  Participants: J. Wu [Hong Kong University of Science and Technology], Z. Zhou [University of Electronic Science and Technology of China], H. Fourati [Contact person], M. Liu [Hong Kong University of Science and Technology].

  In the past several years, there have been several representative attitude determination methods developed using derivative-based optimization algorithms. Optimization techniques e.g. gradient-descent algorithm (GDA), Gauss-Newton algorithm (GNA), LevenbergMarquadt algorithm (LMA) suffer from local optimum in real engineering practices. A brief discussion on the convexity of this problem was presented recently, stating that the problem is neither convex nor concave. In our work, we give analytic proofs on this problem. The results reveal that the target loss function is convex in the common practice of quaternion normalization, which leads to non-existence of local optimum.

- **Behaviors classification based distance measuring system for pedestrians via a foot-mounted multi-inertial sensors**
  Participants: Z. Zhou [University of Electronic Science and Technology of China], S. Mo [University of Electronic Science and Technology of China], J. Wu [Hong Kong University of Science and Technology], H. Fourati [Contact person].

  We developed a foot-mounted pedestrian navigation system prototype with the emphasis on distance measuring with an inertial measurement unit (IMU) which implies the characteristics of pedestrian gait cycle and thus can be used as a crucial step indicator for distance calculation [37]. Conventional methods for step detection and step length estimation cannot adapt well to the general pedestrian applications since the parameters in these methods may vary for different persons and motions. In this paper, an adaptive time- and frequency-domains joint distance measuring method is proposed by utilizing the means of behaviors classification. Two key issues are studied: step detection and step length determination. For the step detection part, first behavior classification along with state transition strategy is designed to identify typical pedestrian behaviors including standing still, walking, running and irregular swing. Then a four-stage step detection method is proposed to adaptively determine both step frequency and threshold in a flexible window. Based on the behavior classification results, a two-segment functional based step length model is established to adapt the walking and running behaviors. Finally, real experiments are carried out to verify our proposed step detection method and step length model. The results show that the proposed method outperforms the existing representative methods and it exhibits the merits of accuracy and adaptability for different persons in real time and significantly improves the accuracy of distance measuring.

- **Human activities and postures recognition: from inertial measurements to quaternion-based approaches**
  Participants: M. Zmitri, H. Fourati [contact person], N. Vuillerme [AGEIS, UGA].

  We present two approaches to assess the effect of the number of inertial sensors and their location placements on recognition of human postures and activities [38]. Inertial and Magnetic Measurement Units (IMMUs)—which consist of a triad of three-axis accelerometer, three-axis gyroscope, and three-axis magnetometer sensors—are used in this work. Five IMMUs are initially used and attached to different body segments. Placements of up to three IMMUs are then considered: back, left foot, and left thigh. The subspace k-nearest neighbors (KNN) classifier is used to achieve the supervised learning process and the recognition task. In a first approach, we feed raw data from three-axis accelerometer and three-axis gyroscope into the classifier without any filtering or pre-processing, unlike what is usually reported in the state-of-the-art where statistical features were computed instead. Results show the efficiency of this method for the recognition of the studied activities and postures. With the proposed algorithm, more than 80% of the activities and postures are correctly

classified using one IMMU, placed on the lower back, left thigh, or left foot location, and more than 90% when combining all three placements. In a second approach, we extract attitude, in term of quaternion, from IMMUs in order to more precisely achieve the recognition process. The obtained accuracy results are compared to those obtained when only raw data is exploited. Results show that the use of attitude significantly improves the performance of the classifier, especially for certain specific activities. In that case, it was further shown that using a smaller number of features, with quaternion, in the recognition process leads to a lower computation time and better accuracy.

- **Improving inertial velocity estimation through magnetic field gradient-based extended kalman filter**
  Participants: M. Zmitri, H. Fourati [contact person], C. Prieur [GIPSA-Lab, UGA].

  We focused on the velocity estimation problem of a rigid body and how to improve it with magnetoinertial sensors-based theory [55]. We provide a continuous-time model that describes the motion of the body and we augment it after by introducing a new magnetic field gradient equation instead of using its value directly as an input for the model, as done usually in the corresponding literature. We investigate the advantage of moving to higher order spatial derivatives of the magnetic field in the estimation of velocity. These derivatives are computed thanks to a determined arrangement of magnetometers array. Within this framework, a specific set configuration of Extended Kalman Filters (EKFs) is proposed to focus mainly on the estimation of velocity and attitude of the body, but includes also an estimation of the magnetic field and its gradient. Some simulations for a specific scenario are proposed to show the improvements that we bring to the velocity estimation.

TRIPOP Project-Team

6. New Results

6.1. Nonlinear waves in granular chains

Participants: Guillaume James, Bernard Brogliato, Kirill Vorotnikov.

Granular chains made of aligned beads interacting by contact (e.g. Newton's cradle) are widely studied in the context of impact dynamics and acoustic metamaterials. In order to describe the response of such systems to impacts or vibrations, it is important to analyze different wave effects such as the propagation of compression waves (solitary waves or fronts) or localized oscillations (traveling breathers), or the scattering of vibrations through the chain. Such phenomena are strongly influenced by contact nonlinearities (Hertz force), spatial inhomogeneities and dissipation.

In the work [8], we analyze the Kuwabara-Kono (KK) model for contact damping, and we develop new approximations of this model which are efficient for the simulation of multiple impacts. The KK model is a simplified viscoelastic contact model derived from continuum mechanics, which allows for simpler calibration (using material parameters instead of phenomenological ones), but its numerical simulation requires a careful treatment due to its non-Lipschitz character. Using different dissipative time-discretizations of the conservative Hertz model, we show that numerical dissipation can be tuned properly in order to reproduce the physical dissipation of the KK model and associated wave effects. This result is obtained analytically in the limit of small time steps (using methods from backward analysis) and is numerically validated for larger time steps. The resulting schemes turn out to provide good approximations of impact propagation even for relatively large time steps.

In addition, G.J. has developed a theoretical method to analyze impacts in homogeneous granular chains with KK dissipation. The idea is to use the exponent α of the contact force as a parameter and derive simpler dynamical equations through an asymptotic analysis, in the limit when α approaches unity and long waves are considered. In that case, different continuum limits of the granular chain can be obtained. When the contact damping constant remains of order unity, wave profiles are well approximated by solutions of a viscous Burgers equation with logarithmic nonlinearity. For small contact damping, dispersive effects must be included and the continuum limit corresponds to a KdV-Burgers equation with logarithmic nonlinearity. By studying traveling wave solutions to these partial differential equations, we obtain analytical approximations of wave profiles such as compression fronts. We observe that these approximations remain meaningful for the classical exponent $\alpha = 3/2$. Indeed, they are close to exact wave profiles computed numerically for the KK model, using both dynamical simulations (response of the chain to a compression by a piston) and the Newton method (computation of exact traveling waves by a shooting method). In addition, in analogy with the Rankine-Hugoniot conditions for hyperbolic systems, we relate the asymptotic states of the KK model (for an infinite granular chain) to the velocity of a propagating front. These results are described in an article in preparation.

6.2. Signal propagation along excitable chains

Participant: Arnaud Tonnelier.

Nonlinear self-sustained waves, or *autowaves*, have been identified in a large class of discrete excitable media. We have proposed a simple continuous-time threshold model for wave propagation in excitable media. The ability of the resulting transmission line to convey a one-bit signal is investigated. Existence and multistability of signals where two successive units share the same waveform is established. We show that, depending on the connectivity of the transmission line, an arbitrary number of distinct signals can be transmitted. More precisely, we prove that, for a one-dimensional information channel with nth-neighbor interactions, a n-fold degeneracy of the speed curve induces the coexistence of $2n$ propagating signals, n of which are stable and allow n distinct symbols transmission. The influence of model parameters (time constants, coupling strength and connectivity) on the traveling signal properties is analyzed. This work is almost finished and is going to be submitted.

6.3. Hybrid Differential Algebraic equations

Participants: Vincent Acary, Bernard Brogliato, Alexandre Rocca.

In [18], [21], we study differential algebraic equations with constraints defined in a piecewise manner using a conditional statement. Such models classically appear in systems where constraints can evolve in a very small time frame compared to the observed time scale. The use of conditional statements or hybrid automata are a powerful way to describe such systems and are, in general, well suited to simulation with event driven numerical schemes. However, such methods are often subject to chattering at mode switch in presence of sliding modes, or can result in Zeno behaviours. In contrast, the representation of such systems using differential inclusions and method from non-smooth dynamics are often closer to the physical theory but may be harder to interpret. Associated time-stepping numerical methods have been extensively used in mechanical modelling with success and then extended to other fields such as electronics and system biology. In a similar manner to the previous application of non-smooth methods to the simulation of piecewise linear ODEs, non-smooth event-capturing numerical scheme are applied to piecewise linear DAEs. In particular, the study of a 2-D dynamical system of index-2 with a switching constraint using set-valued operators, is presented.

6.4. Numerical analysis of multibody mechanical systems with constraints

This scientific theme concerns the numerical analysis of mechanical systems with bilateral and unilateral constraints, with or without friction [1]. They form a particular class of dynamical systems whose simulation requires the development of specific methods for analysis and dedicated simulators [57].

6.4.1. Numerical solvers for frictional contact problems.

Participants: Vincent Acary, Maurice Brémond, Paul Armand.

In [34], we review several formulations of the discrete frictional contact problem that arises in space and time discretized mechanical systems with unilateral contact and three-dimensional Coulomb's friction. Most of these formulations are well–known concepts in the optimization community, or more generally, in the mathematical programming community. To cite a few, the discrete frictional contact problem can be formulated as variational inequalities, generalized or semi–smooth equations, second–order cone complementarity problems, or as optimization problems such as quadratic programming problems over second-order cones. Thanks to these multiple formulations, various numerical methods emerge naturally for solving the problem. We review the main numerical techniques that are well-known in the literature and we also propose new applications of methods such as the fixed point and extra-gradient methods with self-adaptive step rules for variational inequalities or the proximal point algorithm for generalized equations. All these numerical techniques are compared over a large set of test examples using performance profiles. One of the main conclusion is that there is no universal solver. Nevertheless, we are able to give some hints to choose a solver with respect to the main characteristics of the set of tests.

Recently, new developments have been carried out on two new applications of well-known numerical methods in Optimization:

- *Interior point methods* With the visit of Paul Armand, Université de Limoges, we co-supervise a M2 internship, Maksym Shpakovych on the application of interior point methods for quadratic problem with second-order cone constraints. The results are encouraging and a publication in computational mechanics is in progress.
- *Alternating Direction Method of Multipliers*. In collaboration with Yoshihiro Kanno, University of Tokyo, the use of the Alternating Direction Method of Multipliers (ADMM) has been adapted to the discrete frictional contact problems. With the help of some acceleration and restart techniques for first-order optimization methods and a residual balancing technique for adapting the proximal penalty parameter, the method proved to be efficient and robust on our test bench examples. A publication is also in preparation on this subject.

6.4.2. Modeling and numerical methods for frictional contact problems with rolling resistance

Participants: Vincent Acary, Franck Bourrier.

In [19], the Coulomb friction model is enriched to take into account the resistance to rolling, also known as rolling friction. Introducing the rolling friction cone, an extended Coulomb's cone and its dual, a formulation of the Coulomb friction with rolling resistance as a cone complementarity problem is shown to be equivalent to the standard formulation of the Coulomb friction with rolling resistance. Based on this complementarity formulation, the maximum dissipation principle and the bi-potential function are derived. Several iterative numerical methods based on projected fixed point iterations for variational inequalities and block-splitting techniques are given. The efficiency of these method strongly relies on the computation of the projection onto the rolling friction cone. In this article, an original closed-form formulae for the projection on the rolling friction cone is derived. The abilities of the model and the numerical methods are illustrated on the examples of a single sphere sliding and rolling on a plane, and of the evolution of spheres piles under gravity.

6.4.3. Finite element modeling of cable structures

Participants: Vincent Acary, Charlélie Bertrand.

Standard finite element discretization for cable structures suffer from several drawbacks. The first one is related to the mechanical assumption that the cable can not support compression. Standard formulations do not take into account this assumption. The second drawback comes from the high stiffness of the cable model when we deal with large lengths with high Young modulus such as cable ropeways installations. In this context, standard finite element applications cannot avoid compressive solutions and have huge difficulties to converge. In a forthcoming paper, we propose to a formulation based on a piecewise linear modeling of the cable constitutive behavior where the elasticity in compression is canceled. Furthermore, a dimensional analysis help us to formulate a problem that is well-balanced and the conditioning of the problem is diminished. The finite element discretization of this problem yields a robust method where convergence is observed with the number of elements and the nonlinear solver based on nonsmooth Newton strategy is converging up to tight tolerances. The convergence with the number of element allows one to refine the mesh as much as we want that will be of utmost importance for applications with contact and friction. Indeed, a fine discretization with respect to the whole length of the cable will be possible in the contact zone.

6.4.4. Well-posedness of the contact problem

We continue in [3] the analysis of the so-called contact problem for Lagrangian systems with bilateral and unilateral constraints, with set-valued Coulomb's friction. The problem that is analysed this time concerns sticking contacts (in both the normal and the tangential directions), *i.e.*, does there exist a solution (possibly unique) to the contact problem (that takes the form of a complementarity problem) when all contacts are sticking ? An algorithm is proposed that allows in principle to compute solutions. We rely strongly on results of existence and uniqueness of solutions to variational inequality of the second kind, obtained in the team some years ago. Let us note also the erratum/addendum of the monograph [45] in [17], which is regularly updated.

6.5. Analysis and Control of Set-Valued Systems

Participants: Bernard Brogliato, Christophe Prieur, Vincent Acary.

6.5.1. Robust sliding-mode control: continuous and discrete-time

The implicit method for the time-discretization of set-valued sliding-mode controllers was introduced in [29], [31]. The backstepping approach is used in [9] to design a continuous-time and a discrete-time nested set-valued controller that is able to reject unmatched disturbances (a problem that is known to be tough in the sliding-mdoe control community). In [13], [10] we continue the analysis of the implicit discretization of set-valued systems, this time oriented towards the consistency of time-discretizations for homogeneous systems, with one discontinuity at zero (sometimes called quasi-continuous, strangely enough). The discrete-time analysis of the twisting and the super-twisting algorithms are tackled in [7], [4].

6.5.2. Analysis of set-valued Lur'e dynamical systems

Lur'e systems are very popular in the Automatic Control field since their introduction by Lur'e in 1944. In [5] we propose a very complete survey/tutorial on the set-valued version of such dynamical systems (in finite dimension) which mainly consist of the negative feedback interconnection of an ODE with a maximal monotone set-valued operator. The first studies can be traced back to Yakubpovich in 1963 who analysed the stability of a linear time invariant system with positive real constraints, in negative feedback connection with a hysteresis operator. About 600 references are analysed from the point of view of the mathematical formalisms (Moreau's sweeping process, evolution variational inequalities, projected dynamical systems, complementarity dynamical systems, maximal monotone differential inclusions, differential variational inequalities), the relationships between these formalisms, the numerous fields of application, the well-posedness issues (existence, uniqueness and continuous dependence of solutions), and the stability issues (generalized equations for fixed points, Lyapunov stability, invariance principles).

6.5.3. Optimal control of LCS

The quadratic and minimum time optimal control of LCS as in (6) is tackled in [14], [12]. This work relies on the seminal results by Guo and ye (SIAM 2016), and aims at particularizing their results for LCS, so that they become numerically tractable and one can compute optimal controllers and optimal trajectories. The basic idea is to take advantage of the complementarity, to construct linear complementarity problems in the Pontryagin's necessary conditions which can then be integrated numerically, without having to guess a priori the switching instants (the optimal controller can eb discontinuous and the optimal trajectories can visit several modes of the complementarity conditions).

6.6. Dissipative systems

Participant: Bernard Brogliato.

The third edition of the book Dissipative Systems Analysis and Control has been released https://www.springer.com/gp/book/9783030194192. Also a short proof of equivalence of side conditions for strictly positive real (SPR) transfer functions is done in [6], closing a long debate in the Automatic Control community about the characterization of SPR transfer matrices.

AIRSEA Project-Team

6. New Results

6.1. Modeling for Oceanic and Atmospheric flows

6.1.1. Numerical Schemes for Ocean Modelling

Participants: Eric Blayo, Matthieu Brachet, Laurent Debreu, Emilie Duval, Christopher Eldred, Nicholas Kevlahan, Florian Lemarié, Gurvan Madec, Farshid Nazari.

The increase of model resolution naturally leads to the representation of a wider energy spectrum. As a result, in recent years, the understanding of oceanic submesoscale dynamics has significantly improved. However, dissipation in submesoscale models remains dominated by numerical constraints rather than physical ones. Effective resolution is limited by the numerical dissipation range, which is a function of the model numerical filters (assuming that dispersive numerical modes are efficiently removed). As an example, the stabilization of the coupling between barotropic (fast) and baroclinic (slow) modes in a three dimensional ocean model is a source of large numerical dissipation. This has been studied in details in [6].
F. Lemarié and L. Debreu (with H. Burchard, K. Klingbeil and J. Sainte-Marie) have organized the international COMMODORE workshop on numerical methods for oceanic models (Paris, Sept. 17-19, 2018). https://commodore2018.sciencesconf.org/, see [12] for a summary of the scientific discussions. The next COMMODORE meeting is planned for February 2020 and will take place in Hamburg. https://www.conferences.uni-hamburg.de/event/76.

With the increase of resolution, the hydrostatic assumption becomes less valid and the AIRSEA group also works on the development of non-hydrostatic ocean models. The treatment of non-hydrostatic incompressible flows leads to a 3D elliptic system for pressure that can be ill conditioned in particular with non geopotential vertical coordinates. That is why we favour the use of the non-hydrostatic compressible equations that removes the need for a 3D resolution at the price of reincluding acoustic waves [29].
In addition, Emilie Duval started her PhD in September 2018 on the coupling between the hydrostatic incompressible and non-hydrostatic compressible equations.

The team is involved in the HEAT (Highly Efficient ATmospheric Modelling) ANR project. This project aims at developing a new atmospheric dynamical core (DYNAMICO) discretized on an icosahedral grid. This project is in collaboration with Ecole Polytechnique, Meteo-France, LMD, LSCE and CERFACS. In the context of the HEAT project, we worked on the analysis of dispersion analysis of continuous and discontinuous Galerkin methods of arbitrary degree of approximation ([31]), on compatible Galerkin schemes for shallow water model in 2D ([9]). In addition, we worked on the discrete formulation of the thermal rotating shallow water equations. This formulation, based on quasi-Hamiltonian discretizations methods, allows for the first time total mass, buoyancy and energy conservation to machine precision ([8]).

Accurate and stable implementation of bathymetry boundary conditions in ocean models remains a challenging problem. The dynamics of ocean flow often depend sensitively on satisfying bathymetry boundary conditions and correctly representing their complex geometry. Generalized (e.g.) terrain-following coordinates are often used in ocean models, but they require smoothing the bathymetry to reduce pressure gradient errors. Geopotential -coordinates are a common alternative that avoid pressure gradient and numerical diapycnal diffusion errors, but they generate spurious flow due to their “staircase” geometry. In [5], we introduce a new Brinkman volume penalization to approximate the no-slip boundary condition and complex geometry of bathymetry in ocean models. This approach corrects the staircase effect of -coordinates, does not introduce any new stability constraints on the geometry of the bathymetry and is easy to implement in an existing ocean model. The porosity parameter allows modelling subgrid scale details of the geometry. We illustrate the penalization and confirm its accuracy by applying it to three standard test flows: upwelling over a sloping bottom, resting state over a seamount and internal tides over highly peaked bathymetry features.

Figure (1) shows strong improvements obtained when the penalization method is used in comparison with traditional terrain following σ simulations. At 6km resolution, the penalization methods (Figure (1) d)), that takes into account details of bathymetry, allows to recover internal tide wave beams closed to the 3km simulation. (Figure (1) a)).

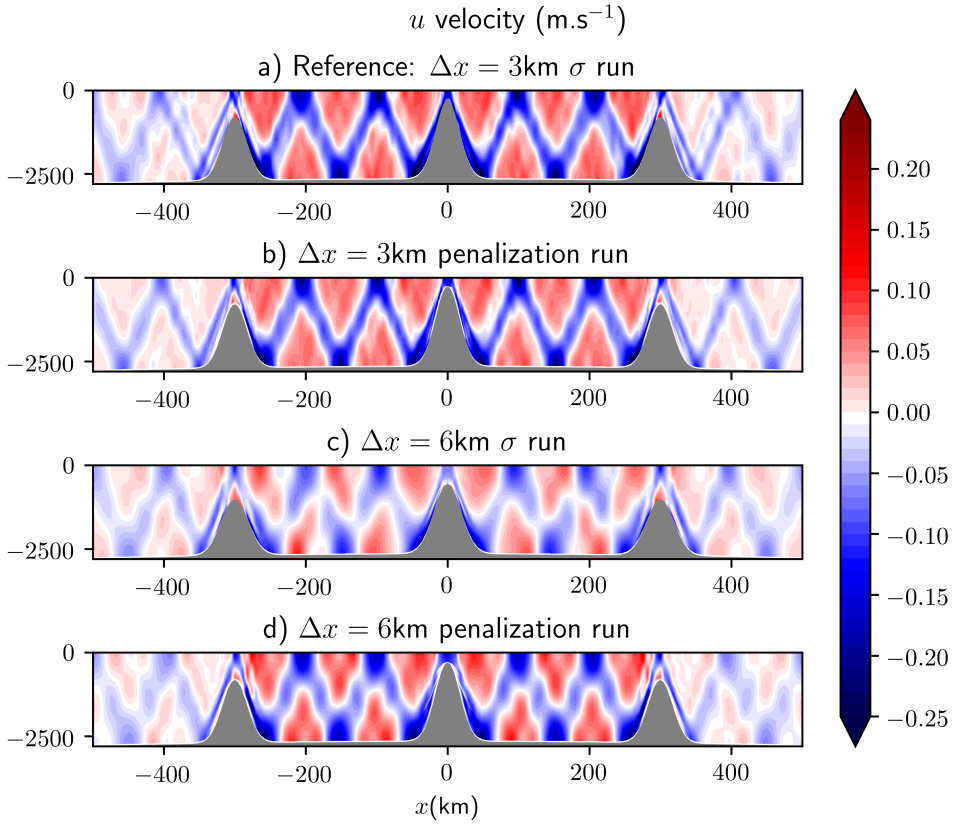


Figure 1. u velocity. Instantaneous solutions of the internal tide test case after 12 M2 tidal cycles of integration. (a) The reference σ coordinate run at 3 km resolution. (b) The penalized run at 3 km resolution. (c) The σ-coordinate run at 3 km resolution. (d) The penalized run at 6 km resolution.

6.1.2. Coupling Methods for Oceanic and Atmospheric Models

Participants: Eric Blayo, Florian Lemarié, Sophie Thery, Simon Clément.

Coupling methods routinely used in regional and global climate models do not provide the exact solution to the ocean-atmosphere problem, but an approximation of one [49]. For the last few years we have been actively working on the analysis of ocean-atmosphere coupling both in terms of its continuous and numerical formulation (see [21] for an overview). Our activities can be divided into four general topics

1. *Stability and consistency analysis of existing coupling methods*: in [49] we showed that the usual methods used in the context of ocean-atmosphere coupling are prone to splitting errors because they correspond to only one iteration of an iterative process without reaching convergence. Moreover, those methods have an additional condition for the coupling to be stable even if unconditionally stable time-stepping algorithms are used. This last remark was further studied in [37] and it turned

out to be a major source of instability in atmosphere-snow coupling.

2. *Study of physics-dynamics coupling*: during the PhD-thesis of Charles Pelletier the scope was on including the formulation of physical parameterizations in the theoretical analysis of the coupling, in particular the parameterization schemes to compute air-sea fluxes [56]. To do so, a metamodel representative of the behavior of the full parameterization but with a continuous form easier to manipulate has been derived thanks to a sensitivity analysis. This metamodel is more adequate to conduct the mathematical analysis of the coupling while being physically satisfactory [57]. More recently we have started to work specifically on the discretization methods for the parameterization of planetary boundary layers in climate models [51] which takes the form of an nonstationary nonlinear parabolic equation. The objective is to derive a discretization for which we could prove nonlinear stability criteria and show robustness to large variations in parabolic Courant number while being consistent with our knowledge of the underlying physical principles (e.g. the Monin-Obukhov theory in the surface layer).This work will continue in the framework of the PhD-thesis of C. Simon.
3. *A simplified atmospheric boundary layer model for oceanic purposes*: Part of our activities within the IMMERSE project is dedicated to the development of a simplified model of the marine atmospheric boundary layer of intermediate complexity between a bulk parameterization and a full three-dimensional atmospheric model and to its integration to the NEMO general circulation model [24]. A constraint in the conception of such a simplified model is to allow an apt representation of the downward momentum mixing mechanism and partial re-energization of the ocean by the atmosphere while keeping the computational efficiency and flexibility inherent to ocean only modeling. A paper is in preparation and will be submitted in early 2020.
4. *Analysis of air-sea-wave interactions in realistic high-resolution realistic simulations*: part of our activity has been in collaboration with atmosphericists and physical oceanographers to study the impact on some modeling assumptions (e.g. [50]) in large-scale realistic ocean-atmosphere coupled simulations [16], [53]. Moreover, within the ALBATROS project [23], we have contributed to the development of a 2-way coupling between an ocean global circulation model (NEMO) with a surface wave model (WW3). Such coupling is not straightforward to implement since it requires modifications of the governing equations, boundary conditions and subgrid scale closures in the oceanic model. A paper is currently under open discussion in Geoscientific Model Development journal on that topic [4].
5. *Efficient coupling methods*: we have been developing coupling approaches for several years, based on so-called Schwarz algorithms. In particular, we addressed the development of efficient interface conditions for multi-physics problems representative of air-sea coupling [61] (paper in preparation). This work is done in the framework of S. Théry PhD (started in fall 2017). During the internship of C. Simon, efficient interface conditions have been derived at a (semi)-discrete level and can thus be systematically compared with the ones obtained from the continuous problem.

These topics are addressed through strong collaborations between the applied mathematicians and the climate community (Meteo-France, Ifremer, LMD, and LOCEAN). Our work on ocean-atmosphere coupling has steadily matured over the last few years and has reached a point where it triggered interest from the climate community. Through the funding of the COCOA ANR project (started in January 2017, PI: E. Blayo), Airsea team members play a major role in the structuration of a multi-disciplinary scientific community working on ocean-atmosphere coupling spanning a broad range from mathematical theory to practical implementations in climate models. An expected outcome of this project should be the design of a benchmark suite of idealized coupled test cases representative of known issues in coupled models. Such idealized test cases should motivate further collaborations at an international level. In this context, a single-column version of the CNRM climate models has been designed and several coupling algorithms have been implemented (work done by S. Valcke, CERFACS). This model will be used to illustrate the relevance of our theoretical work in a semi-realistic context.

6.1.3. Data assimilation for coupled models

In the context of operational meteorology and oceanography, forecast skills heavily rely on proper combination of model prediction and available observations via data assimilation techniques. Historically, numerical weather prediction is made separately for the ocean and the atmosphere in an uncoupled way. However, in recent years, fully coupled ocean-atmosphere models are increasingly used in operational centers to improve the reliability of seasonal forecasts and tropical cyclones predictions. For coupled problems, the use of separated data assimilation schemes in each medium is not satisfactory since the result of such assimilation process is generally inconsistent across the interface, thus leading to unacceptable artefacts. Hence, there is a strong need for adapting existing data assimilation techniques to the coupled framework. As part of our ERACLIM2 contribution three general data assimilation algorithms, based on variational data assimilation techniques, have been developed and applied to a single column coupled model. The dynamical equations of the considered problem are coupled using an iterative Schwarz domain decomposition method. The aim is to properly take into account the coupling in the assimilation process in order to obtain a coupled solution close to the observations while satisfying the physical conditions across the air-sea interface. Results shows significant improvement compared to the usual approach on this simple system. The aforementioned system has been coded within the OOPS framework (Object Oriented Prediction System) in order to ease the transfer to more complex/realistic models.

Following this line of research, CASIS, a collaborative project with Mercator Océan started late 2017 until end of 2019 in order to extend these developments to iterative Kalman smoother data assimilation scheme, in the framework of a coupled ocean-atmospheric boundary layer context.

6.1.4. Optimal control of grids and schemes for ocean model.

Participants: Laurent Debreu, Eugene Kazantsev.

In [28], variational data assimilation technique is applied to a simple bidimensional wave equation that simulates propagation of internal gravity waves in the ocean in order to control grids and numerical schemes. Grid steps of the vertical grid, Brunt-Vaisala frequency and approximation of the horizontal derivative were used as control parameters either separately or in the joint control. Obtained results show that optimized parameters may partially compensate errors committed by numerical scheme due to insufficient grid resolution.

Optimal vertical grid steps and coefficients in horizontal derivative approximation found in the variational control procedure allow us to get the model solution that is rather close to the solution of the reference model. The error in the wave velocity on the coarse grid is mostly compensated in experiments with joint control of parameters while the error in the wave amplitude occurs to be more difficult to correct.

However, optimal grid steps and discretization schemes may be in a disagreement with requirements of other model physics and additional analysis of obtained optimized parameters from the point of view of they agreement with the model is necessary.

6.1.5. Machine learning for parametrisation of the model dissipation.

Participants: Laurent Debreu, Eugene Kazantsev, Arthur Vidard, Olivier Zahm.

Artificial intelligence and machine learning may be considered as a potential way to address unresolved model scales and to approximate poorly known processes such as dissipation that occurs essentially at small scales. In order to understand the possibility to combine numerical model and neural network learned with the aid of external data, we develop a network generation and learning algorithm and use it to approximate nonlinear model operators. Beginning with a simple nonlinear equations like transport-diffusion and Burgers ones, we use artificially generated external data to learn the network by Adam algorithm [47]. Results show the possibility to approximate nonlinear, and even discontinuous dissipation operator with a quite good accuracy, however, several millions iterations are necessary to learn.

6.1.6. Nonhydrostatic Modeling

Participants: Eric Blayo, Laurent Debreu, Emilie Duval.

In the context of the French initiative CROCO (Coastal and Regional Ocean COmmunity model, https://www.croco-ocean.org) for the development of a new oceanic modeling system, Emilie Duval is working on the design of methods to couple local nonhydrostatic models to larger scale hydrostatic ones (PhD started in Oct. 2018). Such a coupling is quite delicate from a mathematical point of view, due to the different nature of hydrostatic and nonhydrostatic equations (where the vertical velocity is either a diagnostic or a prognostic variable). A prototype has been implemented, which allows for analytical solutions in simplified configurations and will make it possible to test different numerical approaches.

6.2. Assimilation of spatially dense observations

Participants: Elise Arnaud, François-Xavier Le Dimet, Arthur Vidard, Long Li, Emilie Rouzies.

6.2.1. Direct assimilation of image sequences

At the present time the observation of Earth from space is done by more than thirty satellites. These platforms provide two kinds of observational information:

- Eulerian information as radiance measurements: the radiative properties of the earth and its fluid envelops. These data can be plugged into numerical models by solving some inverse problems.
- Lagrangian information: the movement of fronts and vortices give information on the dynamics of the fluid. Presently this information is scarcely used in meteorology by following small cumulus clouds and using them as Lagrangian tracers, but the selection of these clouds must be done by hand and the altitude of the selected clouds must be known. This is done by using the temperature of the top of the cloud.

Our current developments are targeted at the use of « Level Sets » methods to describe the evolution of the images. The advantage of this approach is that it permits, thanks to the level sets function, to consider the images as a state variable of the problem. We have derived an Optimality System including the level sets of the images. This approach is being applied to the tracking of oceanic oil spills in the framework of a Long Li's Phd in co-supervision with Jianwei Ma.

6.2.2. Observation error representation

Accounting for realistic observations errors is a known bottleneck in data assimilation, because dealing with error correlations is complex. Following a previous study on this subject, we propose to use multiscale modelling, more precisely wavelet transform, to address this question. In [3] we investigate the problem further by addressing two issues arising in real-life data assimilation: how to deal with partially missing data (e.g., concealed by an obstacle between the sensor and the observed system); how to solve convergence issues associated to complex observation error covariance matrices? Two adjustments relying on wavelets modelling are proposed to deal with those, and offer significant improvements. The first one consists in adjusting the variance coefficients in the frequency domain to account for masked information. The second one consists in a gradual assimilation of frequencies. Both of these fully rely on the multiscale properties associated with wavelet covariance modelling.

A collaborative project started with C. Lauvernet (IRSTEA) in order to make use of this kind of assimilation strategies on the control of pesticide transfer and it led to the co supervision of E. Rouzies PhD, starting in Dec 2019.

6.2.3. Optimal transport for image assimilation

We investigate the use of optimal transport based distances for data assimilation, and in particular for assimilating dense data such as images. The PhD thesis of N. Feyeux studied the impact of using the Wasserstein distance in place of the classical Euclidean distance (pixel to pixel comparison). In a simplified one dimensional framework, we showed that the Wasserstein distance is indeed promising. Data assimilation experiments with the Shallow Water model have been performed and confirm the interest of the Wasserstein distance. This has been extended to water pollutant tracking as part of Long Li's PhD and published in [13]

6.3. Model reduction / multiscale algorithms

6.3.1. Parameter space dimension reduction and Model order reduction

Participants: Mohamed Reda El Amri, Arthur Macherey, Youssef Marzouk, Clémentine Prieur, Alessio Spantini, Ricardo Baptista, Daniele Bigoni, Olivier Zahm.

Numerical models describing the evolution of the system (ocean + atmosphere) contain a large number of parameters which are generally poorly known. The reliability of the numerical simulations strongly depends on the identification and calibration of these parameters from observed data. In this context, it seems important to understand the kinds of low-dimensional structure that may be present in geophysical models and to exploit this low-dimensional structure with appropriate algorithms. We focus in the team, on parameter space dimension reduction techniques, low-rank structures and transport maps techniques for probability measure approximation.

In [25], we proposed a framework for the greedy approximation of high-dimensional Bayesian inference problems, through the composition of multiple low-dimensional transport maps or flows. Our framework operates recursively on a sequence of "residual" distributions, given by pulling back the posterior through the previously computed transport maps. The action of each map is confined to a low-dimensional subspace that we identify by minimizing an error bound. At each step, our approach thus identifies (i) a relevant subspace of the residual distribution, and (ii) a low-dimensional transformation between a restriction of the residual onto this sub-space and a standard Gaussian. We prove weak convergence of the approach to the posterior distribution, and we demonstrate the algorithm on a range of challenging inference problems in differential equations and spatial statistics.

The paper [34] introduces a novel error estimator for the Proper Generalized Decomposition (PGD) approximation of parametrized equations. The estimator is intrinsically random: It builds on concentration inequalities of Gaussian maps and an adjoint problem with random right-hand side, which we approximate using the PGD. The effectivity of this randomized error estimator can be arbitrarily close to unity with high probability, allowing the estimation of the error with respect to any user-defined norm as well as the error in some quantity of interest. The performance of the error estimator is demonstrated and compared with some existing error estimators for the PGD for a parametrized time-harmonic elastodynamics problem and the parametrized equations of linear elasticity with a high-dimensional parameter space.

In the framework of Arthur Macherey's PhD, we have also proposed in [26] algorithms for solving high-dimensional Partial Differential Equations (PDEs) that combine a probabilistic interpretation of PDEs, through Feynman-Kac representation, with sparse interpolation. Monte-Carlo methods and time-integration schemes are used to estimate pointwise evaluations of the solution of a PDE. We use a sequential control variates algorithm, where control variates are constructed based on successive approximations of the solution of the PDE. We are now interested in solving parametrized PDE with stochastic algorithms in the framework of potentially high dimensional parameter space. In [36], we consider gradient-based dimension reduction of vector-valued functions. Multivariate functions encountered in high-dimensional uncertainty quantification problems often vary most strongly along a few dominant directions in the input parameter space. In this work, we propose a gradient-based method for detecting these directions and using them to construct ridge approximations of such functions, in the case where the functions are vector-valued. The methodology consists of minimizing an upper bound on the approximation error, obtained by subspace Poincaré inequalities. We have provided a thorough mathematical analysis in the case where the parameter space is equipped with a Gaussian probability measure. We are now working on the nonlinear generalization of active subspaces. Reduced models are also developed In the framework of robust inversion. In [43], we have combined a new greedy algorithm for functional quantization with a Stepwise Uncertainty Reduction strategy to solve a robust inversion problem under functional uncertainties. An ongoing work aims at further reducing the number of simulations required to solve the same robust inversion problem, based on Gaussian process meta-modeling on the joint input space of deterministic control parameters and functional uncertain variable. These results are applied to automotive depollution. This research axis is conducted in the framework of the Chair OQUAIDO.

6.4. Sensitivity analysis

Participants: Elise Arnaud, Eric Blayo, Laurent Gilquin, Maria Belén Heredia, Adrien Hirvoas, François-Xavier Le Dimet, Henri Mermoz Kouye, Clémentine Prieur, Laurence Viry.

6.4.1. Scientific context

Forecasting geophysical systems require complex models, which sometimes need to be coupled, and which make use of data assimilation. The objective of this project is, for a given output of such a system, to identify the most influential parameters, and to evaluate the effect of uncertainty in input parameters on model output. Existing stochastic tools are not well suited for high dimension problems (in particular time-dependent problems), while deterministic tools are fully applicable but only provide limited information. So the challenge is to gather expertise on one hand on numerical approximation and control of Partial Differential Equations, and on the other hand on stochastic methods for sensitivity analysis, in order to develop and design innovative stochastic solutions to study high dimension models and to propose new hybrid approaches combining the stochastic and deterministic methods.

6.4.2. Global sensitivity analysis

Participants: Elise Arnaud, Eric Blayo, Laurent Gilquin, Maria Belén Heredia, Adrien Hirvoas, Alexandre Janon, Henri Mermoz Kouye, Clémentine Prieur, Laurence Viry.

6.4.2.1. Global sensitivity analysis with dependent inputs

An important challenge for stochastic sensitivity analysis is to develop methodologies which work for dependent inputs. Recently, the Shapley value, from econometrics, was proposed as an alternative to quantify the importance of random input variables to a function. Owen [54] derived Shapley value importance for independent inputs and showed that it is bracketed between two different Sobol' indices. Song et al. [60] recently advocated the use of Shapley value for the case of dependent inputs. In a recent work [55], in collaboration with Art Owen (Standford's University), we show that Shapley value removes the conceptual problems of functional ANOVA for dependent inputs. We do this with some simple examples where Shapley value leads to intuitively reasonable nearly closed form values. We also investigated further the properties of Shapley effects in [30].

6.4.2.2. Extensions of the replication method for the estimation of Sobol' indices

Sensitivity analysis studies how the uncertainty on an output of a mathematical model can be attributed to sources of uncertainty among the inputs. Global sensitivity analysis of complex and expensive mathematical models is a common practice to identify influent inputs and detect the potential interactions between them. Among the large number of available approaches, the variance-based method introduced by Sobol' allows to calculate sensitivity indices called Sobol' indices. Each index gives an estimation of the influence of an individual input or a group of inputs. These indices give an estimation of how the output uncertainty can be apportioned to the uncertainty in the inputs. One can distinguish first-order indices that estimate the main effect from each input or group of inputs from higher-order indices that estimate the corresponding order of interactions between inputs. This estimation procedure requires a significant number of model runs, number that has a polynomial growth rate with respect to the input space dimension. This cost can be prohibitive for time consuming models and only a few number of runs is not enough to retrieve accurate informations about the model inputs.

The use of replicated designs to estimate first-order Sobol' indices has the major advantage of reducing drastically the estimation cost as the number of runs n becomes independent of the input space dimension. The generalization to closed second-order Sobol' indices relies on the replication of randomized orthogonal arrays. However, if the input space is not properly explored, that is if n is too small, the Sobol' indices estimates may not be accurate enough. Gaining in efficiency and assessing the estimate precision still remains an issue, all the more important when one is dealing with limited computational budget.

We designed an approach to render the replication method iterative, enabling the required number of evaluations to be controlled. With this approach, more accurate Sobol' estimates are obtained while recycling previous sets of model evaluations. Its main characteristic is to rely on iterative construction of stratified designs, latin hypercubes and orthogonal arrays [45]

In [11] a new strategy to estimate the full set of first-order and second-order Sobol' indices with only two replicated designs based on orthogonal arrays of strength two. Such a procedure increases the precision of the estimation for a given computation budget. A bootstrap procedure for producing confidence intervals, that are compared to asymptotic ones in the case of first-order indices, is also proposed.

The replicated designs strategy for global sensitivity analysis was also implemented in the applied framework of marine biogeochemical modeling, making use of distributed computing environments [15]. It has allowed to perform a global sensitivity analysis with input space dimension more than eighty, without any screening preliminary step.

6.4.2.3. Green sensitivity for multivariate and functional outputs

Participants: María Belén Heredia, Clémentine Prieur.

Another research direction for global SA algorithm starts with the report that most of the algorithms to compute sensitivity measures require special sampling schemes or additional model evaluations so that available data from previous model runs (e.g., from an uncertainty analysis based on Latin Hypercube Sampling) cannot be reused. One challenging task for estimating global sensitivity measures consists in recycling an available finite set of input/output data. Green sensitivity, by recycling, avoids wasting. These given data have been discussed, e.g., in [59], [58]. Most of the given data procedures depend on parameters (number of bins, truncation argument...) not easy to calibrate with a bias-variance compromise perspective. Adaptive selection of these parameters remains a challenging issue for most of these given-data algorithms. In the context of María Belén Heredia's PhD thesis, we have proposed a non-parametric given data estimator for agregated Sobol' indices, introduced in [48] and further developed in [44] for multivariate or functional outputs. This last work should be submitted soon.

6.4.2.4. Global sensitivity analysis for parametrized stochastic differential equations

Participants: Henri Mermoz Kouye, Clémentine Prieur.

Many models are stochastic in nature, and some of them may be driven by parametrized stochastic differential equations. It is important for applications to propose a strategy to perform global sensitivity analysis (GSA) for such models, in presence of uncertainties on the parameters. In collaboration with Pierre Etoré (DATA department in Grenoble), Clémentine Prieur proposed an approach based on Feynman-Kac formulas [10]. The research on GSA for stochastic simulators is still ongoing, first in the context of the MATH-AmSud project FANTASTIC (Statistical inFerence and sensitivity ANalysis for models described by sTochASTIC differential equations) with Chile and Uruguay, secondly through the PhD thesis of Henri Mermoz Kouye, co-supervised by Clémentine Prieur, in collaboration with INRA Jouy.

6.5. Model calibration and statistical inference

6.5.1. Bayesian calibration

Participants: Maria Belén Heredia, Adrien Hirvoas, Clémentine Prieur.

Physically-based avalanche propagation models must still be locally calibrated to provide robust predictions, e.g. in long-term forecasting and subsequent risk assessment. Friction parameters cannot be measured directly and need to be estimated from observations. Rich and diverse data is now increasingly available from test-sites, but for measurements made along ow propagation, potential autocorrelation should be explicitly accounted for. In the context of María Belén Heredia's PhD, in collaboration with IRSTEA Grenoble, we have proposed in a comprehensive Bayesian calibration and statistical model selection framework with application to an avalanche sliding block model with the standard Voellmy friction law and high rate photogrammetric images. An avalanche released at the Lautaret test-site and a synthetic data set based on the avalanche were used to test the approach. Results have demonstrated i) the effciency of the proposed calibration scheme,

and ii) that including autocorrelation in the statistical modelling definitely improves the accuracy of both parameter estimation and velocity predictions. In the context of the energy transition, wind power generation is developing rapidly in France and worldwide. Research and innovation on wind resource characterisation, turbin control, coupled mechanical modelling of wind systems or technological development of offshore wind turbines floaters are current research topics. In particular, the monitoring and the maintenance of wind turbine is becoming a major issue. Current solutions do not take full advantage of the large amount of data provided by sensors placed on modern wind turbines in production. These data could be advantageously used in order to refine the predictions of production, the life of the structure, the control strategies and the planning of maintenance. In this context, it is interesting to optimally combine production data and numerical models in order to obtain highly reliable models of wind turbines. This process is of interest to many industrial and academic groups and is known in many fields of the industry, including the wind industry, as "digital twin". The objective of Adrien Hirvoas's PhD work is to develop of data assimilation methodology to build the "digital twin" of an onshore wind turbine. Based on measurements, the data assimilation should allow to reduce the uncertainties of the physical parameters of the numerical model developed during the design phase to obtain a highly reliable model. Various ensemble data assimilation approches are currently under consideration to address the problem. In the context of this work, it is necessary to develop algorithms of identification quantifying and ranking all the uncertainty sources. This work in done in collaboration with IFPEN.

6.5.2. Non-Parametric statistical inference for Kinetic Diffusions

Participants: Clémentine Prieur, Jose Raphael Leon Ramos.

This research is the subject of a collaboration with Chile and Uruguay. More precisely, we started working with Venezuela. Due to the crisis in Venezuela, our main collaborator on that topic moved to Uruguay.

We are focusing our attention on models derived from the linear Fokker-Planck equation. From a probabilistic viewpoint, these models have received particular attention in recent years, since they are a basic example for hypercoercivity. In fact, even though completely degenerated, these models are hypoelliptic and still verify some properties of coercivity, in a broad sense of the word. Such models often appear in the fields of mechanics, finance and even biology. For such models we believe it appropriate to build statistical non-parametric estimation tools. Initial results have been obtained for the estimation of invariant density, in conditions guaranteeing its existence and unicity [39] and when only partial observational data are available. A paper on the non parametric estimation of the drift has been accepted recently [40] (see Samson et al., 2012, for results for parametric models). As far as the estimation of the diffusion term is concerned, a paper has been accepted [40], in collaboration with J.R. Leon (Montevideo, Uruguay) and P. Cattiaux (Toulouse). Recursive estimators have been also proposed by the same authors in [41], also recently accepted. In a recent collaboration with Adeline Samson from the statistics department in the Lab, we considered adaptive estimation, that is we proposed a data-driven procedure for the choice of the bandwidth parameters.

In [42], we focused on damping Hamiltonian systems under the so-called fluctuation-dissipation condition. Idea in that paper were re-used with applications to neuroscience in [52].

Note that Professor Jose R. Leon (Caracas, Venezuela, Montevideo, Uruguay) was funded by an international Inria Chair, allowing to collaborate further on parameter estimation.

We recently proposed a paper on the use of the Euler scheme for inference purposes, considering reflected diffusions. This paper could be extended to the hypoelliptic framework.

We also have a collaboration with Karine Bertin (Valparaiso, Chile), Nicolas Klutchnikoff (Université Rennes) and Jose R. León (Montevideo, Uruguay) funded by a MATHAMSUD project (2016-2017) and by the LIA/CNRS (2018). We are interested in new adaptive estimators for invariant densities on bounded domains [38], and would like to extend that results to hypo-elliptic diffusions.

6.6. Modeling and inference for extremes

Participants: Philomène Le Gall, Clémentine Prieur, Patricia Tencaliec.

In [19], we are considering the modeling of precipitation amount with semi-parametric models, modeling both the bulk of the distribution and the tails, but avoiding the arbitrary choice of a threshold. We work in collaboration with Anne-Catherine Favre (LGGE-Lab in Grenoble) and Philippe Naveau (LSCE, Paris).

In the context of Philomène Le Gall's PhD thesis, we are applying the aforementioned modeling of extreme precipitation with the aim of regionalizing extreme precipitation.

6.7. Land Use and Transport Models Calibration

Participants: Clémentine Prieur, Arthur Vidard, Peter Sturm, Elise Arnaud.

Given the complexity of modern urban areas, designing sustainable policies calls for more than sheer expert knowledge. This is especially true of transport or land use policies, because of the strong interplay between the land use and the transportation systems. Land use and transport integrated (LUTI) modelling offers invaluable analysis tools for planners working on transportation and urban projects. Yet, very few local authorities in charge of planning make use of these strategic models. The explanation lies first in the difficulty to calibrate these models, second in the lack of confidence in their results, which itself stems from the absence of any well-defined validation procedure. Our expertise in such matters will probably be valuable for improving the reliability of these models. To that purpose we participated to the building up of the ANR project CITiES led by the STEEP EPI. This project started early 2013 and two PhD about sensitivity analysis and calibration were launched late 2013. Laurent Gilquin defended his PhD in October 2016 [46] and Thomas Capelle defended his in April 2017 and published his latest results in [2].

BEAGLE Project-Team

7. New Results

7.1. Computational Glioscience: A book to review the existing mathematical models of the glial cells

[participant: H. Berry]

Over the last two decades, the recognition that astrocytes - the predominant type of cortical glial cells - could sense neighboring neuronal activity and release neuroactive agents, has been instrumental in the uncovering of many roles that these cells could play in brain processing and the storage of information. These findings initiated a conceptual revolution that leads to rethinking how brain communication works since they imply that information travels and is processed not just in the neuronal circuitry but in an expanded neuron-glial network. On the other hand the physiological need for astrocyte signaling in brain information processing and the modes of action of these cells in computational tasks remain largely undefined. This is due, to a large extent, both to the lack of conclusive experimental evidence, and to a substantial lack of a theoretical framework to address modeling and characterization of the many possible astrocyte functions. This book [https://hal.inria.fr/hal-01995842] aims at filling this gap, providing the first systematic computational approach to the complex, wide subject of neuron-glia interactions. The organization of the book is unique insofar as it considers a selection of "hot topics" in glia research that ideally brings together both the novelty of the recent experimental findings in the field and the modelling challenge that they bear. A chapter written by experimentalists, possibly in collaboration with theoreticians, will introduce each topic. The aim of this chapter, that we foresee less technical in its style than in conventional reviews, will be to provide a review as clear as possible, of what is "established" and what remains speculative (i.e. the open questions). Each topic will then be presented in its possible different aspects, by 2-3 chapters by theoreticians. These chapters will be edited in order to provide a "priming" reference for modeling neuron-glia interactions, suitable both for the graduate student and the professional researcher.

7.2. The impact of tracers on lipid digestion kinetics

[Participant: Carole Knibbe]

Dietary fats are present in the diet under different types of structures, such as spread vs emulsions (notably in processed foods and enteral formula), and interest is growing regarding their digestion and intestinal absorption. In clinical trials, there is often a need to add stable isotope-labeled triacylglycerols (TAGs) as tracers to the ingested fat in order to track its intestinal absorption and further metabolic fate. Because most TAG tracers contain saturated fatty acids, they may modify the physicochemical properties of the ingested labeled fat and thereby its digestion. However, the actual impact of tracer addition on fat crystalline properties and lipolysis by digestive lipases still deserves to be explored. In this context, we monitored the thermal and polymorphic behavior of anhydrous milk fat (AMF) enriched in homogeneous TAGs tracers and further compared it with the native AMF using differential scanning calorimetry and power X-ray diffraction. As tracers, we used a mixture of tripalmitin, triolein and tricaprylin at 2 different concentrations (1.5 and 5.7wt%, which have been used in clinical trials). The addition of TAG tracers modified the AMF melting profile, especially at the highest tested concentration (5.7 wt%). Both AMF and AMF enriched with 1.5wt% tracers were completely melted around 37°C, i.e. close to the body temperature, while the AMF enriched with 5.7wt% tracers remained partially crystallized at this temperature. Similar trends were observed in both bulk and emulsified systems. Moreover, the kinetics of AMF polymorphic transformation was modified in the presence of tracers. While only β' form was observed in the native AMF, the β-form was clearly detected in the AMF containing 5.7wt% tracers. We further tested the impact of tracers on the lipolysis of AMF in bulk using a static in vitro model of duodenal digestion. Lipolysis of AMF enriched with 5.7wt% tracers was delayed compared

with that of AMF and AMF enriched with 1.5wt% tracers. Therefore, low amounts of TAG tracers including tripalmitin do not have a high impact on fat digestion, but one has to be cautious when using higher amounts of these tracers.

7.3. The control of synaptic plasticity by external factors

[participant: H. Berry]

The dorsal striatum exhibits bidirectional corticostriatal synaptic plasticity, NMDAR- and endocannabinoids-(eCB)-mediated, necessary for the encoding of procedural learning. Therefore, characterizing factors controlling corticostriatal plasticity is of crucial importance. Brain-derived neurotrophic factor (BDNF) and its receptor, the tropomyosine receptor kinase- B (TrkB), shape striatal functions and their dysfunction deeply affects basal ganglia. BDNF/TrkB signaling controls NMDAR-plasticity in various brain structures including the striatum. However, despite cross-talk between BDNF and eCBs, the role of BDNF in eCB- plasticity remains unknown. In https://hal.inria.fr/hal-02076121, we show that BDNF/TrkB signaling promotes eCB-plasticity (LTD and LTP) induced by rate-based (low-frequency stimulation) or spike-timing- based (spike-timing-dependent plasticity, STDP) paradigm in striatum. We show that TrkB activation is required for the expression and the scaling of both eCB-LTD and eCB-LTP. Using two-photon imaging of dendritic spines combined with patch-clamp recordings, we show that TrkB activation prolongs intracellular calcium transients, thus increasing eCB synthesis and release. We provide a mathematical model for the dynamics of the signaling pathways involved in corticostriatal plasticity. Finally, we show that TrkB activation enlarges the domain of expression of eCB-STDP. Our results reveal a novel role for BDNF/TrkB signaling in governing eCB-plasticity expression in striatum, and thus the engram of procedural learning.

7.4. A new model for calcium signals in tiny sub-cellular domains

[participants: A. Denizot, H. Soula, H. Berry]

Astrocytes, a glial cell type of the central nervous system, have emerged as detectors and regulators of neuronal information processing. Astrocyte excitability resides in transient variations of free cytosolic calcium concentration over a range of temporal and spatial scales, from sub-microdomains to waves propagating throughout the cell. Despite extensive experimental approaches, it is not clear how these signals are transmitted to and integrated within an astrocyte. The localization of the main molecular actors and the geometry of the system, including the spatial organization of calcium channels IP3R, are deemed essential. However, as most calcium signals occur in astrocytic ramifications that are too fine to be resolved by conventional light microscopy, most of those spatial data are unknown and computational modeling remains the only methodology to study this issue. In https://hal.inria.fr/hal-02184344v2, we propose an IP3R-mediated calcium signaling model for dynamics in such small sub-cellular volumes. To account for the expected stochasticity and low copy numbers, our model is both spatially explicit and particle-based. Extensive simulations show that spontaneous calcium signals arise in the model via the interplay between excitability and stochasticity. The model reproduces the main forms of calcium signals and indicates that their frequency crucially depends on the spatial organization of the IP3R channels. Importantly, we show that two processes expressing exactly the same calcium channels can display different types of calcium signals depending on the spatial organization of the channels. Our model with realistic process volume and calcium concentrations successfully reproduces spontaneous calcium signals that we measured in calcium micro-domains with confocal microscopy and predicts that local variations of calcium indicators might contribute to the diversity of calcium signals observed in astrocytes. To our knowledge, this model is the first model suited to investigate calcium dynamics in fine astrocytic processes and to propose plausible mechanisms responsible for their variability.

7.5. Evolution of genome size

Using the Aevol software, we investigated the dynamics of genome size under different evolutionary pressures (variation of mutation rates and variation of population sizes). The dynamics of the model enabled us to identify a new mutational pressure on genome size that spontaneously increase the fraction of non-coding

sequences. We showed that this mutational pressure interact with the selective pressure for robustness (knibbe et al., 2007), resulting in an equilibrium of genome size and non-coding proportion. Moreover, we showed that this equilibrium can change depending on the size of the population due to the resulting effect on selection intensity. A paper has been published in the proceedings of the ALife 2019 conference (cardes et al, 2019) and an article in in preparation.

7.6. Dynamics of evolutionary innovation

Using a combination of mathematical and computational models (NK-Fitness-Landscapes and Aevol), we investigated the dynamics of innovation in evolving systems. We showed that innovation is often triggered by specific mutational events, typically structural variation of the genome (e.g. duplications, inversions, ...). We further studied this effect and showed that innovation is due to the differences of time scale between the different kinds of mutations: fast mutations (typically point mutations) are rapidly exhausted, resulting in a fitness plateau. However, slow mutations (typically structural variations) can open new evolutionary paths, resulting in the population escaping from the fitness plateau. An article is in preparation in collaboration with Santiago F. Elena (CSIC, Spain).

7.7. Evolution of biological complexity

Using a modified version of the Aevol platform, we studied the evolution of complex features. By evolving population of organisms in conditions where complexity is counter-selected, we showed that complexity accumulates even in these conditions, i.e. even when complex organisms are less fit than simple ones. Moreover we showed that complex organisms are not more robust and not more evolvable than simple ones. This shows that evolution spontaneously initiate a "complexity ratchet" that forces complexity to grow. An article is in press in the Artificial Life Journal (to be published in 2020).

7.8. Dynamics of mutator strains

In a long-lasting collaboration with Utrecht University, we studied the dynamics of mutator strains in constant environments (mutator strains being individuals which mutation rate is increased by several orders of magnitude). Contrary to what is generally admitted, we showed that, although mutators initially suffer from a mutational burden (in coherence with the theory), they are able to quickly recover and avoid the burden. Moreover, we showed that they do so by contracting their coding genome compartment and expanding their non-coding compartment. This result show that mutators can thrive even in a constant environment (rutten et al., 2019).

7.9. Mutiscale phylogenetics models

[Participant: Eric Tannier]

We progressed in the modeling of multi-scale phylogenetic events: we gave an algorithm to infer gene conversions according to a phylogeny [7], a complexity result and an algorithm for transfers with replacements [6], and we devised a simulation tool integrating extinct species and horizontal inheritance [3].

7.10. Evolution of the Drosophila melanogaster Chromatin Landscape and Its Associated Proteins

[participant: A. Crombach]

In the nucleus of eukaryotic cells, genomic DNA associates with numerous protein complexes and RNAs, forming the chromatin landscape. Through a genome-wide study of chromatin-associated proteins in Drosophila cells, five major chromatin types were identified as a refinement of the traditional binary division into hetero- and euchromatin. These five types were given color names in reference to the Greek word chroma. They are defined by distinct but overlapping combinations of proteins and differ in biological and biochemical properties, including transcriptional activity, replication timing, and histone modifications. We assessed the evolutionary relationships of chromatin-associated proteins and presented an integrated view of the evolution and conservation of the fruit fly Drosophila melanogaster chromatin landscape. We combined homology prediction across a wide range of species with gene age inference methods to determine the origin of each chromatin-associated protein. This provided insight into the evolution of the different chromatin types. Our results indicate that for the euchromatic types, YELLOW and RED, young associated proteins are more specialized than old ones; and for genes found in either chromatin type, intron/exon structure is lineage-specific. Next, we provide evidence that a subset of GREEN-associated proteins is involved in a centromere drive in D. melanogaster. Our results on BLUE chromatin support the hypothesis that the emergence of Polycomb Group proteins is linked to eukaryotic multicellularity. In light of these results, we discuss how the regulatory complexification of chromatin links to the origins of eukaryotic multicellularity.

DRACULA Project-Team

5. New Results

5.1. Mathematical models describing the interaction between cancer and immune cells in the lymph node

To study the interplay between tumor progression and the immune response, we develop in [5] two new models describing the interaction between cancer and immune cells in the lymph node. The first model consists of partial differential equations (PDEs) describing the populations of the different types of cells. The second one is a hybrid discrete-continuous model integrating the mechanical and biochemical mechanisms that define the tumor-immune interplay in the lymph node. We use the continuous model to determine the conditions of the regimes of tumor-immune interaction in the lymph node. While we use the hybrid model to elucidate the mechanisms that contribute to the development of each regime at the cellular and tissue levels. We study the dynamics of tumor growth in the absence of immune cells. Then, we consider the immune response and we quantify the effects of immunosuppression and local EGF concentration on the fate of the tumor.

5.2. WASABI: a dynamic iterative framework for gene regulatory network inference

Background Inference of gene regulatory networks from gene expression data has been a long-standing and notoriously difficult task in systems biology. Recently, single-cell transcriptomic data have been massively used for gene regulatory network inference, with both successes and limitations. In the work [8], we propose an iterative algorithm called WASABI, dedicated to inferring a causal dynamical network from time-stamped single-cell data, which tackles some of the limitations associated with current approaches. We first introduce the concept of waves, which posits that the information provided by an external stimulus will affect genes one-by-one through a cascade, like waves spreading through a network. This concept allows us to infer the network one gene at a time, after genes have been ordered regarding their time of regulation. We then demonstrate the ability of WASABI to correctly infer small networks, which have been simulated in silico using a mechanistic model consisting of coupled piecewise-deterministic Markov processes for the proper description of gene expression at the single-cell level. We finally apply WASABI on in vitro generated data on an avian model of erythroid differentiation. The structure of the resulting gene regulatory network sheds a new light on the molecular mechanisms controlling this process. In particular, we find no evidence for hub genes and a much more distributed network structure than expected. Interestingly, we find that a majority of genes are under the direct control of the differentiation-inducing stimulus. Conclusions Together, these results demonstrate WASABI versatility and ability to tackle some general gene regulatory networks inference issues. It is our hope that WASABI will prove useful in helping biologists to fully exploit the power of time-stamped single-cell data.

5.3. A multiscale model of platelet-fibrin thrombus growth in the flow

Thrombosis is a life-threatening clinical condition characterized by the obstruction of blood flow in a vessel due to the formation of a large thrombus. The pathogenesis of thrombosis is complex because the type of formed clots depends on the location and function of the corresponding blood vessel. To explore this phenomenon, we develop in [9] a novel multiscale model of platelet-fibrin thrombus growth in the flow. In this model, the regulatory network of the coagulation cascade is described by partial differential equations. Blood flow is introduced using the Navier–Stokes equations and the clot is treated as a porous medium. Platelets are represented as discrete spheres that migrate with the flow. Each platelet can attach to the thrombus, aggregate, become activated, express proteins on its surface, detach, and/or become non-adhesive. The interaction of platelets with blood flow is captured using the Immersed Boundary Method (IBM). We use the model to

investigate the role of flow conditions in shaping the dynamics of venous and arterial thrombi. We describe the formation of red and white thrombi under venous and arterial flow respectively and highlight the main characteristics of each type. We identify the different regimes of normal and pathological thrombus formation depending on flow conditions.

5.4. Mathematical modeling of platelet production

- In [10], we analyze the existence of oscillating solutions and the asymptotic convergence for a nonlinear delay differential equation arising from the modeling of platelet production. We consider four different cell compartments corresponding to different cell maturity levels: stem cells, megakaryocytic progenitors, megakaryocytes, and platelets compartments, and the quantity of circulating thrombopoietin (TPO), a platelet regulation cytokine.
- In [11], we analyze the stability of a differential equation with two delays originating from a model for a population divided into two subpopulations, immature and mature, and we apply this analysis to a model for platelet production. The dynamics of mature individuals is described by the following nonlinear differential equation with two delays: $x'(t) = -\lambda x(t) + g(x(t-\tau_1)) - g(x(t-\tau_1-\tau_2))e^{-\lambda\tau_2}$. The method of D-decomposition is used to compute the stability regions for a given equilibrium. The center manifold theory is used to investigate the steady-state bifurcation and the Hopf bifurcation. Similarly, analysis of the center manifold associated with a double bifurcation is used to identify a set of parameters such that the solution is a torus in the pseudo- phase space. Finally, the results of the local stability analysis are used to study the impact of an increase of the death rate γ or of a decrease of the survival time τ_2 of platelets on the onset of oscillations. We show that the stability is lost through a small decrease of survival time (from 8.4 to 7 days), or through an important increase of the death rate (from 0.05 to 0.625 days^{-1}).
- In [12], we analyze the stability of a system of differential equations with a threshold-defined delay arising from a model for platelet production. We consider a maturity-structured population of megakaryocyte progenitors and an age-structured population of platelets, where the cytokine thrombopoietin (TPO) increases the maturation rate of progenitors. Using the quasi-steady-state approximation for TPO dynamics and the method of characteristics, partial differential equations are reduced to a system of two differential equations with a state-dependent delay accounting for the variable maturation rate. We start by introducing the model and proving the positivity and boundedness of the solutions. Then we use a change of variables to obtain an equivalent system of two differential equations with a constant delay, from which we prove existence and uniqueness of the solution. As linearization around the unique positive steady state yields a transcendental characteristic equation of third degree, we introduce the main result, a new framework for stability analysis on models with fixed delays. This framework is then used to describe the stability of the megakaryopoiesis with respect to its parameters. Finally, with parameters being obtained and estimated from data, we give an example in which oscillations appear when the death rate of progenitors is increased 10-fold.

5.5. Nonlinear analysis of a model for yeast cell communication

In [13], we study the non-linear stability of a coupled system of two non-linear transport-diffusion equations set in two opposite half-lines. This system describes some aspects of yeast pairwise cellular communication, through the concentration of some protein in the cell bulk and at the cell boundary. We show that it is of bistable type, provided that the intensity of active molecular transport is large enough. We prove the non-linear stability of the most concentrated steady state, for large initial data, by entropy and comparison techniques. For small initial data we prove the self-similar decay of the molecular concentration towards zero. Informally speaking, the rise of a dialog between yeast cells requires enough active molecular transport in this model. Besides, if the cells do not invest enough in the communication with their partner, they do not respond to each other; but a sufficient initial input from each cell in the dialog leads to the establishment of a stable activated state in both cells.

5.6. Alzheimer's disease and prion: An in vitro mathematical model

Alzheimer's disease (AD) is a fatal incurable disease leading to progressive neuron destruction. AD is caused in part by the accumulation in the brain of $A\beta$ monomers aggregating into oligomers and fibrils. Oligomers are amongst the most toxic structures as they can interact with neurons via membrane receptors, including PrP^c proteins. This interaction leads to the misconformation of PrP^c into pathogenic oligomeric prions, PrP^{ol}. In [14], we develop a model describing in vitro $A\beta$ polymerization process. We include interactions between oligomers and PrP^c, causing the misconformation of PrP^c into PrP^{ol}. The model consists of nine equations, including size structured transport equations, ordinary differential equations and delayed differential equations. We analyse the well-posedness of the model and prove the existence and uniqueness of solutions of our model using Schauder fixed point theorem and Cauchy-Lipschitz theorem. Numerical simulations are also provided to give an illustration of the profiles that can be obtained with this model.

5.7. Calibration, Selection and Identifiability Analysis of a Mathematical Model of the in vitro Erythropoiesis in Normal and Perturbed Contexts

The in vivo erythropoiesis, which is the generation of mature red blood cells in the bone marrow of whole organisms, has been described by a variety of mathematical models in the past decades. However, the in vitro erythropoiesis, which produces red blood cells in cultures, has received much less attention from the modelling community. In the paper [15], we propose the first mathematical model of in vitro erythropoiesis. We start by formulating different models and select the best one at fitting experimental data of in vitro erythropoietic differentiation obtained from chicken erythroid progenitor cells. It is based on a set of linear ODE, describing 3 hypothetical populations of cells at different stages of differentiation. We then compute confidence intervals for all of its parameters estimates, and conclude that our model is fully identifiable. Finally, we use this model to compute the effect of a chemical drug called Rapamycin, which affects all states of differentiation in the culture, and relate these effects to specific parameter variations. We provide the first model for the kinetics of in vitro cellular differentiation which is proven to be identifiable. It will serve as a basis for a model which will better account for the variability which is inherent to the experimental protocol used for the model calibration.

5.8. Model-based assessment of the role of uneven partitioning of molecular content on heterogeneity and regulation of differentiation in CD8 T-cell immune responses

Activation of naive CD8 T-cells can lead to the generation of multiple effector and memory subsets. Multiple parameters associated with activation conditions are involved in generating this diversity that is associated with heterogeneous molecular contents of activated cells. Although naive cell polarisation upon antigenic stimulation and the resulting asymmetric division are known to be a major source of heterogeneity and cell fate regulation, the consequences of stochastic uneven partitioning of molecular content upon subsequent divisions remain unclear yet. In [16], we aim at studying the impact of uneven partitioning on molecular-content heterogeneity and then on the immune response dynamics at the cellular level. To do so, we introduce a multiscale mathematical model of the CD8 T-cell immune response in the lymph node. In the model, cells are described as agents evolving and interacting in a 2D environment while a set of differential equations, embedded in each cell, models the regulation of intra and extracellular proteins involved in cell differentiation. Based on the analysis of in silico data at the single cell level, we 1 show that immune response dynamics can be explained by the molecular-content heterogeneity generated by uneven partitioning at cell division. In particular, uneven partitioning acts as a regulator of cell differentiation and induces the emergence of two coexisting sub-populations of cells exhibiting antagonistic fates. We show that the degree of unevenness of molecular partitioning, along all cell divisions, affects the outcome of the immune response and can promote the generation of memory cells.

5.9. Spatial lymphocyte dynamics in lymph nodes predicts the cytotoxic T-Cell frequency needed for HIV infection control

The surveillance of host body tissues by immune cells is central for mediating their defense function. In vivo imaging technologies have been used to quantitatively characterize target cell scanning and migration of lymphocytes within lymph nodes (LNs). The translation of these quantitative insights into a predictive understanding of immune system functioning in response to various perturbations critically depends on computational tools linking the individual immune cell properties with the emergent behavior of the immune system. By choosing the Newtonian second law for the governing equations, we developed in [17] a broadly applicable mathematical model linking individual and coordinated T-cell behaviors. The spatial cell dynamics is described by a superposition of autonomous locomotion, intercellular interaction, and viscous damping processes. The model is calibrated using in vivo data on T-cell motility metrics in LNs such as the translational speeds, turning angle speeds, andmeandering indices. Themodel is applied to predict the impact of T-cell motility on protection against HIV infection, i.e., to estimate the threshold frequency of HIV-specific cytotoxic T cells (CTLs) that is required to detect productively infected cells before the release of viral particles starts. With this, it provides guidance for HIV vaccine studies allowing for the migration of cells in fibrotic LNs.

5.10. Drugs modulating stochastic gene expression affect the erythroid differentiation process

To better understand the mechanisms behind cells decision-making to differentiate, we assessed in [18] the influence of stochastic gene expression (SGE) modulation on the erythroid dif- ferentiation process. It has been suggested that stochastic gene expression has a role in cell fate decision-making which is revealed by single-cell analyses but studies dedicated to demonstrate the consistency of this link are still lacking. Recent observations showed that SGE significantly increased during differentiation and a few showed that an increase of the level of SGE is accompanied by an increase in the differentiation process. However, a consistent relation in both increasing and decreasing directions has never been shown in the same cellular system. Such demonstration would require to be able to experimentally manipulate simultaneously the level of SGE and cell differentiation in order to observe if cell behavior matches with the current theory. We identified three drugs that modulate SGE in primary erythroid progenitor cells. Both Artemisinin and Indomethacin decreased SGE and reduced the amount of differentiated cells. On the contrary, a third component called MB-3 simultaneously increased the level of SGE and the amount of differentiated cells. We then used a dynamical modelling approach which confirmed that differentiation rates were indeed affected by the drug treatment. Using single-cell analysis and modeling tools, we provide experimental evidence that, in a physiologically relevant cellular system, SGE is linked to differentiation.

5.11. Stochastic gene expression with a multistate promoter: breaking down exact distributions

We consider in [19] a stochastic model of gene expression in which transcription depends on a multistate promoter, including the famous two-state model and refractory promoters as special cases, and focus on deriving the exact stationary distribution. Building upon several successful approaches, we present a more unified viewpoint that enables us to simplify and generalize existing results. In particular, the original jump process is deeply related to a multivariate piecewise-deterministic Markov process that may also be of interest beyond the biological field. In a very particular case of promoter configuration, this underlying process is shown to have a simple Dirichlet stationary distribution. In the general case, the corresponding marginal distributions extend the well-known class of Beta products, involving complex parameters that directly relate to spectral properties of the promoter transition matrix. Finally, we illustrate these results with biologically plausible examples.

5.12. Cell generation dynamics underlying naive T-cell homeostasis in adult humans

Thymic involution and proliferation of naive T-cells both contribute to shaping the naive T-cell repertoire as humans age, but a clear understanding of the roles of each throughout a human life span has been difficult to determine. By measuring nuclear bomb test-derived 14 C in genomic DNA, we determined in [22] the turnover rates of CD4 + and CD8 + naive T-cell populations and defined their dynamics in healthy individuals ranging from 20 to 65 years of age. We demonstrate that naive T-cell generation decreases with age because of a combination of declining peripheral division and thymic production during adulthood. Concomitant decline in T-cell loss compensates for decreased generation rates. We investigated putative mechanisms underlying age-related changes in homeostatic regulation of CD4+ naive T-cell turnover, using mass cytometry to profile candidate signaling pathways involved in T-cell activation and proliferation relative to CD31 expression, a marker of thymic proximity for the CD4+ naive T-cell population. We show that basal nuclear factor kB (NF-kB) phosphory-lation positively correlated with CD31 expression and thus is decreased in peripherally expanded naive T-cell clones. Functionally, we found that NF-kB signaling was essential for naive T-cell proliferation to the homeostatic growth factor interleukin (IL)-7, and reduced NF-kB phosphorylation in CD4 + CD31 - naive T cells is linked to reduced homeostatic proliferation potential. Our results reveal an age-related decline in naive T-cell turnover as a putative regulator of naive T-cell diversity and identify a molecular pathway that restricts proliferation of peripherally expanded naive T-cell clones that accumulate with age.

5.13. Erythroid differentiation displays a peak of energy consumption concomitant with glycolytic metabolism rearrangements

Our previous single-cell based gene expression analysis pointed out significant variations of LDHA level during erythroid differentiation. Deeper investigations highlighted that a metabolic switch occurred along differentiation of erythroid cells. More precisely we showed in [26] that self-renewing progenitors relied mostly upon lactate-productive glycolysis, and required LDHA activity, whereas differentiating cells, mainly involved mitochondrial oxidative phosphorylation (OXPHOS). These metabolic rearrangements were coming along with a particular temporary event, occurring within the first 24h of erythroid differentiation. The activity of glycolytic metabolism and OXPHOS rose jointly with oxgene consumption dedicated to ATP production at 12-24h of the differentiation process before lactate-productive glycolysis sharply fall down and energy needs decline. Finally, we demonstrated that the metabolic switch mediated through LDHA drop and OXPHOS upkeep might be necessary for erythroid differentiation. We also discuss the possibility that metabolism, gene expression and epigenetics could act together in a circular manner as a driving force for differentiation.

ERABLE Project-Team

6. New Results

6.1. General comments

We present in this section the main results obtained in 2019.

We tried to organise these along the four axes as presented above. Clearly, in some cases, a result obtained overlaps more than one axis. In such case, we chose the one that could be seen as the main one concerned by such results.

We chose not to detail here the results on more theoretical aspects of computer science when these are initially addressed in contexts not directly related to computational biology even though those on string [11], [36], [40], [41], [23], [45] and graph algorithms in general [35], [39], [38], [17], [43] are relevant for life sciences, such as for instance pan-genome analysis, or could become more specifically so in a near future. One important example of the latter concerns enumeration algorithms that has always been at the heart of the computer science and mathematics interests of the team. In such context, the so-called reconfiguration problem which asks whether one solution can be transformed into the other in a step-by-step fashion such that each intermediate solution is also feasible is of particular relevance. This was explored in the context of a perfect matching problem [37].

A few other results of 2019 are not mentioned in this report, not because the corresponding work is not important, but because it was likewise more specialised [8], [9], [12], [44]. In the same way, also for space reasons, we chose not to detail the results presented in some biological papers of the team when these did not require a mathematical or algorithmic input [16], [22].

On the other hand, we do mention a couple of works that were in preparation or about to be submitted towards the end of 2018.

6.2. Axis 1: Genomics

Transcriptome profiling using Nanopore sequencing Our vision of DNA transcription and splicing has changed dramatically with the introduction of short-read sequencing. These high-throughput sequencing technologies promised to unravel the complexity of any transcriptome. Generally gene expression levels are well-captured using these technologies, but there are still remaining caveats due to the limited read length and the fact that RNA molecules had to be reverse transcribed before sequencing. Oxford Nanopore Technologies has recently launched a portable sequencer which offers the possibility of sequencing long reads and most importantly RNA molecules. In [28], we generated a full mouse transcriptome from brain and liver using such Oxford Nanopore device. As a comparison, we sequenced RNA (RNA-Seq) and cDNA (cDNA-Seq) molecules using both long and short reads technologies and tested the TeloPrime preparation kit, dedicated to the enrichment of full-length transcripts. Using spike-in data, we confirmed in [28] that expression levels are efficiently captured by cDNA-Seq using short reads. More importantly, Oxford Nanopore RNA-Seq tends to be more efficient, while cDNA-Seq appears to be more biased. We further showed that the cDNA library preparation of the Nanopore protocol induces read truncation for transcripts containing internal runs of T's. This bias is marked for runs of at least 15 T's, but is already detectable for runs of at least 9 T's and therefore concerns more than 20% of the expressed transcripts in mouse brain and liver. Finally, we outlined that bioinformatic challenges remain ahead for quantifying at the transcript level, especially when reads are not full-length. Accurate quantification of repeat-associated genes such as processed pseudogenes also remains difficult, and we show in the paper that current mapping protocols which map reads to the genome largely over-estimate their expression, at the expense of their parent gene.

Genotyping and variant detection The amount of genetic variation discovered and characterised in human populations is huge, and is growing rapidly with the widespread availability of modern sequencing technologies. Such a great deal of variation data, that accounts for human diversity, leads to various challenging computational tasks, including variant calling and genotyping of newly sequenced individuals. The standard pipelines for addressing these problems include read mapping, which is a computationally expensive procedure. A few mapping-free tools were proposed in recent years to speed up the genotyping process. While such tools have highly efficient run-times, they focus on isolated, bi-allelic SNPs, providing limited support for multi-allelic SNPs, indels, and genomic regions with high variant density. To address these issues, we introduced MALVA, a fast and lightweight mapping-free method to genotype an individual directly from a sample of reads [10]. MALVA is the first mapping-free tool that is able to genotype multi-allelic SNPs and indels, even in high density genomic regions, and to effectively handle a huge number of variants such as those provided by the 1000 Genome Project. An experimental evaluation on whole-genome data shows that MALVA requires one order of magnitude less time to genotype a donor than alignment-based pipelines, providing similar accuracy. Remarkably, on indels, MALVA provides even better results than the most widely adopted variant discovery tools.

Still on the issue of SNP detection, in [25], we developed the positional clustering theory that (i) describes how the extended Burrows–Wheeler Transform (eBWT) of a collection of reads tends to cluster together bases that cover the same genome position, (ii) predicts the size of such clusters, and (iii) exhibits an elegant and precise LCP array based procedure to locate such clusters in the eBWT. Based on this theory, we designed and implemented an alignment-free and reference-free SNP calling method, and we devised a SNP calling pipeline. Experiments on both synthetic and real data show that SNPs can be detected with a simple scan of the eBWT and LCP arrays as, in agreement with our theoretical framework, they are within clusters in the eBWT of the reads. Finally, our tool intrinsically performs a reference-free evaluation of its accuracy by returning the coverage of each SNP. Based on the results of the experiments on synthetic and real data, we conclude that the positional clustering framework can be effectively used for the problem of identifying SNPs, and it appears to be a promising approach for calling other types of variants directly on raw sequencing data.

Finally, variant detection and various related algorithmic problems were extensively explored in the PhD of Leandro I. S. de Lima [2] defended in April 2019.

Bubble generator Bubbles are pairs of internally vertex-disjoint (s, t)-paths in a directed graph, which have many applications in the processing of DNA and RNA data such as variant calling as presented above. Listing and analysing all bubbles in a given graph is usually unfeasible in practice, due to the exponential number of bubbles present in real data graphs. In [4], we proposed a notion of bubble generator set, *i.e.*, a polynomial-sized subset of bubbles from which all the other bubbles can be obtained through a suitable application of a specific symmetric difference operator. This set provides a compact representation of the bubble space of a graph. A bubble generator can be useful in practice, since some pertinent information about all the bubbles can be more conveniently extracted from this compact set. We provided a polynomial-time algorithm to decompose any bubble of a graph into the bubbles of such a generator in a tree-like fashion. Finally, we presented two applications of the bubble generator on a real RNA-seq dataset.

Genome assembly The continuous improvement of long-read sequencing technologies along with the development of ad-doc algorithms has launched a new *de novo* assembly era that promises high-quality genomes. However, it has proven difficult to use only long reads to generate accurate genome assemblies of large, repeat-rich human genomes. To date, most of the human genomes assembled from long error-prone reads add accurate short reads to further improve the consensus quality (polishing). In a paper to be submitted before the end of 2019 (with as main authors A. di Genova and M.-F. Sagot), we report the development of an algorithm for hybrid assembly, WENGAN, and its application to hybrid sequence datasets from four human samples. WENGAN implements efficient algorithms that exploit the sequence information of short and long reads to tackle assembly contiguity as well as consensus quality. We show that the resulting genome assemblies have high contiguity (contig NG50:16.67-62.06 Mb), few assembly errors (contig NGA50:10.9-45.91 Mb), good consensus quality (QV:27.79-33.61), high gene completeness (BUSCO complete: 94.6-95.1%), and consume few computational resources (CPU hours:153-1027). In particular, the WENGAN assembly of the

haploid CHM13 sample achieved a contig NG50 of 62.06 Mb (NGA50:45.91 Mb), which surpasses the contiguity of the current human reference genome (GRCh38 contig NG50:57.88 Mb). Because of its lower cost, WENGAN is an important step towards the democratisation of the *de novo* assembly of human genomes. WENGAN is available at https://github.com/adigenova/wengan.

On assembly still, although haplotype-aware genome assembly plays an important role in genetics, medicine and various other disciplines, the generation of haplotype-resolved *de novo* assemblies remains a major challenge. Beyond distinguishing between errors and true sequential variants, one needs to assign the true variants to the different genome copies. Recent work has pointed out that the enormous quantities of traditional NGS read data have been greatly underexploited in terms of haplotig computation so far, which reflects the fact that the methodology for reference independent haplotig computation has not yet reached maturity. We presented in [7] a new approach, called POLYploid genome fitTEr (POLYTE) for a *de novo* generation of haplotigs for diploid and polyploid genomes of known ploidy. Our method follows an iterative scheme where in each iteration reads or contigs are joined, based on their interplay in terms of an underlying haplotype-aware overlap graph. Along the iterations, contigs grow while preserving their haplotype identity. Benchmarking experiments on both real and simulated data demonstrate that POLYTE establishes new standards in terms of error-free reconstruction of haplotype-specific sequences. As a consequence, POLYTE outperforms state-of-the-art approaches in various relevant aspects, notably in polyploid settings.

Others Besides the above, we have also explored a proteogenomics workflow for the expert annotation of eukaryotic genomes [18], as well as a technology- and species-independent simulator of sequencing data and genomic variants [42].

6.3. Axis 2: Metabolism and post-transcriptional regulation

Multi-objective metabolic mixed integer optimisation with an application to yeast strain engineering In a paper submitted and already available in bioRxiv (https://www.biorxiv.org/content/early/2018/11/22/476689), we explored the concept of multi-objective optimisation in the field of metabolic engineering when both continuous and integer decision variables are involved in the model. In particular, we proposed a multi-objective model which may be used to suggest reaction deletions that maximise and/or minimise several functions simultaneously. The applications may include, among others, the concurrent maximisation of a bioproduct and of biomass, or maximisation of a bioproduct while minimising the formation of a given by-product, two common requirements in microbial metabolic engineering. Production of ethanol by the widely used cell factory *Saccharomyces cerevisiae* was adopted as a case study to demonstrate the usefulness of the proposed approach in identifying genetic manipulations that improve productivity and yield of this economically highly relevant bioproduct. We did an *in vivo* validation and we could show that some of the predicted deletions exhibit increased ethanol levels in comparison with the wild-type strain. The multi-objective programming framework we developed, called MOMO, is open-source and uses POLYSCIP as underlying multi-objective solver. This is part of the work of Ricardo de Andrade, who was until the end of 2018 postdoc at Unversity of São Paulo with Roberto Marcondes, and in ERABLE. It is joint work with Susana Vinga, external collaborator of ERABLE and partner of the Inria Associated Team Compasso.

Metabolic shifts Analysis of differential expression of genes is often performed to understand how the metabolic activity of an organism is impacted by a perturbation. However, because the system of metabolic regulation is complex and all changes are not directly reflected in the expression levels, interpreting these data can be difficult. In [26], we presented a new algorithm and computational tool that uses a genome-scale metabolic reconstruction to infer metabolic changes from differential expression data. Using the framework of constraint-based analysis, our method produces a qualitative hypothesis of a change in metabolic activity. In other words, each reaction of the network is inferred to have increased, decreased, or remained unchanged in flux. In contrast to similar previous approaches, our method does not require a biological objective function and does not assign on/off activity states to genes. An implementation is provided and is available online at the address https://github.com/htpusa/moomin. We applied the method to three published datasets to show that it successfully accomplishes its two main goals: confirming or rejecting metabolic changes suggested by differentially expressed genes based on how well they fit in as parts of a coordinated metabolic change, as well

as inferring changes in reactions whose genes did not undergo differential expression. The above work was also part of the PhD of Taneli Pusa [3] defended in February 2019.

Metabolic games Game theory is a branch of applied mathematics originally developed to describe and reason about situations where two or more rational agents, the "homo economicus", are faced with choices and have potentially conflicting goals. All participants want to maximise their own well-being, but are doing so taking into account that everyone else is doing the same. Thus paradoxical, suboptimal, outcomes are possible and even common. Evolutionary game theory was born out of the realisation that rational choice can be replaced by natural selection: in the course of evolution the strategy (phenotype) that would "win" the game would prevail by simply proliferating more successfully thanks to its success in the "game". It turns out that phenotype prediction in the context of metabolic networks is exactly the type of problem that evolutionary game theory was meant to answer: given a set of choices (as defined by a metabolic network reconstruction), what will be the actual metabolism observed? In other words, if we culture a set of organisms together in a given medium, which are the phenotype(s) that emerge as winners? In [27], we sought to provide a short introduction to both evolutionary game theory and its use in the context of metabolic modelling. This work was also part of the PhD of Taneli Pusa [3].

6.4. Axis 3: (Co)Evolution

Modelling invasion Nowadays, the most used model in studies of the coevolution of hosts and symbionts is phylogenetic tree reconciliation. A crucial issue in this model is that from a biological point of view, reasonable cost values for an event-based reconciliation are not easily chosen. Different methods have been developed to infer such cost values for a given pair of host and symbiont trees, including one we established in the past. However, a major limitation of these methods is their inability to model the "invasion" of different host species by a same symbiont species (referred to as a spread event), which is often observed in symbiotic relations. Indeed, many symbionts are generalist. For instance, the same species of insect may pollinate different species of plants. In a paper currently in preparation, we propose a method, called AMOCOALA, which for a given pair of host and symbiont trees, estimates the frequency of the cophylogenetic events, in presence of spread events, based on an approximate Bayesian computation (ABC) approach that may be more efficient than a classical likelihood method. The algorithm that we propose on one hand provides more confidence in the set of costs to be used for a given pair of host and symbiont trees, while on the other hand, it allows to estimate the frequency of the events even in the case of large datasets. We evaluated our method on both synthetic and real datasets.

Co-divergence and tree topology In reconstructing the common evolutionary history of hosts and symbionts, the current method of choice is the phylogenetic tree reconciliation. In this model, we are given a host tree H, a symbiont tree S, and a function σ mapping the leaves of S to the leaves of H and the goal is to find, under some biologically motivated constraints, a reconciliation, that is a function from the vertices of S to the vertices of H that respects σ and allows the identification of biological events such as co-speciation, duplication and host switch. The maximum co-divergence problem consists in finding the maximum number of co-speciations in a reconciliation. This problem is NP-hard for arbitrary phylogenetic trees and no approximation algorithm is known. In [14], we considered the influence of tree topology on the maximum co-divergence problem. In particular, we focused on a particular tree structure, namely caterpillar, and showed that in this case the heuristics that are mostly used in the literature provide solutions that can be arbitrarily far from the optimal value. We then proved that finding the max co-divergence is equivalent to computing the maximum length of a subsequence with certain properties of a given permutation. This equivalence leads to two consequences: (i) it shows that we can compute efficiently in polynomial time the optimal time-feasible reconciliation, and (ii) it can be used to understand how much the tree topology influences the value of the maximum number of co-speciations.

6.5. Axis 4: Human and animal health

Rare disease studies Minor intron splicing plays a central role in human embryonic development and survival. Indeed, biallelic mutations in RNU4ATAC, transcribed into the minor spliceosomal U4atac snRNA, are responsible for three rare autosomal recessive multimalformation disorders named Taybi-Linder (TALS/MOPD1), Roifman (RFMN), and Lowry-Wood (LWS) syndromes, which associate numerous overlapping signs of varying severity. Although RNA-seq experiments have been conducted on a few RFMN patient cells, none have been performed in TALS, and more generally no in-depth transcriptomic analysis of the 700 human genes containing a minor (U12-type) intron had been published as yet. We thus sequenced RNA from cells derived from five skin, three amniotic fluid, and one blood biosamples obtained from seven unrelated TALS cases and from age- and sex-matched controls. This allowed us to describe for the first time the mRNA expression and splicing profile of genes containing U12-type introns, in the context of a functional minor spliceosome. Concerning RNU4ATAC-mutated patients, we showed in [15] that as expected, they display distinct U12-type intron splicing profiles compared to controls, but that rather unexpectedly the mRNA expression levels are mostly unchanged. Furthermore, although U12-type intron missplicing concerns most of the expressed U12 genes, the level of U12-type intron retention is surprisingly low in fibroblasts and amniocytes, and much more pronounced in blood cells. Interestingly, we found several occurrences of introns that can be spliced using either U2, U12, or a combination of both types of splice site consensus sequences, with a shift towards splicing using preferentially U2 sites in TALS patients' cells compared to controls.

This work is part of the PhD of Audric Cologne [1] defended in October 2019.

Cancer studies Circular RNAs (circRNAs) are a class of RNAs that is under increasing scrutiny, although their functional roles are debated. In [30], we analysed RNA-seq data of 348 primary breast cancers and developed a method to identify circRNAs that does not rely on unmapped reads or known splice junctions. We identified 95,843 circRNAs, of which 20,441 were found recurrently. Of the circRNAs that match exon boundaries of the same gene, 668 showed a poor or even negative ($R < 0.2$) correlation with the expression level of the linear gene. An *In silico* analysis showed that only a minority (8.5%) of circRNAs could be explained by known splicing events. Both these observations suggest that specific regulatory processes for circRNAs exist. We confirmed the presence of circRNAs of CNOT2, CREBBP, and RERE in an independent pool of primary breast cancers. We identified circRNA profiles associated with subgroups of breast cancers and with biological and clinical features, such as amount of tumour lymphocytic infiltrate and proliferation index. siRNA-mediated knockdown of circCNOT2 was shown to significantly reduce viability of the breast cancer cell lines MCF-7 and BT-474, further underlining the biological relevance of circRNAs. Furthermore, we found that circular, and not linear, CNOT2 levels are predictive for progression-free survival time to aromatase inhibitor (AI) therapy in advanced breast cancer patients, and found that circCNOT2 is detectable in cell-free RNA from plasma. We showed that circRNAs are abundantly present, show characteristics of being specifically regulated, are associated with clinical and biological properties, and thus are relevant in breast cancer.

Other cancer studies have concerned the automatic discovery of the 100-miRNA signature for cancer classification [21], an Integrative and comparative genomic analysis to identify clinically relevant pulmonary carcinoid groups and unveil the supra-carcinoids [5], [complete with 2 papers not yet entered in Hal], and finally the investigation of new therapeutic interventions that are needed to increase the immunogenicity of tumours and overcome the resistance to these immuno-therapies [29].

Infection studies *Mycoplasma hyopneumoniae* is an economically devastating pathogen in the pig farming industry, however little is known about its relation with the swine host. To improve our understanding on this interaction, we infected epithelial cells with *M. hyopneumoniae* to identify the effects of the infection on the expression of swine genes and miRNAs. In addition, we identified miRNAs differentially expressed (DE) in the extracellular milieu and in exosome-like vesicles released by infected cells. A total of 1,268 genes and 170 miRNAs were DE post-infection ($p<0.05$). We identified the up-regulation of genes related to redox homeostasis and antioxidant defense, most of them putatively regulated by the transcription factor NRF2. Down-regulated genes were enriched in cytoskeleton and ciliary function, which could partially explain *M. hyopneumoniae* induced ciliostasis. Our predictions showed that DE miRNAs could be regulating the

aforementioned functions, since we detected down-regulation of miRNAs predicted to target antioxidant genes and up-regulation of miRNAs targeting ciliary and cytoskeleton genes. Based on these observations, *M. hyopneumoniae* seems to elicit an antioxidant response induced by NRF2 in infected cells; in addition, we propose that ciliostasis caused by this pathogen might be related to down-regulation of ciliary genes. The paper presenting these results has been submitted and is in revision.

Others Besides the above, a first step towards deep learning assisted genotype-phenotype association in whole genome-sized data has been explored in the context of predicting amyotrophic lateral sclerosis [34].

IBIS Project-Team

6. New Results

6.1. Analysis of fluorescent reporter gene data

The use of fluorescent and luminescent reporter genes allows real-time monitoring of gene expression, both at the level of individual cells and cell populations (Section 3.2). Over the years, many useful resources have appeared, such as libraries of reporter strains for model organisms and computer tools for designing reporter plasmids. Moreover, the widespread adoption of thermostated microplate readers in experimental laboratories has made it possible to automate and multiplex reporter gene assays on the population level. This has resulted in large time-series data sets, typically comprising $10^5 - 10^6$ measurements of absorbance, fluorescence, and luminescence for 10^3 wells on the microplate. In order to fully exploit these data sets, we need sound mathematical methods to infer biologically relevant quantities from the primary data and computer tools to apply the methods in an efficient and user-friendly manner.

In the past few years we developed novel methods for the analysis of reporter gene data obtained in microplate experiments, based on the use of regularized linear inversion. This allows a range of estimation problems to be solved, notably the inference of growth rate, promoter activity, and protein concentration profiles. The linear inversion methods, published in *Bioinformatics* in 2015 [13], have been implemented in the Python package WELLFARE and integrated in the web application WELLINVERTER. Funded by a grant from the Institut Français de Bioinformatique (IFB), we improved WellInverter by developing a parallel computational architecture with a load balancer to distribute the analysis queries over several back-end servers, a new graphical user interface, and a plug-in system for defining high-level routines for parsing data files produced by microplate readers from different manufacturers. This has resulted in a scalable and user-friendly web service providing a guaranteed quality of service, in terms of availability and response time. The web service has been redeployed on the new IFB cloud and on an Inria server, accompanied by extensive user documentation, online help, and a tutorial. An article on WELLINVERTER, illustrating the use of the tool by analyzing data of the expression of a fluorescent reporter gene controlled by a phage promoter in growing *Escherichia coli* populations, was published in *BMC Bioinformatics* this year [22]. We notably show that the expression pattern in different growth media, supporting different growth rates, corresponds to the pattern expected for a constitutive gene.

6.2. Stochastic modeling and identification of gene regulatory networks in bacteria

At the single-cell level, the processes that govern single-cell dynamics in general and gene expression in particular are better described by stochastic models rather than the deterministic models underlying the linear inversion methods discussed in Section 6.1 . Modern techniques for the real-time monitoring of gene expression in single cells enable one to apply stochastic modelling to study the origins and consequences of random noise in response to various environmental stresses, and the emergence of phenotypic variability. The potential impact of single-cell stochastic analysis and modelling ranges from a better comprehension of the biochemical regulatory mechanisms underlying cellular phenotypes to the development of new strategies for the (computer assisted or genetically engineered) control of cell populations and even of single cells.

Work in IBIS on gene expression and interaction dynamics at the level of individual cells is addressed in terms of identification of parametric intrinsic noise models, on the one hand, and the nonparametric inference of gene expression statistics, on the other hand, from population snapshot data. Along with modelling and inference, identifiability analysis is dedicated special attention. The investigation of the problem of reconstructing promoter activity statistics from reporter gene population snapshot data has led to a full-blown spectral analysis and reconstruction method for reporter gene systems. In the context of the ANR project MEMIP (Section 7.2

), we have characterized reporter systems as noisy linear systems operating on a stochastic input (promoter activity), and developed an inversion method for estimation of promoter activation statistics from reporter population snapshots. The method has been demonstrated on simulated data. Theoretical as well as simulation resuts have been published in *Automatica* this year [15], and will be the object of application to real data.

One of the key limitations of the method is the assumption of stationary promoter activation statistics. In the context of controlled gene expression processes, this may hamper applicability of the method. In response to this, an extension of the method for so-called modulated processes (stationary processes reshaped by a time-varying control input), has been developed and demonstrated on simulations of controlled gene expression. Results were submitted for possible presentation and publication in the proceedings of the IFAC world congress 2020.

6.3. Mathematical analysis of structured branching populations

The investigation of cellular populations at the single-cell level has led to the discovery of important phenomena, such as the co-occurrence of different phenotypes in an isogenic population. Novel experimental techniques, such as time-lapse fluorescence microscopy combined with the use of microfluidic devices (Section 3.2), enable one to take the investigation further by providing time-course profiles of the dynamics of individual cells over entire lineage trees. The development of models that take into account the genealogy of individual cells is an important step in the study of inheritance in bacterial population. As a prerequisite, the efficient analysis of single-cell data relies on the mathematical analysis of those models.

Structured branching processes allow for the study of populations, where the lifecycle of each cell is governed by a given characteristic or trait, such as the concentration of a specific protein inside the cell. The dependence of bacterial phenotypes like cell division times or ageing on such characteristics has been investigated by Aline Marguet using mathematical analysis of the underlying processes. To understand the long-time behavior of structured branching populations, the process describing the trait of a typical individual along its ancestral lineage, called auxiliary process [21] and its asymptotic behavior play a key role. In a publication in *ESAIM: Probability and Statistics* that appeared this year [20], we proved that the empirical measure of the structured branching process converges to the mean value of this auxiliary process. The approach relies on ergodicity arguments for the time-inhomogeneous auxiliary Markov process. The novelty compared to existing spectral methods is that our method allows to consider processes with time-varying rates for the modeling of changing environments. For example, we studied the case of a size-structured population in a varying environment and proved the convergence of the empirical measure in this specific case.

In collaboration with Charline Smadi (IRSTEA Grenoble), Aline Marguet also investigated the long-time behavior of a general class of branching Markov processes. This work, which has been submitted for publication [27], aims at understanding the link between the dynamic of the trait and the dynamic of the population. In the case of a trait modelling the proliferation of a parasite infection in a cellular population, we exhibit conditions on the dynamics of the parasites to survive in the population, despite the cellular divisions that dilute the number of parasites in each cell.

The study of the asymptotic behavior of general non-conservative semigroups is important for several aspects of branching processes, especially to prove the efficiency of statistical procedures. Vincent Bansaye from École Polytechnique, Bertrand Cloez from INRA Montpellier, Pierre Gabriel from Université Versailles Saint-Quentin, and Aline Marguet obtained necessary and sufficient conditions for uniform exponential contraction in weighted total variation norm of non-conservative semigroups. It ensures the existence of Perron eigenelements and provides quantitative estimates of spectral gaps, complementing Krein-Rutman theorems and generalizing recent results relying on probabilistic approaches. This work was submitted for publication this year [26].

6.4. Inference of gene expression parameters on lineage trees

As explained in the previous section, recent technological developments have made it possible to obtain time-course single-cell measurements of gene expression as well as the associated lineage information. However,

most of the existing methods for the identification of mathematical models of gene expression are not well-suited to single-cell data and make the simplifying assumptions that cells in a population are independent, thus ignoring cell lineages. The development of statistical tools taking into account the correlations between individual cells will allow in particular for the investigation of inheritance of traits in bacterial populations.

In the framework of structured branching processes, we studied the statistical reconstruction of parameters. We considered the problem of estimating the division rate from the observations of the trait of the cells at birth. Previous works on the subject considered deterministic dynamics for the evolution of the trait. In collaboration with Marc Hoffmann (Université Paris Dauphine), Aline Marguet investigated the case of a trait evolving according to a diffusion process. The study of the asymptotic behavior of the tagged-chain, corresponding to the trait of a uniformly chosen individual, allowed us to prove the convergence of the empirical measure of the branching process, and the asymptotic minmax efficiency of nonparametric estimators for the density of the transition kernel and the invariant measure of the tagged-chain. For the estimation of the division rate, we proved in a parametric framework the asymptotic efficiency of a standard maximum likelihood proxy estimation. Finally, we demonstrate the validity of our approach on simulated datasets. The results of this work were published in *Stochastic Processes and their Applications* [17].

Along the same lines, modelling and identification of gene expression models with mother-daughter inheritance are being investigated in the context of the ANR project MEMIP. Starting from an earlier work of the group [7], Eugenio Cinquemani, Marc Lavielle (XPOP, Inria Saclay–Île-de-France) and Aline Marguet developed a new model and a method for inference from data for gene expression along tree where the kinetic expression parameters are assumed to be inherited from the mother cell in an autoregressive way. This model generalizes the state-of-the-art mixed-effect models to the case of lineage trees. We implemented the inference procedure in Julia and proved that it provides unbiased estimates of the parameters. The application to the data of osmotic shock response by yeast show that the correlation between the parameter of a cell and its daughter is of 0.6 according to our model, leading to new biological questions such as the understanding of the origin of this inheritance. The results of this study were presented at the major bioinformatics conference ISMB/ECCB 2020 and published in the associated special issue of *Bioinformatics* [19].

6.5. Modeling and inference of RNA degradation

The ability to rapidly respond to changing nutrient availability is crucial for *E. coli* to survive in many environments including the gut. Reorganization of gene expression is the first step for bacteria to adjust their metabolism accordingly. It involves fine-tuning of both transcription and mRNA stability by dedicated regulatory interactions. While transcriptional regulation has been largely studied, the role of mRNA stability during a metabolic switch is poorly understood.

This question was addressed in the framework of the PhD thesis of Manon Morin funded by an INRA-Inria grant. Using combined genome-wide transcriptome and mRNA decay analyses, Manon Morin, Delphine Ropers and colleagues from the Toulouse Biotechnology Institute (ex-LISBP, INRA/INSA Toulouse) investigated the role of mRNA stability in the response of *E. coli* to nutrient changes. They demonstrated that transcript stability increases along metabolic transitions representative of the carbon source fluctuations, the glucose-acetate-starvation transition [9], [10]. Most of the stabilization occurs at glucose-acetate transition when glucose is exhausted. Stabilized mRNAs remain stable during acetate consumption and carbon starvation. Meanwhile, expression of most genes is downregulated. Metabolic control analysis showed that most of gene expression regulation is driven by changes in transcription. Post-transcriptional regulations appear to be important for genes involved in bacterial response to nutrient starvation. These results have been further developed in a paper recently submitted to a biology journal.

The observation of a global stabilization of cellular mRNAs during adaptation to carbon source depletion raises questions about the regulatory mechanisms at work. Known regulators of mRNA stability such as the protein Hfq, the carbon storage regulator Csr, and several small regulatory RNAs, specifically target mRNAs. Are these regulatory mechanisms sufficient to explain the systematic adjustment of mRNA half-lives? The collaboration with Muriel Cocaign-Bousquet and colleagues from the Toulouse Biotechnology Institute has been pursued to answer these questions, in the context of the PhD thesis of Thibault Etienne,

funded by an INRA-Inria PhD grant. The objective is to develop models able to explain how cells coordinate their physiology and the functioning of the degradation machinery following environmental changes. In a paper submitted this year, Thibault Etienne, Delphine Ropers and Muriel Cocaign-Bousquet investigate the possibility that competition between mRNAs for their binding to the degradation machinery is an important mechanism for the regulation of mRNA half-lives. They develop a mathematical model of mRNA degradation and assess the role of competitive effects on mRNA degradation kinetics by numerical simulation and sensitivity analysis. Competition appears to globally increase the stability of cellular mRNAs and to amplify the effect of post-transcriptional regulation. In a follow-up study, the model is currently being used to interpret large data sets corresponding to the degradation kinetics of 4254 mRNAs in *E. coli* cells growing in four different environmental conditions.

6.6. Growth control in bacteria and biotechnological applications

The ability to experimentally control the growth rate is crucial for studying bacterial physiology. It is also of central importance for applications in biotechnology, where often the goal is to limit or even arrest growth. Growth-arrested cells with a functional metabolism open the possibility to channel resources into the production of a desired metabolite, instead of wasting nutrients on biomass production. In recent years we obtained a foundation result for growth control in bacteria [6], in that we engineered an *E. coli* strain where the transcription of a key component of the gene expression machinery, RNA polymerase, is under the control of an inducible promoter. By changing the inducer concentration in the medium, we can adjust the RNA polymerase concentration and thereby switch bacterial growth between zero and the maximal growth rate supported by the medium. The publication also presented a biotechnological application of the synthetic growth switch in which both the wild-type *E. coli* strain and our modified strain were endowed with the capacity to produce glycerol when growing on glucose. Cells in which growth has been switched off continue to be metabolically active and harness the energy gain to produce glycerol at a twofold higher yield than in cells with natural control of RNA polymerase expression.

The experimental work underlying the growth switch has been continued in several directions in the context of the Maximic project by Célia Boyat. Moreover, in collaboration with colleagues from the BIOCORE project-team, we have formulated the maximization of metabolite production by means of the growth switch as a resource reallocation problem that can be analyzed by means of the self-replicator models of bacterial growth in combination with methods from optimal control theory. In a paper published in the *Journal of Mathematical Biology* this year [24], we study various optimal control problems by means of a combination of analytical and computational techniques. We show that the optimal solutions for biomass maximization and product maximization are very similar in the case of unlimited nutrient supply, but diverge when nutrients are limited. Moreover, external growth control overrides natural feedback growth control and leads to an optimal scheme consisting of a first phase of growth maximization followed by a second phase of product maximization. This two-phase scheme agrees with strategies that have been proposed in metabolic engineering. More generally, this work shows the potential of optimal control theory for better understanding and improving biotechnological production processes. Extensions concerning the effect on growth and bioproduction of the (biological or technological) costs associated with discontinuous control strategies, and of the time allotted to optimal substrate utilization, were presented at the European Control Conference (ECC 2019) in Naples this year and published in the proceedings [25].

6.7. Bacterial growth inhibition by acetate

High concentrations of organic acids such as acetate inhibit growth of *Escherichia coli* and other bacteria. This phenomenon is of interest for understanding bacterial physiology but is also of practical relevance. Growth inhibition by organic acids underlies food preservation and causes problems during high-density fermentation in biotechnology. The development of new approaches for the relief of growth inhibition by acetate during high-density fermentation of *E. coli* is one of the motivating assumptions for the work of IBIS in the IPL project COSY (Sections 7.2 and 6.8 below).

What causes growth inhibition by acetate? Classical explanations invoke the uncoupling effect of acetate and the establishment of an anion imbalance. During his PhD thesis, Stéphane Pinhal investigated an alternative hypothesis: the perturbation of acetate metabolism due to the inflow of excess acetate. In an experimental and modelling study published in the *Journal of Bacteriology* [23], Stéphane Pinhal, Delphine Ropers, Hans Geiselmann, and Hidde de Jong developed a set of isogenic strains that remove different parts of the metabolic network involved in acetate metabolism. Analysis of these strains revealed that the inflow of acetate accounts for 20% of the growth-inhibitory effect through a modification of the acetyl phosphate concentration. While the study does not provide a definite answer to the question of what accounts for the remaining 80% of the reduction in growth rate, some of the observations argue against a prominent role of uncoupling in growth inhibition by acetate in the conditions tested.

6.8. Modeling synthetic microbial communities for improving productivity

Modelling, analysis and control of microbial community dynamics is a fast-developing subject with great potential implications in the understanding of natural processes and the enhancement of biootechnological processes. Within the IPL COSY (Section 7.2), we picked up the challenge to design and investigate the dynamics of synthetically engineered microbial communities with a consortium of Inria partners. In IBIS, in particular, we are addressing the design of a bacterial community of two *E.coli strains*, mimicking mutualistic relationships found in nature, and with the potential to outperform a single producer strain in the production of a heterologous protein. During the post-doctoral stay of Marco Mauri, we developed an ODE model of the key growth phenotypes of the community and their interactions, calibrated the model on literature data, and analysed the model for an in-depth understanding of the conditions supporting coexistence and of the tradeoffs encountered in this production process. The results are presented in a paper submitted for publication this year and will be tested experimentally in the framework of the recently-started PhD project of Maaike Sangster. Analysis of optimal community control problems as well as design and deployment of optimal control strategies will follow in synergy with other IPL COSY partners.

6.9. Detection of small non-coding RNAs

Small non-coding RNAs (sRNAs) regulate numerous cellular processes in all domains of life. Several approaches have been developed to identify them from RNA-seq data, which are efficient for eukaryotic sRNAs but remain inaccurate for the longer and highly structured bacterial sRNAs. Together with colleagues from INSA de Lyon, Stéphan Lacour developed APERO, a new algorithm to detect small transcripts from paired-end bacterial RNA-seq data. This algorithm is based on a novel approach, which does not start from the read coverage distribution, but analyzes boundaries of individual sequenced fragments to infer the 5´ and 3´ ends of all transcripts. Validation of the algorithm on *Escherichia coli* and *Salmonella enterica* datasets, based on experimentally validated sRNAs, showed it to outperform all existing methods in terms of sRNA detection and boundary precision. Moreover, APERO was able to identify the small transcript repertoire of *Dickeya dadantii* including putative intergenic RNAs, 5´ UTR or 3´ UTR-derived RNA products and antisense RNAs. This work was published in *Nucleic Acids Research* this year [18]. APERO is freely available as an open source R package (https://github.com/Simon-Leonard/APERO). In other work, together with colleagues from the University of Salento, Lecce (Italy), Stéphan Lacour contributed to RHOTERMPREDICT, an algorithm for predicting Rho-dependent transcription terminators in bacterial genomes [16].

MOSAIC Project-Team

6. New Results

6.1. Dynamical characterization of morphogenesis at cellular scale

Participants: Guillaume Cerutti, Emmanuel Faure [External Collaborator], Christophe Godin, Anuradha Kar, Bruno Leggio, Jonathan Legrand, Patrick Lemaire [External Collaborator], Grégoire Malandain [External Collaborator], Florent Papini, Manuel Petit, Jan Traas [External Collaborator].

- Related Research Axes: RA1 (Representation of biological organisms and their forms in silico) & RA3 (Plasticity & robustness of forms)
- Related Key Modeling Challenges: KMC3 (Realistic integrated digital models)

The modeling of morphogenesis requires to explore the interconnection of different spatial and temporal scales of developing organisms. Non-trivial questions such as whether the observed robustness of morphogenesis is rooted in some highly conserved properties at the cellular level or whether it emerges as a macroscopic phenomenon, necessitate precise, quantitative analyses of complex 3D dynamic structures. The study of dynamical properties at the cellular scale poses at the same time key technical challenges and fundamental theoretical questions. An example of the former category is how to characterize and follow the change of shape of cells within tissues and of tissues within organs, and how to couple this change with, for instance, gene expression dynamics; an illustration of the latter is how to define cell-scale variability of morphogenesis within and between species.

Our team has produced this year several results in this context:

Cell-scale atlases of development. One fundamental question linked to morphogenesis is at which level and timescale tissue or organ development is reproducible and stereotyped. To answer this question, variability must be quantitatively assessed. In the team we have created to this end two morphogenetic atlases: the atlas of gene expression patterns in the Arabidopsis thaliana flower development and the atlas of early embryonic development of the ascidian Phallusia mammillata.

Thanks to the invariant cellular lineage of early development of *P. mammillata* embryos and to 3D reconstruction of their development at cellular resolution, quantitative comparison of their properties from cell to tissue scale has been performed. After fluorescent membrane labelling, several embryos have been imaged for several hours by light-sheet microscopy. These images were then reconstructed through the segmentation pipeline ASTEC, which also automatically tracked each cell over several rounds of cell division. This large amount of data allowed us to create an atlas of geometrical and topological properties at cellular resolution, which gives unprecedented depth of information on the variability of ascidian development. In addition this atlas, coupled to previous knowledge on gene-expression dynamics from the ascidian genetic database (ANISEED), made it possible for us to develop a mathematical and computational model to explore the main drivers of early ascidian development, identified as area-of-contact-mediated cell-cell communications. This model was also validated by experimental manipulations and mutations induced in ascidian embryos. This work is currently under review [26].

On the other hand, developing digital atlases of organism or organs development is a complex challenge for organisms presenting a strong variability in the cellular layout. Indeed contrary to C. Elegans or P. mammillata, for instance, that posses a very strict cell lineage in early phases, the development of most plant organs is under the influence of robust genetic patterns without a unique cellular layout. In that respect, proposing a cell-based atlas of flower development for instance is not straightforward and specific methods have been developed to choose a representative examples of the developing Arabidopsis thaliana flower. Using this representative flower we have generated an atlas in which we have introduced manually the expression patterns of 27 genes. The knowledge generated by the creation of this atlas makes it possible to have a first quantitative (correlative) view on the relation between gene activity and growth.

Robustness of ascidian embryonic development. The image segmentation pipeline ASTEC developed by the team in collaboration with the Inria Morpheme project-team in Sophia Antipolis and the CRBM team in Montpellier, allows the 3D reconstruction and tracking of each cell during early ascidian embryogenesis. This methods allowed us to reconstruct over 50 ascidian embryos, both wild-type and mutants. Exploiting this large database and the fixed cellular lineage of ascidian embryos, we extracted and compared geometrical and topological cellular properties. This allowed us to compare the intra-embryonic (left/right) to the inter-embryonic level of variability of several properties, including cell volume, cell-cell contacts and the structure of the tree seeded by each cell. This study demonstrated that the genetic-induced variability is comparable to the stochastic one, quantitatively showing that ascidian embryonic development is highly canalized, and that the high reproducibility of shapes observed during embryogenesis is rooted in the robustness of cellular geometry and topology. To look for the origin of this canalisation, we developed a mathematical model exploiting our quantitative geometric database and the previously-existing ascidian genetic database ANISEED. This model suggests that the main driver of ascidian development is the cell-cell communication mediated by direct physical contact, and hence dependent of the area-of-contact between neighbouring cells. This means that the robustness of cell topology and geometry is necessary for cell-cell biochemical interactions to give rise to the correct fate restriction events, which in turn we showed to be responsible for major changes of embryo geometry. We also tested and validated this feedback loop between cell contacts, fate restriction events and embryonic geometry predicted by the model by manipulations and mutations induced in ascidian embryos.These results are reported in a paper which is currently under review [26].

Robust extraction and characterization of cellular lineages. The quantification of temporal properties at cellular scale such as volumetric growth rate or strain patterns relies extensively on the identification of cellular lineages in time-lapse acquisitions of living tissues. In the case of plant tissues where the deformations between two consecutive time points can be very important in post-embryonic morphogenesis processes such as early flower development, it remains a real challenge to compute those lineages automatically, and manual user annotation is generally required to produce reliable results.

Building on the previous expertise of the team [25], [28] and on the state-of-the-art computational library for image analysis, timagetk, developed in collaboration with the Morpheme team, we currently develop a set of robust automatic cell lineaging methods for cases ranging from small to highly non-linear deformations. In the course of a M2 internship and the first months of a starting PhD work (Manuel Petit), a first so-called “naive” lineaging method has been implemented and validated on synthetic data with limited deformations. Methods involving optimal flow algorithms on graph structures and iterative image registration are being developed to provide robust results in the case of faster growing tissues. The output of these methods will allow to use the tools developed by the team for the analysis of spatio-temporal properties of growing cells at a much larger scale. This work is part of the Inria IPL Naviscope.

Reconstruction of Arabidopsis ovule development. The ovule is a relatively simple organ, with limited developmental variability, which makes it an excellent case study for the computational modeling of organ development. Given the technical difficulty of producing live-imaging acquisition sequences of ovules, we developed a method to perform a spatial registration of multiple individual ovules at various developmental states and in different global poses. Using the global cylindrical symmetry of the organ and the surface curvature as a key geometrical feature, we aligned individuals on their main axes and on their junction with the underlying placental tissue. Jointly with the 3D segmentation of cells in images, this will allow to evidence the invariant features of ovule development at cellular scale, and to study the robustness of the dynamics of the megaspore mother cell (MMC) across indivduals. This work was part of the Imago project.

6.2. Reconstruction of macroscopic forms from images and characterization of their variability

Participants: Ayan Chaudhury, Christophe Godin, Jonathan Legrand, Katia Mirande.

- Related Research Axes: RA1 (Representations of forms in silico) & RA3 (Plasticity & robustness of forms)
- Related Key Modeling Challenges: KMC3 (Realistic integrated digital models)

To study the variability of macroscopic forms resulting from development, it is necessary to both develop digital reconstruction methods, typically based on image acquisitions, and statistical tools to define notions of distance or average between these forms. The automatic inference of computational representations of forms or organ traits from images of different types is therefore an essential step, for which the use of prior knowledge can be very beneficial. Realistic synthetic models of forms can guide the reconstruction algorithms and/or assess their performances. Computational representations of forms can then be used to analyze how forms vary at the scale of a population, of a species or between species, with potential applications in species identification and genetic or environmental robustness estimation.

Automatized characterization of 3D plant architecture. The digital reconstruction of branching and organ forms and the quantification of phenotypic traits (lengths of internodes, angles between organs, leaf shapes) is of great interest for the analysis of plant morphology at population scale. In collaboration with the ROMI partners from Sony CSL, Paris, we develop an automated processing pipeline that involves the 3D reconstruction of plant architecture from RGB image acquisitions performed by a robot, and the segmentation of the reconstructed plant into organs. We aim at releasing both hardware schematics and the developed software for image reconstruction to be used as cheap open-source solution to phenotype plants. In addition, to provide validation data for the pipeline, we designed a generative model of *Arabidopsis thaliana* simulating the development of the plant architecture at organ scale. This model was used to develop the method for the measurement of angles of organs and test its accuracy:

- RGB images were generated from the model and used as input of the pipeline;
- a physical version of the model has been obtained using 3D printing techniques;

In both cases, knowing the generated phenotypic traits or the model shape allow to test the pipeline ability to reconstruct the plant and quantify its traits of interest

The developed reconstruction and quantification pipeline is not made from scratch but aggregate a number of available third party libraries and codes in addition to three active research topics: spectral clustering, skeleton extraction, and ML segmentation. In a second phase, the model will be used to generate training data for machine learning techniques introduced in the reconstruction methods. This work is part of the *ROMI* project.

6.3. Analysis of tree data

Participants: Romain Azaïs, Christophe Godin, Salah Eddine Habibeche [External Collaborator], Florian Ingels.

- Related Research Axes: RW1 (Representations of forms in silico)
- Related Key Modeling Challenges: KMC1 (A new paradigm for modeling tree structures in biology)

Tree-structured data naturally appear at different scales and in various fields of biology where plants and blood vessels may be described by trees. In the team, we aim to investigate a new paradigm for modeling tree structures in biology in particular to solve complex problems related to the representation of biological organisms and their forms in silico.

In 2019, we investigated the following questions linked to the analysis of tree data. (i) How to control the complexity of the algorithms used to solve queries on tree structures? For example, computing the edit distance matrix of a dataset of large trees is numerically expensive. (ii) How to estimate the parameters within a stochastic model of trees? And finally, (iii) how to develop statistical learning algorithms adapted to tree data? In general, trees do not admit a Euclidean representation, while most of classification algorithms are only adapted to Euclidean data. Consequently, we need to study methods that are specific to tree data.

Approximation of trees by self-nested trees. Complex queries on tree structures (e.g., computation of edit distance, finding common substructures, compression) are required to handle tree objects. A critical question is to control the complexity of the algorithms implemented to solve these queries. One way to address this issue is to approximate the original trees by simplified structures that achieve good algorithmic properties. One can expect good algorithmic properties from structures that present a high level of redundancy in their substructures. Indeed, one can take into account these repetitions to avoid redundant computations on the whole structure. In the team, we think that the class of self-nested trees, that are the most compressed trees by DAG compression scheme, is a good candidate to be such an approximation class.

In [11], we have proved the algorithmic efficiency of self-nested trees through different questions (compression, evaluation of recursive functions, evaluation of edit distance) and studied their combinatorics. In particular, we have established that self-nested trees are roughly exponentially less frequent than general trees. This combinatorics can be an asset in exhaustive search problems. Nevertheless, this result also says that one can not always take advantage of the remarkable algorithmic properties of self-nested trees when working with general trees. Consequently, our aim is to investigate how general trees can be approximated by simplified trees in the class of self-nested trees from both theoretical and numerical perspectives. In [3], we present two approximation algorithms that are optimal but assume that the approximation can be obtained by only adding vertices to the initial data (or by only deleting vertices from the initial data). In [11], we have developed a suboptimal approximation algorithm based on the height profile of a tree that can be used to very rapidly predict the edit distance between two trees, which is a usual but costly operation for comparing tree data in computational biology. Another algorithm based on the efficient simulation of conditioned random walks on the space of trees is currently under development. This work should result in the submission of a paper next year.

It should be noted that the aforementioned strategy and algorithms can only be applied to topological trees. In 2019, we also began a new project on approximation of trees with geometrical attributes on their vertices and with possibly a controlled loss of information during the compression.

Statistical inference. The main objective of statistical inference is to retrieve the unknown parameters of a stochastic model from observations. A Galton-Watson tree is the genealogical tree of a population starting from one initial ancestor in which each individual gives birth to a random number of children according to the same probability distribution, independently of each other. In a recent work [5], we have focused on Galton-Watson trees conditional on their number of nodes. Several main classes of random trees can be seen as conditioned Galton-Watson trees. For instance, an ordered tree picked uniformly at random in the set of all ordered trees of a given size is a conditioned Galton-Watson tree with offspring distribution the geometric law with parameter 1/2. Statistical methods were developed for conditioned Galton-Watson trees in [5]. We have introduced new estimators and stated their consistency. Our techniques improve the existing results both theoretically and numerically.

We continue to explore these questions for subcritical but surviving Galton-Watson trees. The conditioning is a source of bias that must be taken into account to build efficient estimators of the birth distribution. This work should be submitted to a journal next year.

Kernel methods for tree data. Standard statistical techniques – such as SVMs for supervised learning – are usually designed to process Euclidean data. However, trees are typically non-Euclidean, thus preventing using these methods. Kernel methods allow this problem to be overcome by mapping trees in Hilbert spaces. However, the choice of kernel determines the feature space obtained, and thus greatly influences the performance of the different statistical algorithms. Our work is therefore focused on the question of how to build a good kernel.

We first looked in [17] at a kernel of the literature, the subtree kernel, and showed that the choice of the weight function – arbitrarily fixed so far – was crucial for prediction problems. By proposing a new framework to calculate this kernel, based on the DAG compression of trees, we were able to propose a new weight, learned from the data. In particular, on 8 data sets, we have empirically shown that this new weight improves prediction error in 7 cases, and with a relative improvement of more than 50% in 4 of these cases. This work was presented at a national conference [15].

We then tried to generalize our framework by proposing a kernel that is no longer based on subtrees, but on more general structures. To this end, we have developed an algorithm for the exhaustive enumeration of such structures, namely the forest of subtrees with a uniform fringe. This work will be submitted for pre-publication early in the coming year.

6.4. Mechanics of tissue morphogenesis

Participants: Olivier Ali, Arezki Boudaoud [External Collaborator], Guillaume Cerutti, Ibrahim Cheddadi [External Collaborator], Florian Gacon, Christophe Godin, Bruno Leggio, Jonathan Legrand, Hadrien Oliveri, Jan Traas [External Collaborator].

- Related Research Works: RW2 (*Data-driven models*) & RW3 (*Plasticity & robustness of forms*)
- Related Key Modeling Challenges: KMC2 (*Efficient computational mechanical models of growing tissues*) & KMC3 (*Realistic integrated digital models*)

As deformations supporting morphogenesis require the production of mechanical work within tissues, the ability to simulate accurately the mechanical behavior of growing living tissues is a critical issue of the MOSAIC project. From a macroscopic perspective, tissues mechanics can be formalized through the framework of continuum mechanics. However, the fact that they are composed, at the microscopic level, by active building blocks out of equilibrium (namely cells) offers genuine modeling challenges and opportunities. Integrating cellular behaviors such as mechano-sensitivity, intercellular fluxes of materials and cell division into a macroscopic mechanical picture of morphogenesis is the topic of this section.

Flattening mechanism during organogenesis in plants. Many plant species have thin leaf blades and axisymmetric elongating organs, such as stems and roots. From a morphoelastic perspective, such complex shapes are currently believed to emerge from the coordination between strain-based growth and stress-based stiffening at the cellular level.

To study the plausibility of such an hypothesis, we conducted numerical simulations where both a stress-based stiffening mechanism of cell walls [29] and a strain-based growth mechanism [24] have been implemented. We performed such simulations on multicellular and multilayered ellipsoidal structures and track their aspect ratio as they developed under various parametrization sets. One key aspect we wanted to investigate was the effect of an heterogeneous stress-based stiffening mechanism on the overall dynamics: Starting from a given initial shape, can we get significantly different shapes by assuming the stress-based stiffening mechanism active only in specific parts of the structures?

Our results, in accordance with experimental measurements conducted simultaneously by biologist colleagues, showed that: (i) Stress-based stiffening was mandatory to grow flat and axisymmetric organs; (ii) in order to grow flat structures, stress-based stiffening should only be active on anticlinal inner walls.

This work was part of Jan Traas's ERC grant *Morphodynamics*. This work is currently under review, see preprint version [23].

Influence of cell division during flat organogenesis in plants. One key limitation of our 3D modeling approach of leaf-like organogenesis is the lack of cell division implementation. This can be seen as a major flaw in the mechanical understanding of flattening since cell divisions, by increasing the number of load bearing walls, impact significantly the redistribution of mechanical stresses within the tissue.

To alleviate this limitation, we developed a 2D modeling approach to complement the 3D one. This 2D model encompasses the same biophysical processes as the 3D one (described in the previous subsection): a stress-based stiffening and a strain-based growth mechanisms of cell walls; augmented with a cell division module. We used this 2D framework to investigate the flattening dynamics of structures mimicking ellipsoid cross sections of growing organs. Such cross section were described as vertex-based, multicellular and multilayered structures.

We first reproduced the results obtained with the 3D approach to ensure that both models agreed on similar situations, where no cell division was implemented. We tested then several rules of cell division orientation and check which one(s) produced the most efficient flattening process. We were able to show that heterogeneity in the division rule between the epidermis and the inner tissues led to the more efficient flattening process and that a stress-based division rule was the most efficient to produce flat structure.

This analysis is part of the manuscript currently under review and available online in a preprint version [23].

Influence of mechanical stress anisotropy on the orientation of cell divisions in animal tissues. Tight regulation of cell division orientation is fundamental for tissue development. Recently, a great effort has been put into biophysical understanding of the *long-axis* division rules (Hertwig's rule for animal cells, Errera's rule for plant cells) and the systematic deviations from these rules observed *in vivo*. In both plants and animals, such deviations often correlate with anisotropic tensions within the tissue. To what extend these deviations are regulated or simply the result of stochasticity?

To address these questions in animal cells, we modeled theoretically and numerically cell division as an active process in a many-body system. We showed that under isotropic tension a cell's long axis emerges as the energetically optimal division orientation and that anisotropic stresses biased the energetics, leading to systematic deviations form Hertwig's rule. These deviations, as reported experimentally, are correlated to the main direction of stress anisotropy.

Our model successfully predicted division orientation distributions within two experimental systems: epidermis of the ascidian *Phallusia mammillata* (where deviations from Hertwig's rule have been so far eluding explanation) and of the pupal epithelium of the dorsal thorax of *D. melanogaster*.

This work was part of the *Digem* project and was presented in two international conferences: *Mechanobiology and Physics of Life* (Lyon) and *Developmental and Cell Biology of the Future* (Paris); and at the yearly *InriaBio* meeting in Lyon. A paper is currently under review and a preprint is available on bioRxiv [22].

Influence of water fluxes on plant morphogenesis. Since pressure appears as the "engine" behind growth-related deformation in Plants, its regulation by cells is a major control mechanism of morphogenesis. We developed 2D computational models to investigate the morphological consequences of the interplay between cell expansion, water fluxes between cells and tissue mechanics. This interdisciplinary work, between experiments and modeling, address the influence of turgor pressure heterogeneities on relative growth rate between cells. We showed that the coupling between fluxes and mechanics allows to predict observed morphological heterogeneities without any *ad hoc* assumption.

This work was part of the Agropolis fundation project *MecaFruit3D* and Arezki Boudaoud's ERC *PhyMorph*. It resulted in a publication in PLoS Computational Biology [7] that introduces the theoretical model and studies some of its properties. Another paper [27] presents the comparisons with experiments and is currently under review.

Development of *de novo* finite element (F.E.) library dedicated to mechanical simulations performed on complex cellularized structures. In order to compute accurately the mechanical stress field borne by multicellular pressurized 3D structures (such as plant tissues), we needed to update our existing library (*tissueMeca*, see [24]). Three key aspects had to be upgraded (i) the control over the F.E. solver, (ii) tracking of its precision and (iii) integration of the F.E. framework with the rest of our pipeline.

To that end, we decided to switch from *Sofa* to *FEniCS* (https://fenicsproject.org/) as the core F.E. framework used within our simulation pipeline. We started to develop a dedicated library, called *CellFem*, to solve F.E. problems on *PropertyTopomesh* instances (the data structure we developed within the team to describe multicellular plant tissues). *CellFem* provides a high level API to define and resolve variational problems to solve linear as well as non-linear elastic and elasto-plastic problems related to plant tissue morphogenesis.

In parallel, we also started the development of a meshing library (based on the GMSH library (http://gmsh.info/)) called *CellMesh* and dedicated to the triangulation of simplicial complexes. This work is currently under development.

6.5. Signaling and transport for tissue patterning

Participants: Romain Azaïs, Guillaume Cerutti, Christophe Godin, Bruno Leggio, Jonathan Legrand, Teva Vernoux [External Collaborator].

- Related Research Axes: RA1 (Representations of forms in silico) & RA2 (Data-driven models)
- Related Key Modeling Challenges: KMC3 (Realistic integrated digital models)

One central mechanism in the shaping of biological forms is the definition of regions with different genetic identities or physiological properties through bio-chemical processes operating at cellular level. Such patterning of the tissue is often controlled by the action of molecular signals for which active or passive transport mechanisms determine the spatial precision of the targeting. The shoot apical meristem (SAM) of flowering plants is a remarkable example of such finely controlled system where the dynamic interplay between the hormone auxin and the polarization of efflux carriers PIN1 govern the rhythmic patterning of organs, and the consequent emergence of phyllotaxis.

Using *Arabidopsis thaliana* as a model system, we develop an integrated view of the meristem as a self-organizing dynamical form by reconstructing the dynamics of physiological processes from living tissues, and by proposing computational models integrating transport and signaling to study tissue patterning *in silico*.

Automatic quantification of auxin transport polarities. Time-lapse imaging of living SAM tissues marked with various fluorescent proteins allows monitoring the dynamics of cell-level molecular processes. Using a co-visualization of functional fluorescent auxin transporter (PIN1-GFP) with a dye staining of cell walls with propidium iodide (PI), we developed an original method to quantify in 3D the polarization of auxin transport for every anticlinal wall of the first layer of cells in confocal images. The developed method [13] was thoroughly evaluated against super-resolution acquisitions of the same tissue obtained using radial fluctuations (SRRF), and show to provide highly consistent results (less than 10% incorrect polarities, 80% of cells with a polarity vector error lesser than 30°). The digitally reconstructed networks evidenced an overall stable convergence of PIN1 polarities towards the center of the meristem, with a local convergence and divergence pattern that could explain the dynamics of auxin distributions in the meristem [19].

Landmark-based registration for the averaging of meristem patterning. To perform statistics of meristem patterning at the scale of a population, we developed a series of tools to compute a rigid 3D transformation that registers any individual meristem into a common cylindrical reference frame in which point-wise comparison is meaningful. The orignal method relies on the identification of biological landmarks (apex and main symmetry axis of the meristematic dome, position of the lastly emerged organ primordium and direction of the phyllotactic spiral) to compute this transform. These landmarks can be extracted from image acquisitions of meristems carrying the right fluorescent bio-markers (*CLV3* central zone marker for the apex, *DIIV* auxin bio-sensor for the organ primordia) using an original method that relies on the computation of 2D continuous maps of epidermal signal from discrete point clouds. The use of this registration method allowed to evidence key features of the transcriptional response of mersitematic cells to auxin [19].

In a second time, we aim to generalize the method to images without specific bio-markers, using only the geometry of the tissue to identify the relevant landmarks. To do so, machine learning approaches making use of the data processed for [19] are being developed and evaluated. This new landmark-based registration method would drastically improve the ability of comparing different individual meristems, open the way to spatial statistics over of multiple genetic and molecular signals, and contribute to an integrated tissue-level view of meristem patterning.

Computational models of integrated transport and signaling. Guided by new discoveries on auxin patterning dynamics in the shoot apical meristem (SAM) of *A. thaliana*, we developed a theoretical model of active and passive auxin transport. This model, built on existing view of auxin active transport [30], [31], naturally integrates the role of deeper cellular layers in the SAM and the mutual feedbacks between different components of the auxin-transport machinery. Through numerical simulation, the consequences of competing theories on PIN polarisation mechanism on auxin dynamics were explored. These results will serve, in quantitative comparisons with *in vivo* observation, to validate hypotheses on molecular mechanisms of auxin transport and to provide information on the role of memory effects and information fluxes during patterning.

These works were part of the *BioSensors* HFSP project and are carried out in the *Phyllo* ENS-Lyon project. These works gave rise to a journal article which is currently under review and have been partly presented at the *International Worskhop on Image Analysis Methods for the Plant Sciences* in Bron in July 2019.

6.6. Regulation of branching mechanisms in plants

Participants: Romain Azaïs, Frédéric Boudon [External Collaborator], Christophe Godin.

- Research Axes: RA2 (*Data-driven models*) & RA3 (*Plasticity & robustness of forms*)
- Key Modelling Challenges: KMC3 (*Realistic integrated digital models*)

Branching in plants results from the development of apical meristems that recursively produce lateral meristems. These meristems may be more or less differentiated with respect to the apical meristem from which they originate, potentially leading to different types of lateral branches or organs. They also can undergo a more or less long period of inactivation, due to systemic regulation. The understanding of branching systems morphogenesis in plants thus relies on the analysis of the regulatory mechanisms that control both meristem differentiation and activation/inactivation.

Analysis of the diversity of inflorescence architecture in different rice species. Rice is a major cereal for world food security and understanding the genetic and environmental determinants of its branching habits is a timely scientific challenge. The domestication, i.e., the empirical selection by humans, of rice began 10 000 years ago in Asia and 3 000 years ago in Africa. It thus provides a short-term model of the processes of evolution of plants.

Hélène Adam and Stéphane Jouannic from the group Evo-Devo de l'Inflorescence of UMR DIADE at IRD (Montpellier) have collected for years on the different continents an outstanding database of panicle-type inflorescence phenotypes in Asian and African, cultivated and wild, rice species. Classical statistical analysis based on the extraction of characteristic traits for each individual branching system were able to separate wild species from cultivated ones, but could not discriminate between wild species, suggesting that the entire branching structure should be used for classification methods to operate. For this, we are currently developing statistical methods on tree structures (see section 6.3) that should allow us to achieve better discrimination between panicles, based on their branching topology in addition to geometric traits. By coupling the quantitative study of the panicles to genomic analyses carried out by the IRD group, we should be able to highlight which regulation pathways have been selected or altered during the domestication process.

The role of sugars in apical dominance. The outgrowth of axillary buds is a key process in plant branching and which is often shown to be suppressed by the presence of auxin in nodal stems. However, local auxin levels are not always sufficient to explain bud outgrowth inhibition. Recent studies have also identified a contribution of sugar deprivation to this phenomenon. Whether sugars act independently of auxin or other hormones auxin regulates is unknown. Auxin has been shown to induce a decrease of cytokinin levels and to upregulate strigolactone biosynthesis in nodes. Based on rose and pea experiments, both in vitro and in planta, with our collaborators Jessica Bertheloot, Soulaiman Sakr from Institut de Recherche en Horticulture et Semences (IRHS) in Angers, we have shown that sucrose and auxin act antagonistically, dose-dependently, and non-linearly to modulate bud outgrowth. The Angers group provided experimental evidence that sucrose represses bud response to strigolactones but does not markedly affect the action of auxin on cytokinin levels. Using a modeling approach, we tested the ability of this complex regulatory network to explain the observed phenotypes. The computational model can account for various combinations of sucrose and hormones on bud outgrowth in a quantitative manner and makes it possible to express bud outgrowth delay as a simple function of auxin and sucrose levels in the stem. These results provide a simple auxin-sucrose-cytokinin-strigolactone network that accounts for plant adaptation to growing conditions [6] and [10] for a review.

The fractal nature of plants. Inflorescence branching systems are complex and diverse. They result from the interaction between meristem growth and gene regulatory networks that control the flowering transition during morphogenesis. To study these systems, we focused on cauliflower mutants, in which the meristem repeatedly fails in making a complete transition to the flower and for which a complete mechanistic explanation is still lacking.

In collaboration with Eugenio Azpeitia and François Parcy's group in Grenoble, we have developed a first model of the control of floral initiation by genes, refining previous networks from the literature so that they can integrate our hypotheses about the emergence of cauliflower phenotypes. The complete network was validated by multiple analyses, including sensitivity analyses, stable state analysis, mutant analysis, among others. It was then coupled with an architectural model of plant development using L-systems. The coupled model was

used to study how changes in gene dynamics and expression could impact in different ways the architectural properties of plants. The model was then used to study how changes in certain parameters could generate different curd morphologies, including the normal and the fractal-like Romanesco. A paper reporting this work is currently being written.

6.7. Miscellaneous

Participants: Romain Azaïs, Christophe Godin, Bruno Leggio.

Measurements and nonlocal correlations in quantum mechanics. Based on a long standing collaboration between Christophe Godin and Przemyslaw Prusinkiewicz from the University of Calgary on the analysis of connections between computer simulation paradigms and quantum mechanics, we theoretically investigated with the quantum mechanics expertize of Bruno Leggio in the team effects of measurements on quantum systems, mostly in connection with quantum non-locality and entanglement. At the same time, we exploit formal and conceptual analogies between quantum theory and biologically-inspired structures to study the latter under new paradigms.

One fruitful line of research deals with the inherent non-locality of correlations between measurement outcomes, characterizing the quantum world. These phenomena are described by the celebrated Bell inequalities. We study ways to generalize such inequalities to better capture non-local correlations, at the same time shedding light on the origin of the discrepancy between quantum and classical stochasticity. In parallel, we develop and profit from formal analogies between the theory of non-locality and the exploration of fractal structures in the context of simulation of arborescent systems.

Another research line sees the application of parameter-estimation techniques for piecewise deterministic Markovian processes (PDMP), developed by members of the team, to the special case of quantum dynamics: under certain conditions, the evolution of an open quantum system can be described as a PDMP, with a specific and non-trivial structure marking its departure from classical behaviour. We show [21] that approaches to appraise parameter values of the evolving systems, developed in the context of classical dynamics, can be successfully applied to the specific case of quantum systems.

Finally, a third research topic consists of the study of the structure of typical quantum correlations, called entanglement, and its relation to thermal noise induced in a quantum system by its unavoidable interaction with its surrounding environment. We show [9] that the quantitative amount of noise represents a tight upper bound on the amount of bipartite quantum correlation two systems can establish between them.

Statistical analysis and stochastic modelling of penguin diving. The activity at sea of penguins can be reconstructed from measurement devices equipped on the animals during their trips. We study the relative behavior of the time under water with respect to the time spent at the surface from a dataset of about 100 thousands dives of little penguins. We show that dives that form a bout in which the penguin explores a patch of preys show a type of stationarity. We have built a mathematical model of sequences of dives that can be optimized in terms of number of preys catched by the animal under physiological constraints. This reproduces the stationary behavior observed in the data.

NUMED Project-Team (section vide)

STEEP Project-Team

6. New Results

6.1. Analysis of socio-ecological dimensions of human activities – A case study of Beaufort cheese production in the Maurienne Valley

The PhD thesis of Michela Bevione aims at analysing socio-ecological dimensions of human activities creating wealth by coupling quantitative-biophysical approaches and qualitative and socio-economic methodologies to assess territorial metabolism. By focusing on the interactions between flows and actors, the methodology we propose aims at providing a methodological framework for the understanding of a territory and its capability.

As a case study for this thesis, we chose to focus on the production of the AOC-labelled cheese Beaufort in the Maurienne Valley (Savoie department, Auvergne-Rhône-Alpes region, France). Indeed, agriculture play a structuring role for the economic and social dynamics of the valley, and the landscape construction induced by farming activities contributes to create favourable conditions to the development of the touristic sector. Beaufort represents the flagship product of the agricultural sector in the valley and most of farms are dedicated to milk production for the Beaufort industry.

In [7] we represent the circulation of material flows through flow maps, showing the movement of material and monetary resources and products, their direction, source and destination. We focus on the circulation of flows related to the Beaufort industry within the Maurienne Valley and between the valley and other territories. Through Sankey diagrams (a specific kind of flow maps, where the width of the arrows is proportional to the flow quantity) we present the dominant contributions to the overall material flows circulation. This kind of representation is appropriate to characterise the circulation of material flows, the allocation of environmental pressures throughout the Beaufort industry, as well as the monetary dimension and the added value associated to Beaufort production. Mapping the geographical origin of input resources and the destination of output products and incomes allows to evaluate actors' capacity to create wealth through the activation and mobilisation of local resources and/or their dependence on foreign inputs.

Furthermore, results include schematic representations of the relations between local, extra-territorial actors and the circulation of material, environmental and monetary flows. The influence of immaterial resources (informational flows and traditional savoir-faire) and local infrastructures on the circulation of flows, and vice versa, is illustrated. Finally, positive and negative retroactions induced by output products on input resources for Beaufort production are drawn, as well as the interactions with other sub-systems creating wealth in the valley.

6.2. Sensitivity analysis of World3

World3 is a computer tool created to simulate the interactions between the world population, industrial growth and food production within the limits of the planet. It aims to highlight the problems posed by indefinite material growth in a finite world. The first version of this tool was proposed in 1972 by MIT researchers for the first report to the Club of Rome [19]. This report was both highly successful and polemic. The main detractors of the model criticized it for being too approximate in the choice of the parameters and for beeing too simplistic. We started to work on revisiting some aspects of the scientific validation of this model, in a context where growth is widely debated in the scientific and civil community, through a new and more sophisticated sensitivity analysis of the model, compared to what is available in the literature [11].

6.3. Efficient computation of solution space and conflicts detection for linear systems

Our work on Material Flow Analysis (see e.g. the *AF Filières* project), involves the analysis of systems of linear inequalities, $l \leq Ax \leq u$. There are three different but complementary goals for the analysis: (i) given some known variables x_i, efficiently compute the solution space of unknown variables, (ii) if the set of constraints is infeasible, efficiently identify the conflicts, (iii) efficiently classify variables to determine whether they are redundant, just measured, determinable or non-determinable. A baseline implementation for these tasks was available in the team but proved to be too inefficient for larger problem sizes. Through the internship of Alexandre Borthomieu we worked on various improvements, on the algorithmic and implementation side (e.g. choice of programming language), that eventually led to a reduction of execution by three orders of magnitude, compared to the previous implementation.

6.4. Mapping ecosystem services bundles in a heterogeneous mountain region

2019 was the final year of production of the ESNET project (which officially ended in 2017). This and the following section describe our two most complex pieces of work in that project.

Recent institutional and policy frameworks prescribe the incorporation of ecosystem services (ES) into land use management and planning, favouring co-production of ES assessments by stakeholders, land planners and scientists. Incorporating ES into land management and planning requires models to map and analyze ES. Also, because ES do not vary independently, many operational issues ultimately relate to the mitigation of ES trade-offs, so that multiple ES and their interactions need to be considered. Using a highly accurate LULC (Land Use Land Cover) database for the Grenoble urban region (French Alps), we mapped twelve ES using a range of models of varied complexity [5]. A specific, fine-grained (less than 1 ha) LULC database at regional scale (4450 km^2) added great spatial precision in individual ES models, in spite of limits of the typological resolution for forests and semi-natural areas. We analysed ES bundles within three different socio-ecosystems and associated landscape types (periurban, rural and forest areas). Such type-specific bundles highlighted distinctive ES trade-offs and synergies for each landscape. Advanced approaches combining remote sensing, targeted field data collection and expert knowledge from scientists and stakeholders are expected to provide the significant progress that is now required to support the reduction of trade-offs and enhance synergies between management objectives.

6.5. Co-constructing future land-use scenarios for the Grenoble region, France

Physically and socially heterogeneous mountain landscapes support high biodiversity and multiple ecosystem services. But rapid landscape transformation from fast urbanisation and agricultural intensification around cities to abandonment and depopulation in higher and more remote districts, raises urgent environmental and planning issues. For anticipating their future in a highly uncertain socio-economic context, we engaged stakeholders of a dynamic urban region of the French Alps in an exemplary interactive participatory scenario planning (PSP) for co-creating salient, credible and legitimate scenarios. Stakeholders helped researchers adapt, downscale and spatialize four normative visions from the regional government, co-producing four storylines of trend versus break-away futures. Stakeholder input, combined with planning documents and analyses of recent dynamics, enabled parameterisation of high-resolution models of urban expansion, agriculture and forest dynamics. With similar storylines in spite of stakeholders insisting on different governance arrangements, both trend scenarios met current local and European planning objectives of containing urban expansion and limiting loss and fragmentation of agricultural land. Both break-away scenarios induced considerable conversion from agriculture to forest, but with highly distinctive patterns. Under a commonly investigated, deregulated liberal economic context, encroachment was random and patchy across valleys and mountains. A novel reinforced nature protection scenario affecting primarily mountain and hilly areas fostered deliberate consolidation of forested areas and connectivity. This transdisciplinary approach demonstrated the potential of combining downscaled normative scenarios with local, spatially-precise dynamics informed by stakeholders for local appropriation of topdown visions, and for supporting land planning and subsequent assessment of ecosystem service trade-offs. This work is described in [4].

AGORA Project-Team

7. New Results

7.1. Wireless network deployment

Participants: Walid Bechkit, Ahmed Boubrima, Oana Iova, Rodrigue D. Komguem, Abdoul-Aziz Mbacke, Jad Oueis, Hervé Rivano, Razvan Stanica, Fabrice Valois

7.1.1. Deployment of wireless sensor networks for air quality mapping

Wireless sensor networks (WSN) are widely used in environmental applications where the aim is to sense a physical phenomenon such as temperature, air pollution, etc. A careful deployment of sensors is necessary in order to get a better knowledge of these physical phenomena while ensuring the minimum deployment cost [18]. In this work, we focus on using WSN for air pollution mapping and tackle the optimization problem of sensor deployment [3]. Unlike most of the existing deployment approaches, which are either generic or assume that sensors have a given detection range, we define an appropriate coverage formulation based on an interpolation formula that is adapted to the characteristics of air pollution sensing. We derive from this formulation two deployment models for air pollution mapping using integer linear programming while ensuring the connectivity of the network and taking into account the sensing error of nodes. We analyze the theoretical complexity of our models and propose heuristic algorithms based on linear programming relaxation and binary search. We perform extensive simulations on a dataset of the Lyon city, France in order to assess the computational complexity of our proposal and evaluate the impact of the deployment requirements on the obtained results.

7.1.2. Characterization of radio links in case of a ground deployment

In this work, we are interested in characterizing the link properties of a wireless sensor network with nodes deployed at ground level [5]. Such a deployment is fairly common in practice, e.g., when monitoring the vehicular traffic on a road segment or the status of infrastructures such as bridges, tunnels or dams. However, the behavior of off-the-shelf wireless sensor nodes in these settings is not yet completely understood. Through a thorough experimentation campaign, we evaluated not only the impact of the ground proximity on the wireless links, but also the impact of some parameters such as the packet payload, the communication channel frequency and the topography of the deployment area. Our results show that a ground-level deployment has a significant negative impact on the link quality, while parameters such as the packet size produce unexpected consequences. This allows us to parameterize classical theoretical models in order to fit a ground-level deployment scenario. Finally, based on the lessons learned in our field tests, we discuss some considerations that must be taken into account during the design of communication protocols and before the sensor deployment in order to improve network performance.

7.1.3. Sensor deployment in linear wireless sensor networks using the concept of virtual node

In a multi-hop wireless sensor network with a convergecast communication model, there is a high traffic accumulation in the neighborhood of the sink. This area constitutes the bottleneck of the network since the sensors deployed within it rapidly exhaust their batteries. In this work, we consider the problem of sensors deployment for lifetime maximization in a linear wireless sensor network [6]. Existing approaches express the deployment recommendations in terms of distance between consecutive sensors. Solutions imposing such constraints on the deployment may be costly and difficult to manage. We propose a new approach where the network is formed of virtual nodes, each associated to a certain geographical area. An analytical model of the network traffic per virtual node is proposed and a greedy algorithm to calculate the number of sensors that should form each virtual node is presented. Performance evaluation shows that the greedy deployment can improve the network lifetime by up to 40%, when compared to the uniform deployment. Moreover, the proposed approach outperforms the related work when complemented by a scheduling algorithm which

reduces the messages overhearing. It is also shown that the lifetime of the network can be significantly improved if the battery capacity of each sensor is dimensioned taking into account the traffic it generates or relays.

7.1.4. Core network function placement in self-deployable mobile networks

Emerging mobile network architectures (e.g., aerial networks, disaster relief networks) are disrupting the classical careful planning and deployment of mobile networks by requiring specific self-deployment strategies. Such networks, referred to as self-deployable, are formed by interconnected rapidly deployable base stations that have no dedicated backhaul connection towards a traditional core network. Instead, an entity providing essential core network functionalities is co-located with one of the base stations. In this work, we tackle the problem of placing this core network entity within a self-deployable mobile network, i.e., we determine with which of the base stations it must be co-located [9], [15] [15]. We propose a novel centrality metric, the flow centrality, which measures a node capacity of receiving the total amount of flows in the network. We show that in order to maximize the amount of exchanged traffic between the base stations and the core network entity, under certain capacity and load distribution constraints, the latter should be co-located with the base station having the maximum flow centrality. We first compare our proposed metric to other state of the art centralities. Then, we highlight the significant traffic loss occurring when the core network entity is not placed on the node with the maximum flow centrality, which could reach 55% in some cases.

7.1.5. Cyber physical systems and Internet of things: emerging paradigms on smart cities

A city is smart when investment in traditional and modern infrastructure, human and social capital, fuel well being, high quality of life, and sustainable economic development. The Smart City paradigm is driven by technological evolution in the field of Information and Communication Technologies, and more specifically the paradigms of Internet of Things, Industrial Internet of Things and their confluence with Cyber Physical Systems [12]. Smart Cities present a number of application domains that are related to their critical infrastructures, including energy and transport. These domains present needs similar to the industrial manufacturing environment utilizing smart devices and employing control automation for their applications. They could thus be labeled as *industrial domains* in the wider sense. This work presents three application domains associated with Smart Cities, namely Smart Lighting, Smart Buildings / Energy, and Smart Urban Mobility, identifies their requirements and challenges and reviews existing solutions.

7.2. Wireless data collection

Participants: Oana Iova, Abderrahman Ben Khalifa, Razvan Stanica

7.2.1. Reliable and efficient support for downward traffic in RPL

Modern protocols for wireless sensor networks efficiently support multi-hop upward traffic from many sensors to a collection point, a key functionality enabling monitoring applications. However, the ever-evolving scenarios involving low-power wireless devices increasingly require support also for downward traffic, e.g., enabling a controller to issue actuation commands based on the monitored data. The IETF Routing Protocol for Low-power and Lossy Networks (RPL) is among the few tackling both traffic patterns. Unfortunately, its support for downward traffic is significantly unreliable and inefficient compared to its upward counterpart. We tackle this problem by extending RPL with mechanisms inspired by opposed, yet complementary, principles [7]. At one extreme, we retain the route-based operation of RPL and devise techniques allowed by the standard but commonly neglected by popular implementations. At the other extreme, we rely on flooding as the main networking primitive. Inspired by these principles, we define three base mechanisms, integrate them in a popular RPL implementation, analyze their individual and combined performance, and elicit the resulting tradeoffs in scalability, reliability, and energy consumption. The evaluation relies on simulation, using both real-world topologies from a smart city scenario and synthetic grid ones, as well as on testbed experiments validating our findings from simulation. Results show that the combination of all three mechanisms into a novel protocol, T-RPL *i)* yields high reliability, close to the one of flooding, *ii)* with a low energy consumption, similar to route-based approaches, and *iii)* improves remarkably the scalability of RPL w.r.t. downward traffic.

7.2.2. Performance evaluation of LED-to-camera communications

The use of LED-to-camera communication opens the door to a wide range of use cases and applications, with diverse requirements in terms of quality of service. However, while analytical models and simulation tools exist for all the major radio communication technologies, the only way of currently evaluating the performance of a network mechanism over LED-to-camera is to implement and test it. Our work aims to fill this gap by proposing a Markov-modulated Bernoulli process to model the wireless channel in LED-to-camera communications, which is shown to closely match experimental results [11]. Based on this model, we develop and validate *CamComSim*, the first network simulator for LED-to-camera communications.

7.2.3. Performance evaluation of channel access methods for dedicated IoT networks

Networking technologies dedicated for the Internet of Things are different from the classical mobile networks in terms of architecture and applications. This new type of network is facing several challenges to satisfy specific user requirements. Sharing the communication medium between (hundreds of) thousands of connected nodes and one base station is one of these main requirements, hence the necessity to imagine new solutions, or to adapt existing ones, for medium access control. In this work, we start by comparing two classical medium access control protocols, CSMA/CA and Aloha, in the context of Internet of Things dedicated networks [13]. We continue by evaluating a specific adaptation of Aloha, already used in low-power wide area networks, where no acknowledgement messages are transmitted in the network. Finally, we apply the same concept to CSMA/CA, showing that this can bring a number of benefits. The results we obtain after a thorough simulation study show that the choice of the best protocol depends on many parameters (number of connected objects, traffic arrival rate, allowed retransmission number), as well as on the metric of interest (e.g. packet reception probability or energy consumption).

7.2.4. On the use of wide channels in WiFi networks

An increased density of access points is common today in WiFi deployments, and more and more parameters need to be configured in such networks. In this work, we question current industrial guidelines for both residential and enterprise scenarios [14]. More precisely, we investigate the joint channel, power, and carrier sense threshold allocation problem in IEEE 802.11ac networks, showing that the current practice, which is to use narrower channels at maximum power when the deployment is dense, yields much worse performance than a solution using the widest possible channel with a much lower power.

7.3. Network data exploitation

Participants: Florent Delaine, Panagiota Katsikouli, Hervé Rivano, Razvan Stanica

7.3.1. Calibration algorithms for environmental sensor networks

The recent developments in both nanotechnologies and wireless technologies have enabled the rise of small, low cost and energy efficient environmental sensing devices. Many projects involving dense sensor networks deployments have followed, in particular within the Smart City trend. If such deployments are now within economical and technical reach, their maintenance and reliability remain however a challenge. In particular, reaching, then maintaining, the targeted quality of measurement throughout deployment duration is an important issue. Indeed, factory calibration is too expensive for systematic application to low-cost sensors and as these sensors are usually prone to drifting because of premature aging. In addition, there are concerns about the applicability of factory calibration to field conditions [4]. These challenges have fostered many researches on in situ calibration. In situ means that the sensors are calibrated without removing them from their deployment location, preferably without physical intervention, often leveraging their communication capabilities. It is a critical challenge for the economical sustainability of networks with large scale deployments. In this work, we focus on in situ calibration methods for environmental sensor networks. We propose a taxonomy of the methodologies in the literature. Our classification relies on both the architecture of the network of sensors and the algorithmic principles of the calibration methods. This review allows us to identify and discuss two main challenges: how to improve the performance evaluation of such methods and how to enable a quantified comparison of these strategies?

7.3.2. Characterizing and Removing Oscillations in Mobile Phone Location Data

Human mobility analysis is a multidisciplinary research subject that has attracted a growing interest over the last decade. A substantial amount of such recent studies is driven by the availability of original sources of real-world information about individual movement patterns. An important task in the analysis of mobility data is reliably distinguishing between the stop locations and movement phases that compose the trajectories of the monitored subjects. The problem is especially challenging when mobility is inferred from mobile phone location data: here, oscillations in the association of mobile devices to base stations lead to apparent user mobility even in absence of actual movement [10]. In this work, we leverage a unique dataset of spatiotemporal individual trajectories that allows capturing both the user and network operator perspectives in mobile phone location data, and investigate the oscillation phenomenon. We present probabilistic and machine learning approaches for detecting oscillations in mobile phone location data, and a filtering technique for removing those. Our analyses and comparison with state-of-the-art approaches demonstrate the superiority of our solution, both in terms of removed oscillations and of error with respect to ground-truth trajectories.

AVALON Project-Team

6. New Results

6.1. Energy Efficiency in HPC and Large Scale Distributed Systems

Participants: Laurent Lefèvre, Dorra Boughzala, Thierry Gautier.

6.1.1. Performance and Energy Analysis of OpenMP Runtime Systems with Dense Linear Algebra Algorithms

In the article [4], we analyze performance and energy consumption of five OpenMP runtime systems over a non-uniform memory access (NUMA) platform. We also selected three CPU-level optimizations or techniques to evaluate their impact on the runtime systems: processors features Turbo Boost and C-States, and CPU Dynamic Voltage and Frequency Scaling through Linux CPUFreq governors. We present an experimental study to characterize OpenMP runtime systems on the three main kernels in dense linear algebra algorithms (Cholesky, LU, and QR) in terms of performance and energy consumption. Our experimental results suggest that OpenMP runtime systems can be considered as a new energy leverage, and Turbo Boost, as well as C-States, impacted significantly performance and energy. CPUFreq governors had more impact with Turbo Boost disabled, since both optimizations reduced performance due to CPU thermal limits. An LU factorization with concurrent-write extension from libKOMP achieved up to 63% of performance gain and 29% of energy decrease over original PLASMA algorithm using GNU C compiler (GCC) libGOMP runtime. This paper was first published online in 2018-08-09.

6.1.2. Building and Exploiting the Table of Leverages in Large Scale HPC Systems

Large scale distributed systems and supercomputers consume huge amounts of energy. To address this issue, an heterogeneous set of capabilities and techniques that we call leverages exist to modify power and energy consumption in large scale systems. This includes hardware related leverages (such as Dynamic Voltage and Frequency Scaling), middleware (such as scheduling policies) and application (such as the precision of computation) energy leverages. Discovering such leverages, benchmarking and orchestrating them, remains a real challenge for most of the users. We have formally defined energy leverages, and we proposed a solution to automatically build the table of leverages associated with a large set of independent computing resources. We have shown that the construction of the table can be parallelized at very large scale with a set of independent nodes in order to reduce its execution time while maintaining precision of observed knowledge. In 2019 we have explored the leverage energy-efficient non-lossy compression for data-intensive applications [9].

6.2. HPC Component Models and Runtimes

Participants: Thierry Gautier, Christian Perez, Laurent Turpin, Marie Durand, Philippe Virouleau.

6.2.1. Fine-Grained MPI+OpenMP Plasma Simulations: Communication Overlap with Dependent Tasks

In the article [15], we demonstrate how OpenMP 4.5 tasks can be used to efficiently overlap computations and MPI communications based on a case-study conducted on multi-core and many-core architectures. The paper focuses on task granularity, dependencies and priorities, and also identifies some limitations of OpenMP. Results on 64 Skylake nodes show that while 64% of the wall-clock time is spent in MPI communications, 60% of the cores are busy in computations, which is a good result. Indeed, the chosen dataset is small enough to be a challenging case in terms of overlap and thus useful to assess worst-case scenarios in future simulations. Two key features were identified: by using task priority we improved the performance by 5.7% (mainly due to an improved overlap), and with recursive tasks we shortened the execution time by 9.7%. We also illustrate the need to have access to tools for task tracing and task visualization. These tools allowed a fine understanding and a performance increase for this task-based OpenMP+MPI code.

6.2.2. Patches to LLVM compiler

We propose two source code patches to LLVM https://reviews.llvm.org/D63196 and https://reviews.llvm.org/D67447 in order to improve performance of application using numerous fine grain tasks such as [15]. Patches were accepted in 2019.

6.3. Modeling and Simulation of Parallel Applications and Distributed Infrastructures

Participants: Eddy Caron, Zeina Houmani, Frédéric Suter.

6.3.1. Bridging Concepts and Practice in eScience via Simulation-driven Engineering

The CyberInfrastructure (CI) has been the object of intensive research and development in the last decade, resulting in a rich set of abstractions and interoperable software implementations that are used in production today for supporting ongoing and breakthrough scientific discoveries. A key challenge is the development of tools and application execution frameworks that are robust in current and emerging CI configurations, and that can anticipate the needs of upcoming CI applications. In [14] we presented WRENCH, a framework that enables simulation-driven engineering for evaluating and developing CI application execution frameworks. WRENCH provides a set of high-level simulation abstractions that serve as building blocks for developing custom simulators. These abstractions rely on the scalable and accurate simulation models that are provided by the SimGrid simulation framework. Consequently, WRENCH makes it possible to build, with minimum software development effort, simulators that can accurately and scalably simulate a wide spectrum of large and complex CI scenarios. These simulators can then be used to evaluate and/or compare alternate platform, system, and algorithm designs, so as to drive the development of CI solutions for current and emerging applications.

6.3.2. Accurately Simulating Energy Consumption of I/O-intensive Scientific Workflows

While distributed computing infrastructures can provide infrastructure-level techniques for managing energy consumption, application-level energy consumption models have also been developed to support energy-efficient scheduling and resource provisioning algorithms. In [7], we analyze the accuracy of a widely-used application-level model that have been developed and used in the context of scientific workflow executions. To this end, we profile two production scientific workflows on a distributed platform instrumented with power meters. We then conduct an analysis of power and energy consumption measurements. This analysis shows that power consumption is not linearly related to CPU utilization and that I/O operations significantly impact power, and thus energy consumption. We then propose a power consumption model that accounts for I/O operations, including the impact of waiting for these operations to complete, and for concurrent task executions on multi-socket, multi-core compute nodes. We implement our proposed model as part of a simulator that allows us to draw direct comparisons between real-world and modeled power and energy consumption. We find that our model has high accuracy when compared to real-world executions. Furthermore, our model improves accuracy by about two orders of magnitude when compared to the traditional models used in the energy-efficient workflow scheduling literature.

6.4. Cloud Resource Management

Participants: Eddy Caron, Jad Darrous, Christian Perez.

6.4.1. On the Importance of Container Image Placement for Service Provisioning in the Edge

Edge computing promises to extend Clouds by moving computation close to data sources to facilitate short-running and low-latency applications and services. Providing fast and predictable service provisioning time presets a new and mounting challenge, as the scale of Edge-servers grows and the heterogeneity of networks between them increases. Our work [6] is driven by a simple question: can we place container images across Edge-servers in such a way that an image can be retrieved to any Edge-server fast and in a predictable time. To this end, we present KCBP and KCBP-WC, two container image placement algorithms which aim

to reduce the maximum retrieval time of container images. KCBP and KCBP-WC are based on k-Center optimization. However, KCBP-WC tries to avoid placing large layers of a container image on the same Edge-server. Evaluations using trace-driven simulations show that KCBP and KCBP-WC can be applied to various network configurations and reduce the maximum retrieval time of container images by 1.1x to 4x compared to state-of-the-art placements (*i.e.,* Best-Fit and Random).

Data-intensive clusters are heavily relying on distributed storage systems to accommodate the unprecedented growth of data. Hadoop distributed file system (HDFS) is the primary storage for data analytic frameworks such as Spark and Hadoop. Traditionally, HDFS operates under replication to ensure data availability and to allow locality-aware task execution of data-intensive applications. Recently, erasure coding (EC) is emerging as an alternative method to replication in storage systems due to the continuous reduction in its computation overhead. We have conducted an extensive experimental study to understand the performance of data-intensive applications under replication and EC [5], [23]. We use representative benchmarks on the Grid'5000 testbed to evaluate how analytic workloads, data persistency, failures, the back-end storage devices, and the network configuration impact their performances. Our study sheds the light not only on the potential benefits of erasure coding in data-intensive clusters but also on the aspects that may help to realize it effectively.

6.5. Data Stream Processing on Edge Computing

Participants: Eddy Caron, Felipe Rodrigo de Souza, Marcos Dias de Assunção, Laurent Lefèvre, Alexandre Da Silva Veith.

6.5.1. Operator Placement for Data Stream Processing on Fog/Edge Computing

DSP (Data Stream Processing) frameworks are often employed to process the large amount of data generated by the increasing number of IoT devices. A DSP application is commonly structured as a directed graph, or dataflow, whose vertices are operators that perform transformations over the incoming data and edges representing the data dependencies between operators. Such applications are often deployed on the Cloud in order to explore the large number of available resources and its pay-as-you-go business model. Fog computing enables offloading operators from the cloud by placing them close to where the data is generated, whereby reducing the time to process data events. However, fog computing resources often have lower capacity than those available in the Cloud. When offloading operators from the Cloud, the scheduler needs to adjust their level of parallelism and hence decides on the number of operator instances to create during placement in order to achieve a given throughput. This gives rise to two interrelated issues, namely deciding the operators parallelism and computing their placement onto available resources [16].

While addressing the placement problem [8], we proposed an approach consisting of a programming model and real-world implementation of an IoT application. The results show that our approach can minimise the end-to-end latency by at least 38% by pushing part of the IoT application to edge computing resources. Meanwhile, the edge-to-cloud data transfers are reduced by at least 38%, and the messaging costs are reduced by at least 50% when using the existing commercial edge cloud cost models.

In addition, we have designed and validated a discrete event simulation for modelling and simulation of DSP applications on edge computing environments [3].

6.5.2. Multi-Objective Reinforcement Learning for Reconfiguring Data Stream Analytics on Edge Computing

As DSP applications are often long-running, their workload and the infrastructure conditions can change over time. When changes occur, the application must be reconfigured. The operator reconfiguration consists of changing the initial placement by reassigning operators to different devices given target performance metrics. We modelled the operator reconfiguration as a Reinforcement Learning (RL) problem and defined a multi-objective reward considering metrics regarding operator reconfiguration, and infrastructure and application improvement [11]. We also use Monte Carlo Tree Search to organise the episodes generated during simulation and training [12]. Experimental results show that reconfiguration algorithms that minimise only end-to-end processing latency can have a substantial impact on WAN traffic and communication cost. The results also

demonstrate that when reconfiguring operators, RL algorithms improve by over 50% the performance of the initial placement provided by state-of-the-art approaches.

6.6. An Operational Tool for Software Asset Management Improvement

Participants: Eddy Caron, Arthur Chevalier.

6.6.1. Multi-objective algorithm that guarantees license compliance

We have developed a new feature to OpTISAM, an Orange™ software offering tools to perform Software Asset Management (SAM) much more efficiently in order to be able to ensure the full compliance with all contracts from each software and a new type of deployment taking into account these aspects and other additional parameters like energy and performance. Our new feature is a multi-objective algorithm for deploying services in the Cloud that guarantees license compliance while reducing energy consumption but maintaining reasonable performance. In both cases of use and with a significant set of 5000 servers, we were able to show our approach is close to the best values in each criterion while dropping less than 10% of performance each time while keeping a full compliance.

6.7. Platform

Participants: Thierry Gautier, Christian Perez, Simon Delamare, Laurent Lefèvre.

6.7.1. Gemini cluster based on DGX-1 high density computer

The LECO experimental platform is a new medium size scientific instrument funded by DRRT and Inria to investigate research related to BigData and HPC. It is bi-located in Grenoble as part of the the HPCDA computer managed by UMS GRICAD (deployed in 2018) and in Lyon as part of the Grid5K Gemini cluster. The Gemini cluster is composed of two DGX-1 high density computers for HPC and BigData. Each computers has 8 NVIDIA V100 GPGPU cards with 4 infiniband high speed network cards.

CTRL-A Project-Team

7. New Results

7.1. Programming support for Autonomic Computing

7.1.1. Reactive languages

Participants: Gwenaël Delaval, Lucie Muller, Eric Rutten.

Our work in reactive programming for autonomic computing systems is focused on the specification and compilation of declarative control objectives, under the form of contracts, enforced upon classical mode automata as defined in synchronous languages. The compilation involves a phase of Discrete Controller Synthesis, integrating the tool ReaX, in order to obtain an imperative executable code. The programming language Heptagon / BZR (see Section Software and Platforms) integrates our research results [5].

An ongoing topic is on abstraction methods for compilation using discrete controller synthesis (needed for example, in order to program the controllers for systems where the useful data for control can be of arbitrary types (integer, real, ...), or also for systems which are naturally distributed, and require a decentralized controller).

Recent work concerns compilation and diagnosis for discrete controller synthesis. The compilation involving a phase of controller synthesis can fail to find a solution, if the problem is overconstrained. The compiler does notify so to the programmer, but the latter would need a diagnosis in order to understand where and how to debug the program. Such diagnosis is made especially difficult by the declarative nature of the synthesis.

This was the object of the M1 TER internship of Lucie Muller [19].

7.1.2. Domain-specific languages

Participants: Gwenaël Delaval, Soguy Mak Kare Gueye, Eric Rutten.

Our work in Domain-specific languages (DSLs) is founded on our work in component-based programming for autonomic computing systems as examplified by e.g., FRACTAL. We consider essentially the problem of specifying the control of components assembly reconfiguration, with an approach based on the integration within such a component-based framework of a reactive language as in Section 7.1.1 [4]. In recent work, we proposed an extension of a classical Software Architecture Description Languages (ADL) with Ctrl-F, DSL for the specification of dynamic reconfiguration behavior in a [1]. Based on this experience, we also proposed a DSL called Ctrl-DPR [6], allowing designers to easily generate Autonomic Managers for DPR FPGA systems (see Section 7.2.3).

Ongoing work involves a generalization from our past experiences in software components, DPR FPGA, as well as IoT [8], and Cyberphysical Systems. As we observed a similarity in objects and structures (e.g., tasks, implementation versions, resources, and upper-level application layer), we are considering a more general DSL, which could be specialized towards such different target domains, and where the compilation towards reactive models could be studied and improved, especially considering the features of Section 7.1.1 . This direction will also lead us to study the definition of software architecture patterns for multiple loops Autonomic Managers, particularly hierarchical, with lower layers autonomy alleviating management burden from the upper layers as in Section 7.2 .

7.2. Design methods for reconfiguration controller design in computing systems

We apply the results of the previous axes of the team's activity, as well as other control techniques, to a range of infrastructures of different natures, but sharing a transversal problem of reconfiguration control design. From this very diversity of validations and experiences, we draw a synthesis of the whole approach, towards a general view of Feedback Control as MAPE-K loop in Autonomic Computing [7], [9].

7.2.1. Self-adaptative distributed systems

Participants: Quang Pham Tran Anh, Eric Rutten, Hamza Sahli.

Complex Autonomic Computing Systems, as found typically in distributed systems, must involve multiple management loops, addressing different subproblems of the general management, and using different modeling, decision and control approaches (discrete [3], continuous, stochastic, machine-learning based, ...) They are generally addressing deployment and allocation of computations on resources w.r.t. QoS, load, faults, ... but following different, complementary approaches. The similarities and recurring patterns are considered as in Section 7.1.2 . Their execution needs to be distributed w.r.t. different characteristics such as latency (as in Fog and Edge Computing) or load. We are studying Software Architectures to address the design of such complex systems.

7.2.1.1. Self-adaptation of micro-services in Fog/Edge and Cloud computing

Fog systems are a recent trend of distributed computing having vastly ubiquitous architectures and distinct requirements making their design difficult and complex. Fog computing is based on leveraging both resource-scarce computing nodes around the Edge to perform latency and delay sensitive tasks and Cloud servers for the more intensive computation.

In this work, we present a formal model defining spatial and structural aspects of Fog-based systems using Bigraphical Reactive Systems, a fully graphical process algebraic formalism. The model is extended with reaction rules to represent the dynamic behavior of Fog systems in terms of self-adaptation. The notion of bigraph patterns is used in conjunction with boolean and temporal operators to encode spatio-temporal properties inherent to Fog systems and applications. The feasibility of the modelling approach is demonstrated via a motivating case study and various self-adaptation scenarios.

This work is done in cooperation with the Inria team Stack in Nantes, and published in the FOCLASA workshop, co-located with the SFEM conference [13].

7.2.1.2. Autonomic management in Software Defined Networks

In the framework of our cooperation with Nokia Bell-labs (See Section 8.1.2), and the Dyonisos team at Inria Rennes, we are considering the management of Software Defined Networks (SDN), involving Data-Centers and accelerators.

The main approach AI / Machine Learning approaches, developed in Rennes. An ongoing topic is to consider that these reinforcement learning based approaches involve questions of trust, and we are beginning to consider their composition with controllers based e.g. on Control Theory, in order to maintain guarantees on the behaviors of the managed system.

7.2.2. High-Performance Grid Computing

Cloud and HPC (High-Performance Computing) systems have increasingly become more varying in their behavior, in particular in aspects such as performance and power consumption, and the fact that they are becoming less predictable demands more runtime management [10].

7.2.2.1. A Control-Theory based approach to minimize cluster underuse

Participants: Abdul Hafeez Ali, Raphaël Bleuse, Bogdan Robu, Eric Rutten.

One such problem is found in the context of CiGri, a simple, lightweight, scalable and fault tolerant grid system which exploits the unused resources of a set of computing clusters. In this work, we consider autonomic administration in HPC systems for scientific workflows management through a control theoretical approach. We propose a model described by parameters related to the key aspects of the infrastructure thus achieving a deterministic dynamical representation that covers the diverse and time-varying behaviors of the real computing system. We propose a model-predictive control loop to achieve two different objectives: maximize cluster utilization by best-effort jobs and control the file server's load in the presence of external disturbances. The accuracy of the prediction relies on a parameter estimation scheme based on the EKF (Extended Kalman Filter) to adjust the predictive-model to the real system, making the approach adaptive to parametric variations in the infrastructure. The closed loop strategy shows performance improvement and consequently a reduction

in the total computation time. The problem is addressed in a general way, to allow the implementation on similar HPC platforms, as well as scalability to different infrastructures.

This work is done in cooperation with the Datamove team of Inria/LIG, and Gipsa-lab. Some results were published in the CCTA conference [14]. It was the topic of the Master's thesis of Abdul Hafeez Ali [16].

7.2.2.2. Combining Scheduling and Autonomic Computing for Parallel Computing Resource Management

Participants: Raphaël Bleuse, Eric Rutten.

This research topic aims at studying the relationships between scheduling and autonomic computing techniques to manage resources for parallel computing platforms. The performance of such platforms has greatly improved (149 petaflops as of November 2019 [20]) at the cost of a greater complexity: the platforms now contain several millions of computing units. While these computation units are diverse, one has to consider other constraints such as the amount of free memory, the available bandwidth, or the energetic envelope. The variety of resources to manage builds complexity up on its own. For example, the performance of the platforms depends on the sequencing of the operations, the structure (or lack thereof) of the processed data, or the combination of application running simultaneously.

Scheduling techniques offer great tools to study/guaranty performances of the platforms, but they often rely on complex modeling of the platforms. They furthermore face scaling difficulties to match the complexity of new platforms. Autonomic computing manages the platform during runtime (on-line) in order to respond to the variability. This approach is structured around the concept of feedback loops.

The scheduling community has studied techniques relying on autonomic notions, but it has failed to link the notions up. We are starting to address this topic.

7.2.3. High-Performance Embedded Computing

Participants: Soguy Mak Kare Gueye, Stéphane Mocanu, Eric Rutten.

This topics build upon our experience in reconfiguration control in DPR FPGA [2].

Implementing self-adaptive embedded systems, such as UAV drones, involves an offline provisioning of the several implementations of the embedded functionalities with different characteristics in resource usage and performance in order for the system to dynamically adapt itself under uncertainties. We propose an autonomic control architecture for self-adaptive and self-reconfigurable FPGA-based embedded systems. The control architecture is structured in three layers: a mission manager, a reconfiguration manager and a scheduling manager. This work is in the framework of the ANR project HPeC (see Section 9.2.1).

7.2.3.1. DPR FPGA and discrete control for reconfiguration

In this work we focus on the design of the reconfiguration manager. We propose a design approach using automata-based discrete control. It involves reactive programming that provides formal semantics, and discrete controller synthesis from declarative objectives.

Ongoing work concerns experimental validation, where upon the availability of hardware implementations of vision, detection and tracking tasks, a demonstrator is being built integrating our controller.

7.2.3.2. Mission management and stochastic control

In the Mission Management workpackage of the ANR project HPeC, a concurrent control methodology is constructed for the optimal mission planning of a U.A.V. in stochastic environnement. The control approach is based on parallel ressource sharing Partially Observable Markov Decision Processes modeling of the mission. The parallel POMDP are reduced to discrete Markov Decision Models using Bayesian Networks evidence for state identification. The control synthesis is an iterative two step procedure : first MDP are solved for the optimisation of a finite horizon cost problem ; then the possible ressource conflicts between parallel actions are solved either by a priority policy or by a QoS degradation of actions, e.g., like using a lower resolution version of the image processing task if the ressource availability is critical.

This work was performed in the framework of the PhD of Chabha Hireche, defended in nov. 2019 [17].

7.2.4. IoT and Cyberphysical Systems

Participants: Neil Ayeb, Ayan Hore, Fabien Lefevre, Stéphane Mocanu, Jan Pristas, Eric Rutten, Gaetan Sorin, Mohsen Zargarani.

7.2.4.1. Device management

The research topic is targeting an adaptative and decentralized management for the IoT. It will contribute design methods for processes in virtualized gateways in order to enhance IoT infrastructures. More precisely, it concerns Device Management (DM) in the case of large numbers of connected sensors and actuators, as can be found in Smart Home and Building, Smart Electricity grids, and industrial frameworks as in Industry 4.0.

Device Management is currently industrially deployed for LAN devices, phones and workstation management. Internet of Things (IoT) devices are massive, dynamic, heterogeneous, and inter-operable. Existing solutions are not suitable for IoT management. This work in an industrial environment addresses these limitations with a novel autonomic and distributed approach for the DM.

This work is in the framework of the Inria/Orange labs joint laboratory (see Section 8.1.1), and supported by the CIFRE PhD thesis grant of Neïl Ayeb, starting dec. 2017. It was awarded a best paper distinction at the Doctoral Symposium of ICAC 2019 [12].

7.2.4.2. Security in SCADA industrial systems

We focus mainly on vulnerability search, automatic attack vectors synthesis and intrusion detection [11]. Model checking techniques are used for vulnerability search and automatic attack vectors construction. Intrusion detection is mainly based on process-oriented detection with a technical approach from run-time monitoring. The LTL formalism is used to express safety properties which are mined on an attack-free dataset. The resulting monitors are used for fast intrusion detections. A demonstrator of attack/defense scenario in SCADA systems has been built on the existing G-ICS lab (hosted by ENSE3/Grenoble-INP). This work is in the framework of the ANR project Sacade on cybersecurity of industrial systems (see Section 9.2.2).

One of important results is the realization of a Hardware-in-the-loop SCADA Cyberange based on a electronic interface card that allows to interface real-world PLC with a software simulation [21]. The entire system is available in open-source including the electronic card fabrication files (http://gics-hil.gforge.inria.fr/). Interfacing system allow connection with various commercial simulation software but also with “home made” simulators [15]. The work is also supported by Grenoble Alpes Cybersecurity Institute (see Section 9.1.1) and Pulse program of IRT NANOELEC.

Ongoing work concerns the complementary topic of analysis and identification of reaction mechanisms for self-protection in cybersecurity, where, beyond classical defense mechanisms that detect intrusions and attacks or assess the kind of danger that is caused by them, we explore models and control techniques for the automated reaction to attacks, in order to use detection information to take the appropriate defense and repair actions. A first approach was developed in the M2R internship by Ayan Hore [18]

DANTE Project-Team

7. New Results

7.1. Graph Signal Processing and Machine Learning

Participants: Paulo Gonçalves, Rémi Gribonval, Marion Foare, Thomas Begin, Esteban Bautista Ruiz, Gaetan Frusque, Amélie Barbe, Mikhail Tsitsvero, Marija Stojanova, Márton Karsai, Sébastien Lerique, Jacobo Levy Abitbol.

7.1.1. L^γ -PageRank for Semi-Supervised Learning

Participants: Paulo Gonçalves, Esteban Bautista Ruiz.

PageRank for Semi-Supervised Learning has shown to leverage data structures and limited tagged examples to yield meaningful classification. Despite successes, classification performance can still be improved, particularly in cases of fuzzy graphs or unbalanced labeled data. To address such limitations, a novel approach based on powers of the Laplacian matrix L^γ ($\gamma > 0$), referred to as L^γ-PageRank, is proposed. Its theoretical study shows that it operates on signed graphs, where nodes belonging to one same class are more likely to share positive edges while nodes from different classes are more likely to be connected with negative edges. It is shown that by selecting an optimal γ, classification performance can be significantly enhanced. A procedure for the automated estimation of the optimal γ, from a unique observation of data, is devised and assessed. Experiments on several datasets demonstrate the effectiveness of both L^γ-PageRank classification and the optimal γ estimation. [11]

7.1.2. Designing Convex Combination of Graph Filters

Participant: Paulo Gonçalves.

In this work, we studied the problem of parametric modeling of network-structured signals with graph filters. Unlike the popular polynomial graph filters, which are based on a single graph shift operator, we considered convex combinations of graph shift operators particularly adapted to directed graphs. As the resulting modeling problem is not convex, we reformulated it as a convex optimization problem which can be solved efficiently. Experiments on real-world data structured by undirected and directed graphs were conducted. The results showed the effectiveness of this method compared to other methods reported in the literature. [18]

7.1.3. Optimal transport under regularity constraints for domain adaptation between graphs with attributes

Participants: Paulo Gonçalves, Amélie Barbe.

In this work, we addresses the problem of domain adaptation between two graphs by optimal transport. We aimed at benefiting from the knowledge of a labeled source graph to improve the classification of nodes in an unlabeled target graph. We focused on the setting where a set of features is associated to each node of the graphs. We proposed an original method that optimizes a transportation plan from the source to the target that *(i)* preserves the structures transported between the graphs and *(ii)* prevents the mapping from transporting two source nodes with different labels to the same destination. [30]

7.1.4. Sparse tensor dimensionality reduction with application to the clustering of functional connectivity in the brain

Participants: Paulo Gonçalves, Gaetan Frusque.

Functional connectivity (FC) is a graph-like data structure commonly used by neuroscientists to study the dynamic behaviour of the brain activity. However, these analyses rapidly become complex and time-consuming, as the number of connectivity components to be studied is quadratic with the number of electrodes. In our work, we addressed the problem of clustering FC into relevant ensembles of simultaneously activated components that reveal characteristic patterns of the epileptic seizures of a given patient. While $k-$means is certainly the most popular method for data clustering, it is known to perform badly on large dimensional data sets, and to be highly sensitive to noise. To overcome the co-called curse of dimensionality, we proposed a new tensor decomposition to reduce the size of the data set formed by FC time series recorded for several seizures, before applying k-means. Our contribution is twofold: First, we derived a method that we compared to the state of the art, emphasizing one variant that imposes sparsity constraints. Second, we conducted a real case study, applying the proposed sparse tensor decomposition to epileptic data in order to infer the functional connectivity graph dynamics corresponding to the different stages of an epileptic seizure. [31], [47]

7.1.5. Graph signal processing to model WLANs performances

Participants: Paulo Gonçalves, Thomas Begin, Marija Stojanova.

As WLANs have become part of our everyday life, there is an increasing need for more transmission capacity and wireless coverage. In response to this growing need, network administrators tend to intensify the deployment of Access Points (APs). However, if not correctly done, this AP densification may lead to badly planned and uncoordinated networks with sub-optimal use of the available resources. In this work, we propose a data-driven approach using graph signal processing and a set of input/output signals to capture the behavior of a WLAN and derive a predictive performance model. Given the simplicity and the novelty of the proposed model, we believe that its relative error of around 10-20% in modeling and 25% in prediction may represent a promising start for new approaches in the modeling of WLANs. [33]

7.1.6. Joint embedding of structure and features via graph convolutional networks

Participants: Márton Karsai, Sébastien Lerique.

We propose *AN2VEC*, a node embedding method which ultimately aims at disentangling the information shared by the structure of a network and the features of its nodes. Building on the recent developments of Graph Convolutional Networks (GCN), we develop a multitask GCN Variational Autoencoder where different dimensions of the generated embeddings can be dedicated to encoding feature information, network structure, and shared feature-network information. We explore the interaction between these disentangled characters by comparing the embedding reconstruction performance to a baseline case where no shared information is extracted. We use synthetic datasets with different levels of interdependency between feature and network characters and show (i) that shallow embeddings relying on shared information perform better than the corresponding reference with unshared information, (ii) that this performance gap increases with the correlation between network and feature structure, and (iii) that our embedding is able to capture joint information of structure and features. Our method can be relevant for the analysis and prediction of any featured network structure ranging from online social systems to network medicine. [51]

7.2. Computational Human Dynamics and Temporal Networks

Participants: Márton Karsai, Sébastien Lerique, Jacobo Levy Abitbol, Samuel Unicomb, Sicheng Dai.

7.2.1. Optimal Proxy Selection for Socioeconomic Status Inference on Twitter

Participants: Márton Karsai, Jacobo Levy Abitbol.

The socioeconomic status of people depends on a combination of individual characteristics and environmental variables, thus its inference from online behavioral data is a difficult task. Attributes like user semantics in communication, habitat, occupation, or social network are all known to be determinant predictors of this feature. In this paper we propose three different data collection and combination methods to first estimate and, in turn, infer the socioeconomic status of French Twitter users from their online semantics. Our methods are based on open census data, crawled professional profiles, and remotely sensed, expert annotated information on living environment. Our inference models reach similar performance of earlier results with the advantage of relying on broadly available datasets and of providing a generalizable framework to estimate socioeconomic status of large numbers of Twitter users. These results may contribute to the scientific discussion on social stratification and inequalities, and may fuel several applications. [19]

7.2.2. Randomized reference models for temporal networks

Participant: Márton Karsai.

In this paper we propose a unified framework for classifying and understanding microcanonical RRMs (MRRMs). Focusing on temporal networks, we use this framework to build a taxonomy of MRRMs that proposes a canonical naming convention, classifies them, and deduces their effects on a range of important network features. We furthermore show that certain classes of compatible MRRMs may be applied in sequential composition to generate over a hundred new MRRMs from the existing ones surveyed in this article. We provide two tutorials showing applications of the MRRM framework to empirical temporal networks: 1) to analyze how different features of a network affect other features and 2) to analyze how such features affect a dynamic process in the network. We finally survey applications of MRRMs found in literature. [48]

7.2.3. Reentrant phase transitions in threshold driven contagion on multiplex networks

Participants: Márton Karsai, Samuel Unicomb.

Models of threshold driven contagion explain the cascading spread of information, behavior, systemic risk, and epidemics on social, financial and biological networks. At odds with empirical observation, these models predict that single-layer unweighted networks become resistant to global cascades after reaching sufficient connectivity. We investigate threshold driven contagion on weight heterogeneous multiplex networks and show that they can remain susceptible to global cascades at any level of connectivity, and with increasing edge density pass through alternating phases of stability and instability in the form of reentrant phase transitions of contagion. Our results provide a novel theoretical explanation for the observation of large scale contagion in highly connected but heterogeneous networks. [23]

7.2.4. Interactional and informational attention on Twitter

Twitter may be considered as a decentralized social information processing platform whose users constantly receive their followees' information feeds, which they may in turn dispatch to their followers. This decentralization is not devoid of hierarchy and heterogeneity, both in terms of activity and attention. In particular, we appraise the distribution of attention at the collective and individual level, which exhibits the existence of attentional constraints and focus effects. We observe that most users usually concentrate their attention on a limited core of peers and topics, and discuss the relationship between interactional and informational attention processes – all of which, we suggest, may be useful to refine influence models by enabling the consideration of differential attention likelihood depending on users, their activity levels and peers' positions. [10]

7.2.5. Efficient limited time reachability estimation in temporal networks

Participant: Márton Karsai.

In this paper we propose a probabilistic counting algorithm, which gives simultaneous and precise estimates of the in- and out-reachability (with any chosen waiting-time limit) for every starting event in a temporal network. Our method is scalable allowing measurements for temporal networks with hundreds of millions of events. This opens up the possibility to analyse reachability, spreading processes, and other dynamics in large temporal networks in completely new ways; to compute centralities based on global reachability for all events; or to find with high probability the exact node and time, which could lead to the largest epidemic outbreak. [52]

7.2.6. weg2vec: Event embedding for temporal networks

Participant: Márton Karsai.

Network embedding techniques are powerful to capture structural regularities in networks and to identify similarities between their local fabrics. However, conventional network embedding models are developed for static structures, commonly consider nodes only and they are seriously challenged when the network is varying in time. Temporal networks may provide an advantage in the description of real systems, but they code more complex information, which could be effectively represented only by a handful of methods so far. Here, we propose a new method of event embedding of temporal networks, called weg2vec, which builds on temporal and structural similarities of events to learn a low dimensional representation of a temporal network. This projection successfully captures latent structures and similarities between events involving different nodes at different times and provides ways to predict the final outcome of spreading processes unfolding on the temporal structure. [53]

7.3. Communication Networks

Participants: Thomas Begin, Anthony Busson, Isabelle Guérin Lassous, Marion Foare, Philippe Nain, Lafdal Abdelwedoud, Marija Stojanova, Rémy Grünblatt, Juan Pablo Astudillo.

7.3.1. Quantum communications

In [29] we investigate the performance of a quantum switch serving a set of users. The function of the switch is to convert bipartite entanglement generated over individual links connecting each user to the switch, into bipartite or tripartite entangled states among (pairs or groups of) users at the highest possible rates at a fixed ratio. Such entanglement can then be converted to quantum-secure shared secret bits among pairs or triples of users using E91-like Quantum Key Distribution (QKD) protocols. The switch can store a certain number of qubits in a quantum memory for a certain length of time, and can make two-qubit Bell-basis measurements or three-qubit GHZ-basis projective measurements on qubits held in the memory. We model a set of randomized switching policies. Discovering that some are better than others, we present analytical results for the case where the switch stores one qubit per user at a given time step, and find that the best policies outperform a time division multiplexing (TDM) policy for sharing the switch between bipartite and tripartite entanglement generation. This performance improvement decreases as the number of users grows. The model is easily augmented to study the capacity region in the presence of qubit decoherence, obtaining similar results. Moreover, decoherence appears to have little effect on capacity. We also study a smaller class of policies when the switch can store two qubits per user.

7.3.2. Resource Allocation

In [28] we consider assignment policies that allocate resources to users, where both resources and users are located on a one-dimensional line $[0, \infty\infty)$. First, we consider unidirectional assignment policies that allocate resources only to users located to their left. We propose the Move to Right (MTR) policy, which scans from left to right assigning nearest rightmost available resource to a user, and contrast it to the Unidirectional Gale-Shapley (UGS) matching policy. While both policies among all unidirectional policies minimize the expected distance traveled by a request (request distance), MTR is fairer. Moreover, we show that when user and resource locations are modeled by statistical point processes, and resources are allowed to satisfy more than one user, the spatial system under unidirectional policies can be mapped into bulk service queueing systems, thus allowing the application of many queueing theory results that yield closed-form expressions. As we consider a case where different resources can satisfy different numbers of users, we also generate new results for bulk service queues. We also consider bidirectional policies where there are no directional restrictions on resource allocation and develop an algorithm for computing the optimal assignment which is more efficient than known algorithms in the literature when there are more resources than users. Finally, numerical evaluation of performance of unidirectional and bidirectional allocation schemes yields design guidelines beneficial for resource placement.

7.3.3. VoD broadcasting over vehicular networks

Participants: Thomas Begin, Anthony Busson, Isabelle Guérin Lassous.

We consider a VoD (Video on-Demand) platform designed for vehicles traveling on a highway or other major roadway. Typically, cars or buses would subscribe to this delivery service so that their passengers get access to a catalog of movies and series stored on a back-end server. The network infrastructure comprises IEEE 802.11p RSUs (Road Side Units) that are deployed along the highway and deliver video content to traveling vehicles. In this paper, we propose a simple analytical and yet accurate solution to estimate two key performance parameters for a VoD platform: (i) the average download data rate experienced by vehicles over their journey and (ii) the average "interruption time", which corresponds to the fraction of time the video playback of a given vehicle is interrupted because of an empty buffer. Through multiple examples, we investigate the influence of several parameters (e.g., the video bit rate, the number of vehicles, the distance between RSUs, the vehicle velocity) on these two performance parameters whose outcome may help the sizing of an IEEE 802.11p-based VoD platform [12].

7.3.4. Performance Evaluation of Channel Bonding in IEEE 802.11ac

Participants: Thomas Begin, Anthony Busson, Marija Stojanova.

WLANs grow in popularity in home, public, and work environments, resulting in constantly increasing demands for wireless coverage and capacity. There exist two dominant strategies that help solve the problem of WLAN capacity: the deployment of more APs and enhancement of the standards in use. These policies result in WLANs containing a larger number of more complex devices, making the prediction of the network's behavior an even more elaborate problem. Because of these issues, WLANs are prone to inefficient configurations. In this paper, we propose a Markovian continuous time model that aims at predicting the throughputs achieved by all the WLAN's APs as a function of the network's topology and the AP's throughput demands. By means of simulation, we show that our model achieves mean relative errors of less than 10% for networks of different sizes and with diverse node configurations. The model is adapted to the specificities of the IEEE 802.11ac standard amendment and can be used to solve problems such as channel assignment or channel bonding. We derive guidelines on the best practice in channel bonding given a performance metric and for different MCS indexes, frame aggregation rates, saturation levels, and network topologies. We then put our findings to the test by identifying the optimal channel bonding combination in a WLAN containing a diverse set of nodes.

7.3.5. Distributed Congestion Control mechanism for NANs

Participants: Thomas Begin, Anthony Busson, Juan Pablo Astudillo.

The need for significant improvements in the management and efficient use of electrical energy has led to the evolution from the traditional electrical infrastructures towards modern Smart Grid networks. Taking into account the critical importance of this type of networks, multiple research groups focus their work on issues related to the generation, transport and consumption of electrical energy. One of the key research points is the data communication network associated with the electricity transport infrastructure, and specifically the network that interconnects the devices in consumers' homes, the so-called Neighborhood Area Networks (NANs). In this paper, a new distributed congestion control mechanism is proposed, implemented and evaluated for NANs. Besides, different priorities have been considered for the traffic flows transmitted by different applications. The main goal is to provide with the needed Quality of Service (QoS) to all traffic flows, especially in high traffic load situations. The proposal is evaluated in the context of a wireless ad hoc network made up by a set of smart meter devices, using the Ad hoc On-Demand Distance Vector (AODV) routing protocol and the IEEE 802.11ac physical layer standard. The application of the proposed congestion control mechanism, together with the necessary modifications made to the AODV protocol, lead to performance improvements in terms of packet delivery ratio, network throughput and transit time, fairness between different traffic sources and QoS provision [35].

7.3.6. Simulation and Performance Evaluation of the Intel Rate Adaptation Algorithm

Participants: Rémy Grünblatt, Isabelle Guérin-Lassous.

With the rise of the complexity of the IEEE 802.11 standard, rate adaptation algorithms have to deal with a large set of values for all the different parameters having an impact on the network throughput. Simple trial-and-error algorithms can no longer explore solution space in reasonable time and smart solutions are required. Most of the WiFi controllers rely on proprietary code and the used rate adaptation algorithms in these controllers are unknown. Very few WiFi controllers expose their rate adaptation algorithms if they do not rely on the MINSTREL-HT algorithm which is implemented in the mac80211 component of the Linux kernel. Intel WiFi controllers come with their own rate adaptation algorithms that are implemented in the Intel IWLWIFI Linux Driver which is open-source.

In this work, we have reverse-engineered the Intel rate adaptation mechanism from the source code of the IWLWIFI Linux driver, and we give, in a comprehensive form, the underlying rate adaptation algorithm named IWL-MVM-RS. We describe the different mechanisms used to seek the best throughput adapted to the network conditions. We have also implemented the IWL-MVM-RS algorithm in the NS-3 simulator. Thanks to this implementation, we can evaluate the performance of IWL-MVM-RS in different scenarios (static and with mobility, with and without fast fading). We also compare the performances of IWL-MVM-RS with the ones of MINSTREL-HT and IDEALWIFI, also implemented in the NS-3 simulator [26], [32].

7.3.7. A Passive Method to Infer the Weighted Conflict Graph of a IEEE 802.11 Network

Participants: Lafdal Abdelwedoud, Anthony Busson, Isabelle Guérin-Lassous, Marion Foare.

Wi-Fi networks often consist of several Access Points (APs) to form an Extended Service Set. These APs may interfere with each other as soon as they use the same channel or overlapping channels. A classical model to describe interference is the conflict graph. As the interference level varies in the network and in time, we consider a weighted conflict graph. In this work, we propose a method to infer the weights of the conflict graph of a Wi-Fi network.

Weights represent the proportion of activity from a neighbor detected by the Clear Channel Assessment mechanism. Our method relies on a theoretical model based on Markov networks applied to a decomposition of the original conflict graph. The input of our solution is the activity measured at each AP, measurements available in practice. The proposed method is validated through ns-3 simulations performed for different scenarios. Results show that our solution is able to accurately estimate the weights of the conflict graph. [24], [34].

DATAMOVE Project-Team

7. New Results

7.1. Integration of High Performance Computing and Data Analytics

7.1.1. In Situ Processing Model

The work in [2] focuses on proposing a model for in situ analysis taking into account memory constraints. This model is used to provide different scheduling policies to determine both the number of resources that should be dedicated to analysis functions, and that schedule efficiently these functions. We evaluate them and show the importance of considering memory constraints when choosing in between in situ and in transit resource allocation.

7.1.2. I/O Characterization

I/O operations are the bottleneck of several HPC applications due to the difference between process- ing and data access speeds. Hence, it is important to understand and characterize the typical I/O behavior of these applications, so we can identify problems in HPC architectures and propose solutions. In [3], we conducted an extensive analysis to collect and analyze information about applications that run in the Santos Dumont supercomputer, deployed in the National Laboratory for Scientific Computing (LNCC), in Brazil. In [9], we propose an I/O characterization approach that uses unsupervised learning to cluster jobs with similar I/O behavior, using information from high-level aggregated traces.

7.1.3. Online adaptation of the I/O stack to applications

I/O optimization techniques such as request scheduling can improve performance mainly for the access patterns they target, or they depend on the precise tune of parameters. In [19], we propose an approach to adapt the I/O forwarding layer of HPC systems to the application access patterns by tuning a request scheduler. Our case study is the TWINS scheduling algorithm, where performance improvements depend on the time window parameter, which depends on the current workload. Our approach uses a reinforcement learning technique to make the system capable of learning the best parameter value to each access pattern during its execution, without a previous training phase. Our approach can achieve a precision of 88% on the parameter selection in the first hundreds of observations of an access pattern. After having observed an access pattern for a few minutes (not necessarily contiguously), the system will be able to optimize its performance for the rest of the life of the system (years).

Such an auto-tuning approach requires a classification of application access patterns,to separate situations where the optimization techniques will have a different performance behavior. Such a classification is not available in the stateless server-side, hence it has to be estimated from metrics on recent accesses. In [8], we evaluate three machine learning techniques to automatically detect the I/O access pattern of HPC applications at run time: decision trees, random forests, and neural networks. We also proposed in [15] a pattern matching approach for server-side access pattern detection for the HPC I/O stack. The goal is to empower the system to learn a classification during the execution of the system, by representing access patterns by all relevant metrics. We build a time series to represent accesses spatiality, and use a pattern matching algorithm, in addition to an heuristic, to compare it to known patterns.

7.1.4. Data management for workflow execution

In [11], we studied a typical scenario in research facilities. Instrumental data is generated by lab equipment such as microscopes, collected by researchers into USB devices, and analyzed in their own computers. In this scenario, an instrumental data management framework could store data in a institution-level storage infrastructure and allow to execute tasks to analyze this data in some available processing nodes. This setup has the advantages of promoting reproducible research and the efficient usage of the expensive lab equipment (in addition to increasing researchers productivity). We detailed the requirements for such a framework regarding the needs of our case study of the CEA, analyzed performance limitations of the proposed architecture, and pointed to the connection between centralized storage and the processing nodes as the critical point.

In order to alleviate this bottleneck, we investigated using the storage devices of the processing nodes as a cache for the remote storage, and replication strategies to maximize data locality for tasks. A simulator called RepliSim was developed for this research.

7.2. Data Aware Batch Scheduling

We obtained in 2018 two important results on on-line scheduling using resource augmentation. The main idea is that the algorithm is applied to a more powerful environment than that of the adversary. We focused more specifically on the mechanism of rejection based on the concept of duality for mathematical programming applied for the analysis of the algorithm's performance. More precisely, we proposed a primal-dual algorithm for the online scheduling problem of minimizing the total weighted flow time of jobs on unrelated machines when the preemption of jobs is not allowed. This analysis concerned usual sequential jobs. These results have been distinguished among the most significant ones on the annual ACM review of on-line algorithms. We extended this work on a practical side by applying the analysis to actual batch schedulers with parallel jobs, rejection was interpreted as redirecting jobs to some predefined machines.

Machine Learning is a hot topic which received recently a great attention for dealing with the huge amount of data produced by the explosion of the digital applications and for dealing with uncertainties. The members of DataMove promoted a methodology based on simulation and machine learning to obtain efficient dynamic scheduling policies. The main idea is to focus the learning scheme targeting the policies them-selves, and not the specific parameters of the problem. Today, this methodology is mature and it is applied in several project like ANR Energumen (performances and replaced by energy saving). We also launched a new project at MIAI on edge Intelligence. The idea is to propose an alternative to the high-consuming classical IA by doing most of the computations close the the place where the data are produced. We are developing both an efficient task orchestration framework and distributed learning algorithms.

We wrote a survey [20] on scheduling on heterogeneous machines where we provided a complete benchmark suite and we recoded all existing algorithms and compared them.

MARACAS Team

7. New Results

7.1. Results of axis 1: fundamental limits

We worked in 2019 on the following main research directions:

1. Fundamental limits of IoT networks

Table 1.

Principal Investigators:	Malcolm Egan, Samir Perlaza, Jean-Marie Gorce
Students:	Dadja Toussaint Anade-Akbo, Lélio Chetot
Funding:	Orange Labs, ANR Arburst
Partners:	Philippe Mary (IETR, Rennes), Laurent Clavier (IRCICA, Lille) JM Kélif (Orange Labs) H. Vincent Poor (Princeton University, NJ, USA)
Publications:	[34], [48], [36], [49], [35]

One of the main figures of merit in an IoT cell is the capability to support a massive access from distributed nodes, but with very small information quantity [12]. This perspective raises fundamental questions relative to the theoretical limits and performance of this kind of very large scale deployments. Fundamental limits are neither well known nor even well formulated. What is the maximal number of IoT nodes we may deploy in a given environment? At which energetic cost? With which transmission reliability or latency? These multiple questions highlight that the problem is not unique and the capacity is not the only (and even not the main) challenge to be addressed. We aim at establishing the fundamental limits of a decentralized system in a bursty regime which includes short packets of information and impulsive interference regime. We are targeting the fundamental limits and their mathematical expression, according to the usual information theory framework capturing the capacity region by establishing converse and achievability theorems.

2. Stability and sensitivity of fundamental limits

Table 2.

Principal Investigator:	Malcolm Egan, Samir Perlaza
Students:	-
Funding:	
Partners:	H. Vincent Poor, Alex Disto, Princeton University Vyacheslav Kungurtsev, Czech Technical University
Publications:	[8],[33]

The analysis of the fundamental limits on communications systems is performed under some assumptions including Gaussian noise, channel input symbols with average power, among others. Nonetheless, despite that these constraints were well suited for describing communications systems in the early 90's, the evolution of these systems make these assumptions vacuous today. Often, noise is better described by α–stable stochastic processes in IoT networks and channel inputs are subject to constraints in the amplitude, energy harvesting etc. From this perspective, our contributions are based on the notion of capacity sensitivity to study the capacity of continuous memoryless point-to-point channels. The capacity sensitivity reflects how the capacity changes with small perturbations in any of the parameters describing the channel, e.g., cost constraints on the input distribution as well as on the noise distribution.

3. Energy self-sustained wireless networks

Table 3.

Principal Investigator:	Samir Perlaza
Students:	Nizar Khalfet
Funding:	H2020 ComMed
Partners:	I. Kikridis (U. of Cyprus)
Publications:	[29], [42], [43], [50]

The main scientific challenge is to set up a theoretical framework for designing and developing fully decentralized energy-self-sustained communications systems. The main motivation stems from the fact that wireless networks deployed in hard-to-reach places, e.g., remote geographical areas, concrete structures, human body or war zones are often limited by the lifetime of their batteries. This contrasts with the fact that hardware is built to last for very long periods. One of the solutions being considered today for solving the energy limitation problem is the use of energy harvesting (EH) techniques. Within this context, our work focuses on the study of wireless communications systems based on EH sources. EH is expected to be the enabler of energy self-sustainability by eliminating the critical dependence on manual battery recharging.

However, a solid answer on whether or not EH is a viable solution can be given only if the corresponding fundamental limits of data transmission based on EH are known. This is mainly because these limits are based on the laws of Physics and thus, determine the barrier between feasible and unfeasible systems. We study the fundamental limits of three strongly correlated problems regarding the energy supply of future wireless networks: (i) Data transmission over centralized and decentralized EH multi-user channels; (ii) Simultaneous energy and information transmission in multi-user channels; and (iii) Energy cooperation. In a near future, we expect to exploit these results to design algorithms and protocols and later to perform a proof of concept on FIT/CorteXlab. We believe that a solid theoretical framework may help to drive the future design and performance evaluation of applications involving EH based wireless communications systems within smart buildings, smart cities.

4. Security and Privacy

Table 4.

Principal Investigator:	Samir Perlaza
Students:	David Kibloff
Funding:	Inria-DGA PhD
Partners:	Guilaume Villemaud (Socrate), Ligong Wang (ETIS, Cergy) Raphael Shaeffer (TU Berlin)
Publications:	[44], [51]

Information theory is also well adapted to study the fundamental limits of privacy and secrecy. Indeed, the wiretap channel and the covert communication [53] models have been shown to be appropriate for privacy preserving communications in wireless communications. With the PhD of David Kibloff defended in October 2019, we explored the following problem. Given a code used to send a message to two receivers through a degraded discrete memoryless broadcast channel (DM-BC), the sender wishes to alter the codewords to achieve the following goals: (i) the original broadcast communication continues to take place, possibly at the expense of a tolerable increase of the decoding error probability; and (ii) an additional covert message can be transmitted to the stronger receiver such that the weaker receiver cannot detect the existence of this message. The main results are: (a) feasibility of covert communications is proven by using a random coding argument for

general DM-BCs; and (b) necessary conditions for establishing covert communications are described and an impossibility (converse) result is presented for a particular class of DM-BCs. Together, these results characterize the asymptotic fundamental limits of covert communications for this particular class of DM-BCs within an arbitrarily small gap. Future extensions will concern the Gaussian and other continuous channels, or more complex scenarios where some subsets of nodes are willing to communicate while some external observers cannot even detect the existence of these messages. Covert communication allows to introduce a side constraint that prevent a network to be attacked.

5. Structured Codes for Quantization and Channel Estimation

Table 5.

Principal Investigator:	Malcolm Egan
Publications:	[25]

Finite frames are sequences of vectors in finite dimensional Hilbert spaces that play a key role in signal processing and coding theory. In this work, we study the class of tight unit-norm frames for $\mathbb{C}^d$ that also form regular schemes, which we call tight regular schemes (TRS). Many common frames that arising in vector quantization and channel state estimation, such as equiangular tight frames and mutually unbiased bases, fall in this class. We investigate characteristic properties of TRSs and prove that for many constructions, they are intimately connected to weighted 1-designs—arising from cubature rules for integrals over spheres in $\mathbb{C}^d$—with weights dependent on the Voronoi regions of each frame element. Aided by additional numerical evidence, we conjecture that all TRSs in fact satisfy this property.

7.2. Results of axis 2: algorithms

1. Massive random access in LPWAN

Table 6.

Principal Investigator:	Jean-Marie Gorce, Claire Goursaud
Students:	Diane Duchemin, Lélio Chetot
Funding:	ANR Ephyl, Inria-Nokia common lab
Partners:	Sequans, Supelec Rennes, ISEP, CEA Leti, Nokia
Publications:	[30], [31], [37], [47]

The optimization of IoT access techniques was the objective of the ANR Ephyl collaborative project, where we studied different solutions at the PHY and MAC layers as presented in [47].

The main question Maracas group addressed in this research is the detection of simultaneous random transmissions from distributed nodes. The underlying mechanism is a coded slotted Aloha allowing to avoid hand-skake mechanisms. Each node can transmit randomly and the receiver tries to detect several packets simultaneously. Our objective is to identify a good code family, and to determine the fundamental trade-off in terms of nodes density versus reliability. During this year, we focused on the detection of a small subset of simultaneous active nodes, exploiting optimal detection. We developed a MAP based iterative detector at a multi-antennas receiver in [30]. We also proposed a low complexity detector in [37].

This joint coding-decoding optimization problem will be also investigated from extensive simulations and experimental data (see section 3.4), and represents an interesting problem to evaluate deep learning based approaches.

2. Interference management

Table 7.

Principal Investigator:	Léonardo Cardoso, Jean-Marie Gorce
Students:	Hassan Khalam
Funding:	Fed4PMR (PIA)
Partners:	Thales
Publications:	[41]

Interference management and resource management is a very complex problem in wireless environment (e.g. [55]). The capacity region is known for some specific scenarios and some specific channel conditions. But the optimal performance relies on perfect feedback mechanisms, to get channel state information at the transmitters and to coordinate them. As proposed by Jafar et al, topological interference management (TIM) [56] is a seducing framework to balance performance with feedback complexity. In the context of the Fed4PMR project, we develop new algorithms to allow partial coordination between interfering transmitters [41], relying only on some partial interference information. This approach suits particularly well with the requirements of PMR networks, since their deployments is not optimized. The algorithm relies on an association of degrees of freedom evaluation, graph theory and interference alignment.

Based on this first study, we will explore the suitability of TIM in other application scenarios (especially for the standard IEE802.11ax under preparation). For short, TIM allows to build optimal graph representations of a wireless networks, with reduced coordination needs. TIM can be seen as an approach to optimally quantize a complex interfering graph and to distribute its knowledge in an optimal fashion.

3. Learning in radio systems

Table 8.

Principal Investigator:	Léonardo Cardoso, Malcolm Egan, Jean-Marie Gorce
Student:	Cyrille Morin, Mathieu Goutay
Funding:	ADR Analytics, Inria-Nokia common lab AI chair ANR program (applied)
Partners:	Jakob Hoydis, Nokia Bell Labs
Publications:	[45]

Following the artificial intelligence tsunami, the research community in wireless systems (both industry and academia) is engaged in a strong competition to determine how this revolution could change the paradigm of wireless networks. Following the preliminary studies made by Jakob Hoydis [54], we investigate in this research action, the potential of deep learning in radio communications. The central question is to identify which processing could take advantage from neural networks against classical approaches.

Our joint strategy with Nokia follows: we target the production of a huge set of experimental data with FIT/CorteXlab to facilitate the comparison of different solutions and to train neural networks on real data. We currently investigate three original problems : transmitter identification from its RF signature (Cyrille Morin PhD) [45], self-synchronization procedures based on neural networks (Cyrille Morin PhD) and dirty RF compensation (Mathieu Goutay PhD, patents submitted). Last but not least, we believe that an intelligent radio should be able to learn from its environment and to adapt its behavior. Therefore, in the future, we will explore reinforcement principles associated to neural networks and applied to learning based radio.

This topic is very hot, and most top ranked conference have special sessions on this topic. We believe that our partnership with Nokia, our data sets from FIT/CorteXlab and our experience in estimation theory let us be highly competitive.

7.3. Results of axis 3: experimental assessment

During 2019-2020, our experimental work was mostly devoted to the development of new functions of FIT/CorteXlab, and to the development of experimental evaluations with external partners.

1. Development of a user and administrative graphical interface

Table 9.

Principal Investigator:	Pascal Girard, Matthieu Imbert, Léonardo Cardoso
Funding:	FIT/CorteXlab
Partners:	FIT consortium

The objective is to develop a web-based user-friendly interface for using CorteXlab. Several modules are planned and the first module is the user management module, which aims at easing platform usage and improving the metadata that we can associate with each experimenter and experiment. This metadata aims at improving the metrics we can gather about the platform's usage.

2. Development of a docker-based experiment conducting middleware.

Table 10.

Principal Investigator:	Matthieu Imbert, Léonardo Cardoso
Funding:	FIT/CorteXlab
Partners:	FIT consortium

CorteXlab relies on Minus, an experiment conducting middleware which allows users to submit experimental tasks to the platform, handles the automatic execution of these experiments, and gathers their results. The initial design for Minus relies on a fixed toolchain (mainly composed of GNURadio, hardware drivers, and additional external or in-house software or GNURadio blocks, FPGA tools, etc.). Experimenters are supposed to use this fixed toolchain in a batch-like workflow. It is hard for experimenters to extend the limits of the fixed toolchain (e.g. to use a custom library or software, or a different version of GNURadio), and the development phase of an experiment can be painful due to the batch-like interface. To improve this, we have developed a new experimental workflow based on docker [61] images and containers which allows experimenters to use our in-house provided docker images [52], adapt them if needed, or even create completely custom ones. These images have the benefit that they can be used identically on the experimenters' workstations, on the CorteXlab platform, or another platform, and they can be used interactively if needed, even on CorteXlab. This increases greatly the ease of use of the platform, the reproducibility and share-ability of experiments, and the breadth of its usage.

3. Reference scenario for massive IoT access

Table 11.

Principal Investigator:	Othmane Oubejja, Jean-Marie Gorce Matthieu Imbert, Léonardo Cardoso
Funding:	ANR Ephyl, ANR ARburst
Partners:	CEA Leti, Supelec Rennes, Sequans
Publications:	[46]

In this work we developed an experimental setup for dense IoT access evaluation, as part of the project "Enhanced Physical Layer for Cellular IoT" (EPHYL), using FIT/CorteXlab radio testbed. The aim of this work is to provide a customizable and open source design for IoT networks prototyping in a massive multi-user, synchronized and reproducible environment thanks to the

hardware and software capabilities of the testbed. The massive access feature is managed by emulating a base station and several sensors per radio nodes. As shown in Fig.4 , two categories of modular network components are used in our design: a base station unit and a multi-sensor emulator unit. These components are separately hosted in dedicated and remotely accessible radio nodes.

The features of this design can be accessed through customizable demos as documentation and resources are available online. As a result, it is possible for any interested user to plug custom algorithms, evaluate diverse communication scenarios and perform necessary physical measurements.

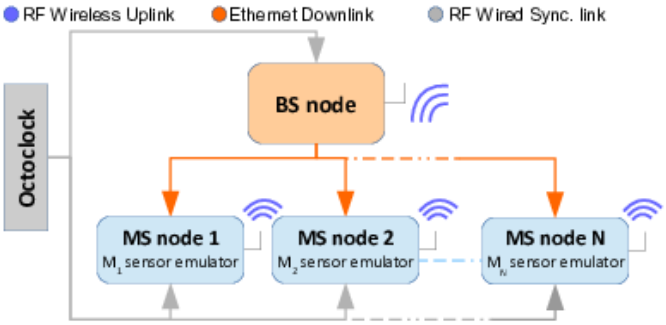


Figure 4. EPHYL IoT network representation

7.4. Results of axis 4: other application fields

1. Smart Grid

Table 12.

Principal Investigators:	Samir Perlaza
Student:	Matei Moldoveanu (visitor)
Partners:	Inaki Esnaola
Publications:	[40]

We study the recovery of missing data from multiple smart grid datasets within a matrix completion framework. The datasets contain the electrical magnitudes required for monitoring and control of the electricity distribution system. Each dataset is described by a low rank matrix. Different datasets are correlated as a result of containing measurements of different physical magnitudes generated by the same distribution system. To assess the validity of matrix completion techniques in the recovery of missing data, we characterize the fundamental limits when two correlated datasets are jointly recovered. We then proceed to evaluate the performance of Singular Value Thresholding (SVT) and Bayesian SVT (BSVT) in this setting. We show that BSVT outperforms SVT by simulating the recovery for different correlated datasets. The performance of BSVT displays the tradeoff behaviour described by the fundamental limit, which suggests that BSVT exploits the correlation between the datasets in an efficient manner.

2. Molecular Communications

Some of the most ambitious applications of molecular communications are expected to lie in nanomedicine and advanced manufacturing. In these domains, the molecular communication system is surrounded by a range of biochemical processes, some of which may be sensitive to chemical species used for communication. Under these conditions, the biological system and the molecular communication system impact each other. As such, the problem of coexistence arises, where both

Table 13.

Principal Investigators:	Malcolm Egan
Postdoc:	Bayram Akdeniz
Funding:	Inria Projet Recherche Exploratoire (PRE)
Partners:	Valeria Loscri (FUN Team, Inria) Marco Di Renzo (CNRS), Bao Tang (University of Graz, Austria) Trung Duong (Queen's University Belfast) Ido Nevat (TUMCREATE, Singapore)
Publications:	[38], [39], [24], [26]

the reliability of the molecular communication system and the function of the biological system must be ensured. In this paper, we study this problem with a focus on interactions with biological systems equipped with chemosensing mechanisms, which arises in a large class of biological systems. We motivate the problem by considering chemosensing mechanisms arising in bacteria chemo-taxis, a ubiquitous and well-understood class of biological systems. We then propose strategies for a molecular communication system to minimize disruption of biological system equipped with a chemosensing mechanism. This is achieved by exploiting tools from the theory of chemical reaction networks. To investigate the capabilities of our strategies, we obtain fundamental information theoretic limits by establishing a new connection with the problem of covert communications.

3. Intelligent Transportation

Table 14.

Principal Investigators:	Malcolm Egan
Partners:	Michel Jakob (Czech Technical University in Prague), Nir Oren (University of Aberdeen)
Publications:	[27]

Market mechanisms are now playing a key role in allocating and pricing on-demand transportion services. In practice, most such services use posted-price mechanisms, where both passengers and drivers are offered a journey price which they can accept or reject. However, providers such as Liftago and GrabTaxi have begun to adopt a mechanism whereby auctions are used to price drivers. These latter mechanisms are neither posted-price nor classical double auctions, and can instead be considered a hybrid mechanism. In this work, we develop and study the properties of a novel hybrid on-demand transport mechanism. Due to the need for incorporating statistical knowledge and communication of system state information, communication-theoretic methods can play a useful role.

In particular, as these mechanisms require knowledge of passenger demand, we analyze the data-profit tradeoff as well as how passenger and driver preferences influence mechanism performance. We show that the revenue loss for the provider scales with $\sqrt{n \log n}$ for n passenger requests under a multi-armed bandit learning algorithm with beta distributed preferences. We also investigate the effect of subsidies on both profit and the number of successful journeys allocated by the mechanism, comparing these with a posted-price mechanism, showing improvements in profit with a comparable number of successful requests.

POLARIS Project-Team

7. New Results

7.1. Design of Experiments

Performance engineering of scientific HPC applications requires to measure repeatedly the performance of applications or of computation kernels, which consume a large amount of time and resources. It is essential to design experiments so as to reduce this cost as much as possible. Our contribution along this axis is twofold: (1) the investigation sound exploration techniques and (2) the control of experiments to ensure the measurements are as representative as possible of real workload.

Writing, porting, and optimizing scientific applications makes autotuning techniques fundamental to lower the cost of leveraging the improvements on execution time and power consumption provided by the latest software and hardware platforms. Despite the need for economy, most autotuning techniques still require large budgets of costly experimental measurements to provide good results, while rarely providing exploitable knowledge after optimization. In [16], we investigate the use of *Design of Experiments* to propose a user-transparent autotuning technique that operates under tight budget constraints by significantly reducing the measurements needed to find good optimizations. Our approach enables users to make informed decisions on which optimizations to pursue and when to stop. We present an experimental evaluation of our approach and show it is capable of leveraging user decisions to find the best global configuration of a GPU Laplacian kernel using half of the measurement budget used by other common autotuning techniques. We show that our approach is also capable of finding speedups of up to $50\times$, compared to gcc's -O3, for some kernels from the SPAPT benchmark suite, using up to $10\times$ fewer measurements than random sampling. Although the results are very encouraging, our approach relies on assumptions on the geometry of the search space that are difficult to test in very large dimension. We are thus currently pursuing this line of research using non parametric approaches based on gaussian process regression, space filling designs and iteratively selecting configurations that yield the best expected improvement.

Our second contribution is related to the control of measurements. In [40], we relate a surprising observation on the performance of the highly optimized and regular DGEMM function on modern processors. The DGEMM function is a widely used implementation of the matrix product. While the asymptotic complexity of the algorithm only depends on the sizes of the matrices, we show that the performance is significantly impacted by the matrices content. Although it would be expected that special values like 1 or 0 may yield to specific behevior, we show that arbitrary constant values are no different and that random values incur a significant performance drop. Our experiments show that this may be due to bit flips in the CPU causing an energy consumption overhead. Such phenomenon reminds the importance of thoroughly randomizing every single parameter of experiments to avoid bias toward specific behavior.

7.2. Predictive Simulation of HPC Applications

Finely tuning MPI applications (number of processes, granularity, collective operation algorithms, topology and process placement) is critical to obtain good performance on supercomputers. With a rising cost of modern supercomputers, running parallel applications at scale solely to optimize their performance is extremely expensive. Using SimGrid, we work toward providing a methodology allowing to provide inexpensive but faithful predictions of expected performance.

The methodology we propose relies on SimGrid/SMPI and captures the complexity of adaptive applications by emulating the MPI code while skipping insignificant parts. In [18] we demonstrate its capability with High Performance Linpack (HPL), the benchmark used to rank supercomputers in the TOP500 and which requires a careful tuning. We explain (1) how we both extended the SimGrid's SMPI simulator and slightly modified the open-source version of HPL to allow a fast emulation on a single commodity server at the scale of a

supercomputer and (2) how to model the different components (network, BLAS, ...) of the system. We show that a careful modeling of both spatial and temporal node variability allows us to obtain predictions within a few percents of real experiments. The modeling of BLAS operations is particularly important and we have thus started investigating in the context of simulating a sparse direct solver how to automatically performance models for commonly used BLAS kernels [33]. A key difficulty remains the acquisition of faithful performance measurements as modern processors are often quite unstable. This effort is therefore particularly related to the aforementioned "Design of Experiments" line of research.

7.3. Simulation of Smart Grids

In [35], we present ASGriDS, an asynchronous Smart Grid simulation framework. ASGriDS is multi-domain, it simultaneously models the power network along with its physical loads/generators, controllers, and communication infrastructure. ASGriDS provides a unified workflow in a pythonic environment, to describe, run and control complex SmartGrid deployment scenarios. ASGriDS is an event-driven simulator that can run in either real-time or accelerated real-time. As it is modular and its components interact asynchronously, it can run either locally on a distributed infrastructure, also in hardware-in-the- loop setups, and on top of emulated/physical communication links. In this paper, we present the design of our simulator and we demonstrate its use with a generation control problem on a low voltage network. We use ASGriDS to deploy a real-time controller based on optimal power flow, on top of TCP and UDP based communication network, under various packet loss conditions.

7.4. Batch Scheduling

Despite the impressive growth and size of super-computers, the computational power they provide still cannot match the demand. Efficient and fair resource allocation is a critical task. Super-computers use Resource and Job Management Systems to schedule applications, which is generally done by relying on generic index policies such as First Come First Served and Shortest Processing time First in combination with Backfilling strategies. Unfortunately, such generic policies often fail to exploit specific characteristics of real workloads.

In [36], we focus on improving the performance of online schedulers by studying mixed policies, which are created by combining multiple job characteristics in a weighted linear expression, as opposed to classical pure policies which use only a single characteristic. This larger class of scheduling policies aims at providing more flexibility and adaptability. We use space coverage and black-box optimization techniques to explore this new space of mixed policies and we study how can they adapt to the changes in the workload. We perform an extensive experimental campaign through which we show that (1) the best pure policy is far from optimal and that (2) using a carefully tuned mixed policy would allow to significantly improve the performance of the system. (3) We also provide empirical evidence that there is no one size fits all policy, by showing that the rapid workload evolution seems to prevent classical online learning algorithms from being effective.

A careful investigation of why such mixed strategy fail to globally exploit weekly workload features reveal that some users sometimes provide widely inaccurate information, which dramatically fools the batch scheduling heuristic. Indeed, users typically provide a loose upper bound estimate for job execution times that are hardly useful. Previous studies attempted to improve these estimates using regression techniques. Although these attempts provide reasonable predictions, they require a long period of training data. Furthermore, aiming for perfect prediction may be of limited use for scheduling purposes. In [50], we propose a simpler approach by classifying jobs as small or large and prioritizing the execution of small jobs over large ones. Indeed, small jobs are the most impacted by queuing delays but they typically represent a light load and incur a small burden on the other jobs. The classifier operates online and learns by using data collected over the previous weeks, facilitating its deployment and enabling fast adaptations to changes in workload characteristics. We evaluate our approach using four scheduling policies on six HPC platform workload traces. We show that: (i) incorporating such classification significantly reduces the average bounded slowdown of jobs in all scenarios, and (ii) the obtained improvements are comparable, in most scenarios, to the ideal hypothetical situation where the scheduler would know the exact running time of jobs in advance.

7.5. Load Balancing

In distributed systems, load balancing is a powerful concept to improve the distribution of jobs across multiple computing resources and to control performance metrics such as delays and throughputs while avoiding the overload of any single resource. This section describes three contributions:

- In multi-server distributed queueing systems, the access of stochastically arriving jobs to resources is often regulated by a dispatcher, also known as load balancer. A fundamental problem consists in designing a load balancing algorithm that minimizes the delays experienced by jobs. During the last two decades, the power-of-d-choice algorithm, based on the idea of dispatching each job to the least loaded server out of d servers randomly sampled at the arrival of the job itself, has emerged as a breakthrough in the foundations of this area due to its versatility and appealing asymptotic properties. In [8], we consider the power-of-d-choice algorithm with the addition of a local memory that keeps track of the latest observations collected over time on the sampled servers. Then, each job is sent to a server with the lowest observation. We show that this algorithm is asymptotically optimal in the sense that the load balancer can always assign each job to an idle server in the large-system limit. This holds true if and only if the system load λ is less than $1 - 1/d$. If this condition is not satisfied, we show that queue lengths are tightly bounded by $\lceil \frac{-\log(1-\lambda)}{\log(\lambda d+1)} \rceil$. This is in contrast with the classic version of the power-of-d-choice algorithm, where at the fluid scale a strictly positive proportion of servers containing i jobs exists for all $i \geq 0$, in equilibrium. Our results quantify and highlight the importance of using memory as a means to enhance performance in randomized load balancing.
- When dispatching jobs to parallel servers, or queues, the highly scalable round-robin (RR) scheme reduces the variance of interarrival times at all queues to a great extent but has no impact on the variances of service processes. Contrariwise, size-interval task assignment (SITA) routing has little impact on the variances of interarrival times but makes the service processes as deterministic as possible. In [6], we unify both 'static' approaches to design a scalable load balancing framework able to control the variances of the arrival and service processes jointly. It turns out that the resulting combination significantly improves performance and is able to drive the mean job delay to zero in the large-system limit; it is known that this property is not achieved when both approaches are considered separately. Within realistic parameters, we show that the optimal number of size intervals that partition the support of the job size distribution is small with respect to the system size. This enhances the applicability of the proposed load balancing scheme at a large scale. In fact, we find that adding a little bit of information about job sizes to a dispatcher operating under RR improves performance a lot. Under the optimal scaling of size intervals and assuming highly variable job sizes, numerical simulations indicate that the proposed algorithm is competitive with the (less scalable) join-the-shortest-workload algorithm even when the system size grows large.
- Size-based routing provides robust strategies to improve the performance of computer and communication systems with highly variable workloads because it is able to isolate small jobs from large ones in a static manner. The basic idea is that each server is assigned all jobs whose sizes belong to a distinct and continuous interval. In the literature, dispatching rules of this type are referred to as SITA (Size Interval Task Assignment) policies. Though their evident benefits, the problem of finding a SITA policy that minimizes the overall mean (steady-state) waiting time is known to be intractable. In particular it is not clear when it is preferable to balance or unbalance server loads and, in the latter case, how. In [7], we provide an answer to these questions in the celebrated limiting regime where the system capacity grows linearly with the system demand to infinity. Within this framework, we prove that the minimum mean waiting time achievable by a SITA policy necessarily converges to the mean waiting time achieved by SITA-E, the SITA policy that equalizes server loads, provided that servers are homogeneous. However, within the set of SITA policies we also show that SITA-E can perform arbitrarily bad if servers are heterogeneous. In this case we prove that there exist exactly C! asymptotically optimal policies, where C denotes the number of server types, and all of them are linked to the solution of a single strictly convex optimization problem. It turns out that the mean

waiting time achieved by any of such asymptotically optimal policies does not depend on how job-size intervals are mapped to servers. Our theoretical results are validated by numerical simulations with respect to realistic parameters and suggest that the above insights are also accurate in small systems composed of a few servers, i.e., ten.

7.6. FoG Computing

To this day, the Internet of Things (IoT) continues its explosive growth. Nevertheless, with the exceptional evolution of traffic demand, existing infrastructures are struggling to resist. In this context, Fog computing is shaping the future of IoT applications. It offers nearby computational, networking and storage resources to respond to the stringent requirements of these applications. However, despite its several advantages, Fog computing raises new challenges which slow its adoption down. Hence, there is a lack of practical solutions to enable the exploitation of this novel concept.

In [19], we propose FITOR, an orchestration system for IoT applications in the Fog environment. This solution builds a realistic Fog environment while offering efficient orchestration mechanisms. In order to optimize the provisioning of Fog-Enabled IoT applications, FITOR relies on O-FSP, an optimized fog service provisioning strategy which aims to minimize the provisioning cost of IoT applications, while meeting their requirements. Based on extensive experiments, the results obtained show that O-FSP optimizes the placement of IoT applications and outperforms the related strategies in terms of i) provisioning cost ii) resource usage and iii) acceptance rate. In [46], we propose a novel strategy, which we call GO-FSP and which optimizes the placement of IoT application components while coping with their strict performance requirements. To do so, we first propose an Integer Linear Programming (ILP) formulation for the IoT application provisioning problem. The latter targets to minimize the deployment cost while ensuring a load balancing between heterogeneous devices. Then, a GRASP-based approach is proposed to achieve the aforementioned objectives. Finally, we make use of the FITOR orchestration system to evaluate the performance of our solution under real conditions. Obtained results show that our scheme outperforms the related strategies. We are currently comparing such strategy with other strategies based on online learning mechanisms under various information scenarios (delayed and noisy feedback, inaccurate application load information, etc.).

Last, fog computing also extends the capacities of the cloud to the edge of the network, near the physical world, so that Internet of Things (IoT) applications can benefit from properties such as short delays, real-time and privacy. Unfortunately, devices in the Fog-IoT environment are usually unstable and prone to failures. In this context, the consequences of failures may impact the physical world and can, therefore, be critical. In [28], we present a framework for end-to-end resilience of Fog-IoT applications. The framework was implemented and experimented on a smart home testbed.

7.7. Research Management: Research Reproducibility and Credit

We are actively promoting better research practices, in particular in term of research reproducibility and contribution recognition. Our contribution this year is threefold

First, we have participated to the writing of a book introducing reproducible research [39]. For a researcher, there is nothing more frustrating than the failure to reproduce major results obtained a few months back. The causes of such disappointments can be multiple and insidious. This phenomenon plays an important role in the so-called "research reproducibility crisis". This book takes a current perspective onto a number of potentially dangerous situations and practices, to examplify and highlight the symptoms of non-reproducibility in research. Each time, it provides efficient solutions ranging from good-practices that are easily and immediately implementable to more technical tools, all of which are free and have been put to the test by the authors themselves. Students and engineers and researchers should find efficient and accessible ways leading them to improve their reproducible research practices.

Second, to allow students and engineers and researchers to receive proper training in reproducible research, we have run the second session of the Mooc "Reproducible research: Methodological principles for a transparent science" on the FUN platform from April, 1 to June, 13 2019. This MOOC allows scientists to learn modern and reliable tools such as Markdown for taking structured notes, Desktop search applications, GitLab for version control and collaborative working, and Computational notebooks (Jupyter, RStudio, and Org-Mode) for efficiently combining the computation, presentation, and analysis of data. More than 2,100 persons registered to this session and we are currently working on a third session which is expected to start in the beginning of the year 2020.

Third, software is a fundamental pillar of modern scientific research, not only in computer science, but actually across all fields and disciplines. However, there is a lack of adequate means to cite and reference software, for many reasons. An obvious first reason is software authorship, which can range from a single developer to a whole team, and can even vary in time. The panorama is even more complex than that, because many roles can be involved in software development: software architect, coder, debugger, tester, team manager, and so on. Arguably, the researchers who have invented the key algorithms underlying the software can also claim a part of the authorship. And there are many other reasons that make this issue complex. We provide in [5] a contribution to the ongoing efforts to develop proper guidelines and recommendations for software citation, building upon the internal experience of Inria, the French research institute for digital sciences. As a central contribution, we make three key recommendations. (1) We propose a richer taxonomy for software contributions with a qualitative scale. (2) We claim that it is essential to put the human at the heart of the evaluation. And (3) we propose to distinguish citation from reference which is particularly important in the context of reproducible research.

7.8. Mean Field Games and Control

In [10], we consider mean field games with discrete state spaces (called discrete mean field games in the following) and we analyze these games in continuous and discrete time, over finite as well as infinite time horizons. We prove the existence of a mean field equilibrium assuming continuity of the cost and of the drift. These conditions are more general than the existing papers studying finite state space mean field games. Besides, we also study the convergence of the equilibria of N -player games to mean field equilibria in our four settings. On the one hand, we define a class of strategies in which any sequence of equilibria of the finite games converges weakly to a mean field equilibrium when the number of players goes to infinity. On the other hand, we exhibit equilibria outside this class that do not converge to mean field equilibria and for which the value of the game does not converge. In discrete time this non-convergence phenomenon implies that the Folk theorem does not scale to the mean field limit.

In [20], we consider a class of nonlinear systems of differential equations with uncertainties, i.e., with lack of knowledge in some of the parameters that is represented by a time-varying unknown bounded functions. An under-approximation of such systems consists of a subset of its reachable set, for any value of the unknown parameters. By relying on optimal control theory through Pontryagin's principle, we provide an algorithm for the under-approximation of a linear combination of the state variables in terms of a fully automated tool-chain named UTOPIC. This allows to establish tight under-approximations of common benchmarks models with dimensions as large as sixty-five.

7.9. Energy and Network Optimization

This section describes four contributions on energy and network optimization.

- One of the key challenges in Internet of Things (IoT) networks is to connect many different types of autonomous devices while reducing their individual power consumption. This problem is exacerbated by two main factors: first, the fact that these devices operate in and give rise to a highly dynamic and unpredictable environment where existing solutions (e.g., water-filling algorithms) are no longer relevant; and second, the lack of sufficient information at the device end. To address these issues, we propose a regret-based formulation that accounts for arbitrary network dynamics: this allows us to derive an online power control scheme that is provably capable of adapting to such

changes, while relying solely on strictly causal feedback. In so doing, we identify an important tradeoff between the amount of feedback available at the transmitter side and the resulting system performance: if the device has access to unbiased gradient observations, the algorithm's regret after T stages is $O(T^{-1/2})$ (up to logarithmic factors); on the other hand, if the device only has access to scalar, utility-based information, this decay rate drops to $O(T^{-1/4})$. The above is validated by an extensive suite of numerical simulations in realistic channel conditions, which clearly exhibit the gains of the proposed online approach over traditional water-filling methods. This contribution appeared in [11].

- Many businesses possess a small infrastructure that they can use for their computing tasks, but also often buy extra computing resources from clouds. Cloud vendors such as Amazon EC2 offer two types of purchase options: on-demand and spot instances. As tenants have limited budgets to satisfy their computing needs, it is crucial for them to determine how to purchase different options and utilize them (in addition to possible self-owned instances) in a cost-effective manner while respecting their response-time targets. In this paper, we propose a framework to design policies to allocate self-owned, on-demand and spot instances to arriving jobs. In particular, we propose a near-optimal policy to determine the number of self-owned instances and an optimal policy to determine the number of on-demand instances to buy and the number of spot instances to bid for at each time unit. Our policies rely on a small number of parameters and we use an online learning technique to infer their optimal values. Through numerical simulations, we show the effectiveness of our proposed policies, in particular that they achieve a cost reduction of up to 64.51% when spot and on-demand instances are considered and of up to 43.74% when self-owned instances are considered, compared to previously proposed or intuitive policies. This contribution appeared in [13].
- In [22], we consider the classical problem of minimizing offline the total energy consumption required to execute a set of n real-time jobs on a single processor with varying speed. Each real-time job is defined by its release time, size, and deadline (all integers). The goal is to find a sequence of processor speeds, chosen among a finite set of available speeds, such that no job misses its deadline and the energy consumption is minimal. Such a sequence is called an optimal speed schedule. We propose a linear time algorithm that checks the schedulability of the given set of n jobs and computes an optimal speed schedule. The time complexity of our algorithm is in $O(n)$, to be compared with $O(n \log(n))$ for the best known solutions. Besides the complexity gain, the main interest of our algorithm is that it is based on a completely different idea: instead of computing the critical intervals, it sweeps the set of jobs and uses a dynamic programming approach to compute an optimal speed schedule. Our linear time algorithm is still valid (with some changes) with an arbitrary power function (not necessarily convex) and arbitrary switching times
- Network utility maximization (NUM) is an iconic problem in network traffic management which is at the core of many current and emerging network design paradigms - and, in particular, software-defined networks (SDNs). Thus, given the exponential growth of modern-day networks (in both size and complexity), it is crucial to develop scalable algorithmic tools that are capable of providing efficient solutions in time which is dimension-free, i.e., independent-or nearly-independent-on the size of the system. To do so, we leverage a suite of modified gradient methods known as "mirror descent" and we derive a scalable and efficient algorithm for the NUM problem based on gradient exponentiation. We show that the convergence speed of the proposed algorithm only carries a logarithmic dependence on the size of the network, so it can be implemented reliably and efficiently in massively large networks where traditional gradient methods are prohibitively slow. These theoretical results are sub-sequently validated by extensive numerical simulations showing an improvement of several order of magnitudes over standard gradient methods in large-scale networks. This contribution appeared in [31].
- In the DNS resolution process, packet losses and ensuing retransmission timeouts induce marked latencies: the current UDP-based resolution process takes up to 5 seconds to detect a loss event. In [24], [24], we find that persistent DNS connections based on TCP or TLS can provide an elegant solution to this problem. With controlled experiments on a testbed, we show that persistent DNS

connections significantly reduces worst-case latency. We then leverage a large-scale platform to study the performance impact of TCP/TLS on recursive resolvers. We find that off-the-shelf software and reasonably powerful hardware can effectively provide recursive DNS service over TCP and TLS, with a manageable performance hit compared to UDP.

7.10. Privacy, Fairness, and Transparency in Online Social Medias

This section describes four contributions on privacy, fairness and transparency in online social medias

- The Facebook advertising platform has been subject to a number of controversies in the past years regarding privacy violations, lack of transparency, as well as its capacity to be used by dishonest actors for discrimination or propaganda. In this study, we aim to provide a better understanding of the Facebook advertising ecosystem, focusing on how it is being used by advertisers. We first analyze the set of advertisers and then investigate how those advertisers are targeting users and customizing ads via the platform. Our analysis is based on the data we collected from over 600 real-world users via a browser extension that collects the ads our users receive when they browse their Facebook timeline, as well as the explanations for why users received these ads. Our results reveal that users are targeted by a wide range of advertisers (e.g., from popular to niche advertisers); that a non-negligible fraction of advertisers are part of potentially sensitive categories such as news and politics, health or religion; that a significant number of advertisers employ targeting strategies that could be either invasive or opaque; and that many advertisers use a variety of targeting parameters and ad texts. Overall, our work emphasizes the need for better mechanisms to audit ads and advertisers in social media and provides an overview of the platform usage that can help move towards such mechanisms.

  This contribution appeared in [14].

- To help their users to discover important items at a particular time, major websites like Twitter, Yelp, TripAdvisor or NYTimes provide Top-K recommendations (e.g., 10 Trending Topics, Top 5 Hotels in Paris or 10 Most Viewed News Stories), which rely on crowd-sourced popularity signals to select the items. However, diferent sections of a crowd may have diferent preferences, and there is a large silent majority who do not explicitly express their opinion. Also, the crowd often consists of actors like bots, spammers, or people running orchestrated campaigns. Recommendation algorithms today largely do not consider such nuances, hence are vulnerable to strategic manipulation by small but hyper-active user groups. To fairly aggregate the preferences of all users while recommending top-K items, we borrow ideas from prior research on social choice theory, and identify a voting mechanism called Single Trans-ferable Vote (STV) as having many of the fairness properties we desire in top-K item (s)elections. We develop an innovative mechanism to attribute preferences of silent majority which also make STV completely operational. We show the generalizability of our approach by implementing it on two diferent real-world datasets. Through extensive experimentation and comparison with state-of-the-art techniques, we show that our proposed approach provides maximum user satisfaction, and cuts down drastically on items disliked by most but hyper-actively promoted by a few users.

  This contribution appeared in [17].

- The rise of algorithmic decision making led to active researches on how to define and guarantee fairness, mostly focusing on one-shot decision making. In several important applications such as hiring, however, decisions are made in multiple stage with additional information at each stage. In such cases, fairness issues remain poorly understood. In this paper we study fairness in k-stage selection problems where additional features are observed at every stage. We first introduce two fairness notions, local (per stage) and global (final stage) fairness, that extend the classical fairness notions to the k-stage setting. We propose a simple model based on a probabilistic formulation and show that the locally and globally fair selections that maximize precision can be computed via a linear program. We then define the price of local fairness to measure the loss of precision induced by local constraints; and investigate theoretically and empirically this quantity. In particular, our experiments show that the price of local fairness is generally smaller when the sensitive attribute

is observed at the first stage; but globally fair selections are more locally fair when the sensitive attribute is observed at the second stage—hence in both cases it is often possible to have a selection that has a small price of local fairness and is close to locally fair.

This contribution appeared in [21].

- Most social platforms offer mechanisms allowing users to delete their posts, and a significant fraction of users exercise this right to be forgotten. However, ironically, users' attempt to reduce attention to sensitive posts via deletion, in practice, attracts unwanted attention from stalkers specifically to those (deleted) posts. Thus, deletions may leave users more vulnerable to attacks on their privacy in general. Users hoping to make their posts forgotten face a "damned if I do, damned if I don't" dilemma. Many are shifting towards ephemeral social platform like Snapchat, which will deprive us of important user-data archival. In the form of intermittent withdrawals, we present, Lethe, a novel solution to this problem of (really) forgetting the forgotten. If the next-generation social platforms are willing to give up the uninterrupted availability of non-deleted posts by a very small fraction, Lethe provides privacy to the deleted posts over long durations. In presence of Lethe, an adversarial observer becomes unsure if some posts are permanently deleted or just temporarily withdrawn by Lethe; at the same time, the adversarial observer is overwhelmed by a large number of falsely flagged un-deleted posts. To demonstrate the feasibility and performance of Lethe, we analyze large-scale real data about users' deletion over Twitter and thoroughly investigate how to choose time duration distributions for alternating between temporary withdrawals and resurrections of non-deleted posts. We find a favorable trade-off between privacy, availability and adversarial overhead in different settings for users exercising their right to delete. We show that, even against an ultimate adversary with an uninterrupted access to the entire platform, Lethe offers deletion privacy for up to 3 months from the time of deletion, while maintaining content availability as high as 95% and keeping the adversarial precision to 20%.

  This contribution appeared in [27],

7.11. Optimization Methods

This section describes six contributions on optimization.

- In [9], we propose an interior-point method for linearly constrained – and possibly nonconvex – optimization problems. The proposed method – which we call the Hessian barrier algorithm (HBA) – combines a forward Euler discretization of Hessian Riemannian gradient flows with an Armijo backtracking step-size policy. In this way, HBA can be seen as an alternative to mirror descent (MD), and contains as special cases the affine scaling algorithm, regularized Newton processes, and several other iterative solution methods. Our main result is that, modulo a non-degeneracy condition, the algorithm converges to the problem's critical set; hence, in the convex case, the algorithm converges globally to the problem's minimum set. In the case of linearly constrained quadratic programs (not necessarily convex), we also show that the method's convergence rate is $O(1/k^\rho)$ for some $\rho \in (0, 1]$ that depends only on the choice of kernel function (i.e., not on the problem's primitives). These theoretical results are validated by numerical experiments in standard non-convex test functions and large-scale traffic assignment problems.
- In[15], Lipschitz continuity is a central requirement for achieving the optimal $O(1/T)$ rate of convergence in monotone, deterministic variational inequalities (a setting that includes convex minimization, convex-concave optimization, nonatomic games, and many other problems). However, in many cases of practical interest, the operator defining the variational inequality may exhibit singularities at the boundary of the feasible region, precluding in this way the use of fast gradient methods that attain this optimal rate (such as Nemirovski's mirror-prox algorithm and its variants). To address this issue, we propose a novel regularity condition which we call Bregman continuity, and which relates the variation of the operator to that of a suitably chosen Bregman function. Leveraging this condition, we derive an adaptive mirror-prox algorithm which attains the optimal $O(1/T)$ rate of convergence in problems with possibly singular operators, without any prior knowledge of the

degree of smoothness (the Bregman analogue of the Lipschitz constant). We also show that, under Bregman continuity, the mirror-prox algorithm achieves a $O(1/\sqrt{T})$ convergence rate in stochastic variational inequalities.

- In [23] Variational inequalities have recently attracted considerable interest in machine learning as a flexible paradigm for models that go beyond ordinary loss function minimization (such as generative adversarial networks and related deep learning systems). In this setting, the optimal O(1/t) convergence rate for solving smooth monotone variational inequalities is achieved by the Extra-Gradient (EG) algorithm and its variants. Aiming to alleviate the cost of an extra gradient step per iteration (which can become quite substantial in deep learning applications), several algorithms have been proposed as surrogates to Extra-Gradient with a *single* oracle call per iteration. In this paper, we develop a synthetic view of such algorithms, and we complement the existing literature by showing that they retain a O(1/t) ergodic convergence rate in smooth, deterministic problems. Subsequently, beyond the monotone deterministic case, we also show that the last iterate of single-call, *stochastic* extra-gradient methods still enjoys a O(1/t) local convergence rate to solutions of non-monotone variational inequalities that satisfy a second-order sufficient condition.
- In [25], we study a class of online convex optimization problems with long-term budget constraints that arise naturally as reliability guarantees or total consumption constraints. In this general setting, prior work by Mannor et al. (2009) has shown that achieving no regret is impossible if the functions defining the agent's budget are chosen by an adversary. To overcome this obstacle, we refine the agent's regret metric by introducing the notion of a "K-benchmark", i.e., a comparator which meets the problem's allotted budget over any window of length K. The impossibility analysis of Mannor et al. (2009) is recovered when $K = T$; however, for $K = o(T)$, we show that it is possible to minimize regret while still meeting the problem's long-term budget constraints. We achieve this via an online learning policy based on Cautious Online Lagrangiant Descent (COLD) for which we derive explicit bounds, in terms of both the incurred regret and the residual budget violations.
- In [26], owing to their connection with generative adversarial networks (GANs), saddle-point problems have recently attracted considerable interest in machine learning and beyond. By necessity, most theoretical guarantees revolve around convex-concave (or even linear) problems; however, making theoretical inroads towards efficient GAN training depends crucially on moving beyond this classic framework. To make piecemeal progress along these lines, we analyze the behavior of mirror descent (MD) in a class of non-monotone problems whose solutions coincide with those of a naturally associated variational inequality - a property which we call coherence. We first show that ordinary, "vanilla" MD converges under a strict version of this condition, but not otherwise; in particular, it may fail to converge even in bilinear models with a unique solution. We then show that this deficiency is mitigated by optimism: by taking an "extra-gradient" step, optimistic mirror descent (OMD) converges in all coherent problems. Our analysis generalizes and extends the results of Daskalakis et al. (2018) for optimistic gradient descent (OGD) in bilinear problems, and makes concrete headway for establishing convergence beyond convex-concave games. We also provide stochastic analogues of these results, and we validate our analysis by numerical experiments in a wide array of GAN models (including Gaussian mixture models, as well as the CelebA and CIFAR-10 datasets).
- In [30], we develop a new stochastic algorithm with variance reduction for solving pseudo-monotone stochastic variational inequalities. Our method builds on Tseng's forward-backward-forward algorithm, which is known in the deterministic literature to be a valuable alternative to Korpelevich's extragradient method when solving variational inequalities over a convex and closed set governed with pseudo-monotone and Lipschitz continuous operators. The main computational advantage of Tseng's algorithm is that it relies only on a single projection step, and two independent queries of a stochastic oracle. Our algorithm incorporates a variance reduction mechanism, and leads to a.s. convergence to solutions of a merely pseudo-monotone stochastic variational inequality problem. To the best of our knowledge, this is the first stochastic algorithm achieving this by using only a single projection at each iteration.

7.12. Learning

This section describes three contributions on machine learning.

- In [12], we examine the convergence of no-regret learning in games with continuous action sets. For concreteness, we focus on learning via "dual averaging", a widely used class of no-regret learning schemes where players take small steps along their individual payoff gradients and then "mirror" the output back to their action sets. In terms of feedback, we assume that players can only estimate their payoff gradients up to a zero-mean error with bounded variance. To study the convergence of the induced sequence of play, we introduce the notion of variational stability, and we show that stable equilibria are locally attracting with high probability whereas globally stable equilibria are globally attracting with probability 1. We also discuss some applications to mixed-strategy learning in finite games, and we provide explicit estimates of the method's convergence speed.
- Resource allocation games such as the famous Colonel Blotto (CB) and Hide-and-Seek (HS) games are often used to model a large variety of practical problems, but only in their one-shot versions. Indeed, due to their extremely large strategy space, it remains an open question how one can efficiently learn in these games. In this work, we show that the online CB and HS games can be cast as path planning problems with side-observations (SOPPP): at each stage, a learner chooses a path on a directed acyclic graph and suffers the sum of losses that are adversarially assigned to the corresponding edges; and she then receives semi-bandit feedback with side-observations (i.e., she observes the losses on the chosen edges plus some others). We propose a novel algorithm, EXP3-OE, the first-of-its-kind with guaranteed efficient running time for SOPPP without requiring any auxiliary oracle. We provide an expected-regret bound of EXP3-OE in SOPPP matching the order of the best benchmark in the literature. Moreover, we introduce additional assumptions on the observability model under which we can further improve the regret bounds of EXP3-OE. We illustrate the benefit of using EXP3-OE in SOPPP by applying it to the online CB and HS games.

  This contribution appeared in [29], [49]. In an earlier article [38], we also looked at the sequential Colonel Blotto game under bandit feedback and we proposed a blackbox optimization based method to optimize the exploration distribution of the classical COMBAND algorithm.
- In [32], we study nonzero-sum hypothesis testing games that arise in the context of adversarial classification, in both the Bayesian as well as the Neyman-Pearson frameworks. We first show that these games admit mixed strategy Nash equilibria, and then we examine some interesting concentration phenomena of these equilibria. Our main results are on the exponential rates of convergence of classification errors at equilibrium, which are analogous to the well-known Chernoff-Stein lemma and Chernoff information that describe the error exponents in the classical binary hypothesis testing problem, but with parameters derived from the adversarial model. The results are validated through numerical experiments.

ROMA Project-Team

7. New Results

7.1. Creation of the start-up "Mumps Technologies SAS"

In January 2019, Jean-Yves L'Excellent left the ROMA team to co-found with Patrick Amestoy and Chiara Puglisi the company "Mumps Technologies" around the free software library MUMPS (Cecill-C licence). MUMPS solves large systems of sparse linear equations on high-performance computers in a robust and effective way. Mumps Technologies carries on collaborations and R&D activities to keep the MUMPS software library state-of-the-art and freely available, while offering to its clients a set of services.

7.2. Scheduling independent stochastic tasks under deadline and budget constraints

This work discusses scheduling strategies for the problem of maximizing the expected number of tasks that can be executed on a cloud platform within a given budget and under a deadline constraint. The execution times of tasks follow IID probability laws. The main questions are how many processors to enroll and whether and when to interrupt tasks that have been executing for some time. We provide complexity results and an asymptotically optimal strategy for the problem instance with discrete probability distributions and without deadline. We extend the latter strategy for the general case with continuous distributions and a deadline and we design an efficient heuristic which is shown to outperform standard approaches when running simulations for a variety of useful distribution laws.

The findings were published in a journal [8].

7.3. Online scheduling of task graphs on heterogeneous platforms

Modern computing platforms commonly include accelerators. We target the problem of scheduling applications modeled as task graphs on hybrid platforms made of two types of resources, such as CPUs and GPUs. We consider that task graphs are uncovered dynamically, and that the scheduler has information only on the available tasks, i.e., tasks whose predecessors have all been completed. Each task can be processed by either a CPU or a GPU, and the corresponding processing times are known. Our study extends a previous $4\sqrt{m/k}$-competitive online algorithm by Amaris et al. [46], where m is the number of CPUs and k the number of GPUs ($m \geq k$). We prove that no online algorithm can have a competitive ratio smaller than $\sqrt{m/k}$. We also study how adding flexibility on task processing, such as task migration or spoliation, or increasing the knowledge of the scheduler by providing it with information on the task graph, influences the lower bound. We provide a $(2\sqrt{m/k}+1)$-competitive algorithm as well as a tunable combination of a system-oriented heuristic and a competitive algorithm; this combination performs well in practice and has a competitive ratio in $\Theta(\sqrt{m/k})$. We also adapt all our results to the case of multiple types of processors. Finally, simulations on different sets of task graphs illustrate how the instance properties impact the performance of the studied algorithms and show that our proposed tunable algorithm performs the best among the online algorithms in almost all cases and has even performance close to an offline algorithm.

The findings were published in a journal [9].

7.4. A generic approach to scheduling and checkpointing workflows

This work deals with scheduling and checkpointing strategies to execute scientific workflows on failure-prone large-scale platforms. To the best of our knowledge, this work is the first to target fail-stop errors for arbitrary workflows. Most previous work addresses soft errors, which corrupt the task being executed by a processor but do not cause the entire memory of that processor to be lost, contrarily to fail-stop errors. We revisit classical mapping heuristics such as HEFT and MINMIN and complement them with several checkpointing strategies. The objective is to derive an efficient trade-off between checkpointing every task (CKPTALL), which is an overkill when failures are rare events, and checkpointing no task (CKPTNONE), which induces dramatic re-execution overhead even when only a few failures strike during execution. Contrarily to previous work, our approach applies to arbitrary workflows, not just special classes of dependence graphs such as M-SPGS (Minimal Series-Parallel Graphs). Extensive experiments report significant gain over both CKPTALL and CKPTNONE, for a wide variety of workflows.

The findings were published in a journal [10].

7.5. Limiting the memory footprint when dynamically scheduling DAGs on shared-memory platforms

Scientific workflows are frequently modeled as Directed Acyclic Graphs (DAGs) of tasks, which represent computational modules and their dependences in the form of data produced by a task and used by another one. This formulation allows the use of runtime systems which dynamically allocate tasks onto the resources of increasingly complex computing platforms. However, for some workflows, such a dynamic schedule may run out of memory by processing too many tasks simultaneously. This paper focuses on the problem of transforming such a DAG to prevent memory shortage, and concentrates on shared memory platforms. We first propose a simple model of DAGs which is expressive enough to emulate complex memory behaviors. We then exhibit a polynomial-time algorithm that computes the maximum peak memory of a DAG, that is, the maximum memory needed by any parallel schedule. We consider the problem of reducing this maximum peak memory to make it smaller than a given bound. Our solution consists in adding new fictitious edges, while trying to minimize the critical path of the graph. After proving that this problem is NP-complete, we provide an ILP solution as well as several heuristic strategies that are thoroughly compared by simulation on synthetic DAGs modeling actual computational workflows. We show that on most instances we are able to decrease the maximum peak memory at the cost of a small increase in the critical path, thus with little impact on the quality of the final parallel schedule.

The findings were published in a journal [12].

7.6. Scheduling independent stochastic tasks on heterogeneous cloud platforms

This work introduces scheduling strategies to maximize the expected number of independent tasks that can be executed on a cloud platform within a given budget and under a deadline constraint. The cloud platform is composed of several types of virtual machines (VMs), where each type has a unit execution cost that depends upon its characteristics. The amount of budget spent during the execution of a task on a given VM is the product of its execution length by the unit execution cost of that VM. The execution lengths of tasks follow a variety of standard probability distributions (exponential, uniform, half-normal, etc.), which is known beforehand and whose mean and standard deviation both depend upon the VM type. Finally, there is a global available budget and a deadline constraint, and the goal is to successfully execute as many tasks as possible before the deadline is reached or the budget is exhausted (whichever comes first). On each VM, the scheduler can decide at any instant to interrupt the execution of a (long) running task and to launch a new one, but the budget already spent for the interrupted task is lost. The main questions are which VMs to enroll, and whether and when to interrupt tasks that have been executing for some time. We assess the complexity of the problem by showing its NP-completeness and providing a 2-approximation for the asymptotic case where budget and deadline both tend to infinity. Then we introduce several heuristics and compare their performance by running an extensive set of simulations.

This work has been presented at the Cluster 2019 conference [17].

7.7. Improved energy-aware strategies for periodic real-time tasks under reliability constraints

This work revisited the real-time scheduling problem recently introduced by Haque, Aydin and Zhu [62]. In this challenging problem, task redundancy ensures a given level of reliability while incurring a significant energy cost. By carefully setting processing frequencies, allocating tasks to processors and ordering task executions, we improve on the previous state-of-the-art approach with an average gain in energy of 20%. Furthermore, we establish the first complexity results for specific instances of the problem.

This work has been accepted at the RTSS 2019 conference [18].

7.8. Multilevel algorithms for acyclic partitioning of directed acyclic graphs

We investigate the problem of partitioning the vertices of a directed acyclic graph into a given number of parts. The objective function is to minimize the number or the total weight of the edges having end points in different parts, which is also known as the edge cut. The standard load balancing constraint of having an equitable partition of the vertices among the parts should be met. Furthermore, the partition is required to be acyclic; i.e., the interpart edges between the vertices from different parts should preserve an acyclic dependency structure among the parts. In this work, we adopt the multilevel approach with coarsening, initial partitioning, and refinement phases for acyclic partitioning of directed acyclic graphs. We focus on two-way partitioning (sometimes called bisection), as this scheme can be used in a recursive way for multiway partitioning. To ensure the acyclicity of the partition at all times, we propose novel and efficient coarsening and refinement heuristics. The quality of the computed acyclic partitions is assessed by computing the edge cut. We also propose effective ways to use the standard undirected graph partitioning methods in our multilevel scheme. We perform a large set of experiments on a dataset consisting of (i) graphs coming from an application and (ii) some others corresponding to matrices from a public collection. We report significant improvements compared to the current state of the art.

This work is published in a journal [11].

7.9. A multi-dimensional Morton-ordered block storage for mode-oblivious tensor computations

Computation on tensors, treated as multidimensional arrays, revolve around generalized basic linear algebra subroutines (BLAS). We propose a novel data structure in which tensors are blocked and blocks are stored in an order determined by Morton order. This is not only proposed for efficiency reasons, but also to induce efficient performance regardless of which mode a generalized BLAS call is invoked for; we coin the term mode-oblivious to describe data structures and algorithms that induce such behavior. Experiments on one of the most bandwidth-bound generalized BLAS kernel, the tensor–vector multiplication, not only demonstrate superior performance over two state-of-the-art variants by up to 18%, but additionally show that the proposed data structure induces a 71% less sample standard deviation for tensor–vector multiplication across d modes, where d varies from 2 to 10. Finally, we show our data structure naturally expands to other tensor kernels and demonstrate up to 38% higher performance for the higher-order power method.

This work is published in a journal [13]

7.10. Effective heuristics for matchings in hypergraphs

The problem of finding a maximum cardinality matching in a d-partite, d-uniform hypergraph is an important problem in combinatorial optimization and has been theoretically analyzed. We first generalize some graph matching heuristics for this problem. We then propose a novel heuristic based on tensor scaling to extend the matching via judicious hyperedge selections. Experiments on random, synthetic and real-life hypergraphs show that this new heuristic is highly practical and superior to the others on finding a matching with large cardinality.

This work is published in the proceedings of SEA^2, where it has received the best paper award [16].

7.11. Karp-Sipser based kernels for bipartite graph matching

We consider Karp–Sipser, a well known matching heuristic in the context of data reduction for the maximum cardinality matching problem. We describe an efficient implementation as well as modifications to reduce its time complexity in worst case instances, both in theory and in practical cases. We compare experimentally against its widely used simpler variant and show cases for which the full algorithm yields better performance .

This work appears in the proceedings of ALENEX2020 [20]

7.12. Efficient and effective sparse tensor reordering

This paper formalizes the problem of reordering a sparse tensor to improve the spatial and temporal locality of operations with it, and proposes two reordering algorithms for this problem, which we call BFS-MCS and Lexi-Order. The BFS-MCS method is a Breadth First Search (BFS)-like heuristic approach based on the maximum cardinality search family; Lexi-Order is an extension of doubly lexical ordering of matrices to tensors. We show the effects of these schemes within the context of a widely used tensor computation, the Candecomp/Parafac decomposition (CPD), when storing the tensor in three previously proposed sparse tensor formats: coordinate (COO), compressed sparse fiber (CSF), and hierarchical coordinate (HiCOO). A new partition-based superblock scheduling is also proposed for HiCOO format to improve load balance. On modern multicore CPUs, we show Lexi-Order obtains up to 4.14$\times$ speedup on sequential HiCOO-Mttkrp and 11.88$\times$ speedup on its parallel counterpart. The performance of COO-and CSF-based Mttkrps also improves. Our two reordering methods are more effective than state-of-the-art approaches.

This work appears in the proceedings of ICS2019 [21].

7.13. High performance tensor–vector multiplication on shared-memory systems

Tensor–vector multiplication is one of the core components in tensor computations. We have recently investigated high performance, single core implementation of this bandwidth-bound operation. Here, we investigate its efficient, shared-memory implementations. Upon carefully analyzing the design space, we implement a number of alternatives using OpenMP and compare them experimentally. Experimental results on up to 8 socket systems show near peak performance for the proposed algorithms.

This work appears in the proceedings of PPAM2019 and is supported with a technical report [22], [36].

7.14. Matrix symmetrization and sparse direct solvers

We investigate algorithms for finding column permutations of sparse matrices in order to have large diagonal entries and to have many entries symmetrically positioned around the diagonal. The aim is to improve the memory and running time requirements of a certain class of sparse direct solvers. We propose efficient algorithms for this purpose by combining two existing approaches and demonstrate the effect of our findings in practice using a direct solver. We show improvements in a number of components of the running time of a sparse direct solver with respect to the state of the art on a diverse set of matrices.

This work will appear in the proceedings of CSC2020 [23].

7.15. A scalable clustering-based task scheduler for homogeneous processors using DAG partitioning

When scheduling a directed acyclic graph (DAG) of tasks on computational platforms, a good trade-off between load balance and data locality is necessary. List-based scheduling techniques are commonly used greedy approaches for this problem. The downside of list-scheduling heuristics is that they are incapable of making short-term sacrifices for the global efficiency of the schedule. In this work, we describe new list-based scheduling heuristics based on clustering for homogeneous platforms, under the realistic duplex single-port communication model. Our approach uses an acyclic partitioner for DAGs for clustering. The clustering enhances the data locality of the scheduler with a global view of the graph. Furthermore, since the partition is acyclic, we can schedule each part completely once its input tasks are ready to be executed. We present an extensive experimental evaluation showing the trade-offs between the granularity of clustering and the parallelism, and how this affects the scheduling. Furthermore, we compare our heuristics to the best state-of-the-art list-scheduling and clustering heuristics, and obtain more than three times better makespan in cases with many communications.

This work appears in the proceedings of IPDPS 2019 [25].

7.16. Improving Locality-Aware Scheduling with Acyclic Directed Graph Partitioning

We investigate efficient execution of computations, modeled as Directed Acyclic Graphs (DAGs), on a single processor with a two-level memory hierarchy, where there is a limited fast memory and a larger slower memory. Our goal is to minimize execution time by minimizing redundant data movement between fast and slow memory. We utilize a DAG partitioner that finds localized, acyclic parts of the whole computation that can fit into fast memory, and minimizes the edge cut among the parts. We propose a new scheduler that executes each part one-by-one, obeying the dependency among parts, aiming at reducing redundant data movement needed by cut-edges. Extensive experimental evaluation shows that the proposed DAG-based scheduler significantly reduces redundant data movement.

This work will appear in the proceedings of PPAM 2019 [24].

7.17. Replication Is More Efficient Than You Think

We revisit replication coupled with checkpointing for fail-stop errors. Replication enables the application to survive many fail-stop errors, thereby allowing for longer checkpointing periods. Previously published works use replication with the no-restart strategy, which never restart failed processors until the application crashes. We introduce the restart strategy where failed processors are restarted after each checkpoint. which may introduce additional overhead during checkpoints but prevents the application configuration from degrading throughout successive checkpointing periods. We show how to compute the optimal checkpointing period for this strategy, which is much larger than the one with no-restart, thereby decreasing I/O pressure. We show through simulations that using the restart strategy significantly decreases the overhead induced by replication, in terms of both total execution time and energy consumption.

This work appears in the proceedings of SC 2019 [15], [28].

7.18. Generic matrix multiplication for multi-GPU accelerated distributed-memory platforms over PaRSEC

We introduce a generic and flexible matrix-matrix multiplication algorithm $C = A \times B$ for state-of-the-art computing platforms. Typically, these platforms are distributed-memory machines whose nodes are equipped with several accelerators. To the best of our knowledge, SLATE is the only library that provides a publicly available implementation on such platforms, and it is currently limited to problem instances where the C

matrix can entirely fit in the memory of the GPU accelerators. Our algorithm relies on the classical tile-based outer-product algorithm, but enhances it with several control dependencies to increase data re-use and to optimize communication flow from/to the accelerators within each node. The algorithm is written with the PARSEC runtime system, which allows for a fast and generic implementation, while achieving close-to-peak performance.

This work appears in the proceedings of Scala 2019 [19].

7.19. Reservation strategies for stochastic jobs

We are interested in scheduling stochastic jobs on a reservation-based platform. Specifically, we consider jobs whose execution time follows a known probability distribution. The platform is reservation-based, meaning that the user has to request fixed-length time slots. The cost then depends on both (i) the request duration (pay for what you ask); and (ii) the actual execution time of the job (pay for what you use).

A reservation strategy determines a sequence of increasing-length reservations, which are paid for until one of them allows the job to successfully complete. The goal is to minimize the total expected cost of the strategy. We provide some properties of the optimal solution, which we characterize up to the length of the first reservation. We then design several heuristics based on various approaches, including a brute-force search of the first reservation length while relying on the characterization of the optimal strategy, as well as the discretization of the target continuous probability distribution together with an optimal dynamic programming algorithm for the discrete distribution.

We evaluate these heuristics using two different platform models and cost functions: The first one targets a cloud-oriented platform (e.g., Amazon AWS) using jobs that follow a large number of usual probability distributions (e.g., Uniform, Exponential, LogNormal, Weibull, Beta), and the second one is based on interpolating traces from a real neuroscience application executed on an HPC platform. An extensive set of simulation results show the effectiveness of the proposed reservation-based approaches for scheduling stochastic jobs.

This work appears in the proceedings of IPDPS 2019 [14].

SOCRATE Project-Team

5. New Results

5.1. Flexible Radio Front-End

Activities in this axis could globally be divided in three main topics: wake-up radio and wireless power transfer, RFID systems and combination of spatial modulation and full-duplex.

5.1.1. Wake-Up radio and wireless power transfer

The ubiquity of wireless sensor networks (WSN), as well as the rapid development of the Internet of Things (IoT), impel new approaches to reduce the energy consumption of the connected devices. The wake-up radio receivers (WuRx) were born in this context to reduce as much as possible the energy consumption of the radio communication part. We aim at proposing a low-cost, high-efficiency rectifier to improve a quasi-passive WuRx performance in terms of communication range. By optimizing the wideband matching circuit and the proposed rectifier's load impedance, the sensitivity was increased by 5 dB, corresponding to an increase of the communication range (13 meters in free space) [10].

We also studied an original solution to maximize the DC power collected in the case of a wireless power transfer (WPT) scenario. Using state-space model representation, the WPT System is considered as a feedback approach in order to maximize the amount of harvested energy. To do this, a global simulation is performed to show the importance of taking into account the propagation channel and the rectifier circuit aspects in the case of optimizing the waveform to increase the harvested energy. By using an optimized multi-sine signal with zero phase as the excitation, taking into account the characteristics of the channel and the physical contributions of the rectifier, we managed to obtain better output DC values compared to a single tone source or a multi-sine signal without optimization, with the same average power input [14].

We plan now to apply this optimized WPT technique to feed Wireless sensors in the particular case of ventilation ducts (HVAC) [24].

5.1.2. RFID

The ARA (Auvergne Rhone Alpes) RAFTING project mainly deals with the design and analysis of wire antennas for RFID tags in the context of wearable electronics. More specifically, an helical dipole antenna dedicated to the smart textile yarn applications has been designed. Moreover, the performance was analysized with respect to mechanical constraints, together with the extraction of accurate electrical models. This work was done in collaboration with Primo 1D company. In perspective, the integration of the NFC protocol together with RFID UHF and the integration of sensing capabilities is envisaged [6], [19], [7], [12], [21].

The Spie ICS- INSA Lyon chair on IoT has granted us for a PhD thesis on Scatter Radio and RFID tag-to-tag communications. Some seminal results have shown that it is actually possible to create a communication between two RFID tags, just using ambient radiowaves or a dedicated distant radio source, without the need of generating a signal from the tag itself. Theoretical and simulated performance have been studied.

5.1.3. Combination of spatial modulation and full-duplex

Spatial modulation (SM) as a new MIMO technique is based on transmitting part of the information by activating different emitting antennas. SM increases spectral efficiency and uses only one radio frequency chain. Moreover, for full-duplex (FD) communication systems, self-interference (SI) is always a central problem. Therefore, combining FD and SM can dramatically reduce the difficulty of SIC (Self-interference Cancellation) because of the single SI chain. A Full Duplex Spatial Modulation (FDSM) system is proposed and an active analog SIC is designed in this work. Moreover, the impact of SIC accuracy on the system performance is studied. The results demonstrate that the accuracy requirement will increase as the INR (Self-interference-to-noise Ratio) increases. The FDSM system is less sensitive than the FD system, which can get a better BER (Bit Error Rate) performance as errors increase. Furthermore, an SI detector is proposed to resolve the influence of the number of detected symbols.

5.2. Software Radio Programming Model

5.2.1. Transiently powered systems and Non-Volatile Memory

Socrate is studying the new NVRAM (Non-Volatile Radom Access Memory) technology and its use in ultra-low power context. Non-Volatile memory has been extisting for a while (Nand Flash for instance) but was not sufficiently fast to be used as main memory. Many emerging technologies are forseen for Non-Volatile RAM to replace current RAM [32].

Socrate has started a work on the applicability of NVRAM for *transiently powered systems*, i.e. systems which may undergo power outage at any time. This study resulted in the Sytare software published in IEEE Transaction on Computer [3] and is also studied in an Inria Project Lab ZEP (https://project.inria.fr/iplzep/teams/).

The Sytare software introduces a checkpointing system that takes into account peripherals (ADC, leds, timer, radio communication, etc.) present on all embedded systems. Checkpointing is the natural solution to power outage: regularly save the state of the system in NVRAM so as to restore it when power is on again. However, no work on checkpointing took into account the restoration of the states of peripherals, Sytare provides this possibility.

Another acheivement in this domain is the PhD of Tristan Delizy that concerns memory heterogeneity that results from new NVM technologies. While emerging memory technologies may offer power reduction and high integration density, they come with major drawbacks such as high latency or limited endurance. As a result, system designers tend to juxtapose several memory technologies on the same chip. We aim to provide the embedded application programmer with a transparent software mechanism to leverage this memory heterogeneity. The work of Tristan Delizy studies the interaction between dynamic memory allocation and memory heterogeneity. He provides cycle accurate simulation of embedded platforms with various memory technologies and shows that different dynamic allocation strategies have a major impact on performance. He demonstrates that interesting performance gains can be achieved even for a low fraction of memory using low latency technology, but only with a clever placement strategy between memory banks. This work will soon be proposed to publication.

5.2.2. Sytare integration in Riot

The ADT SytaRiot has been granted to provide transient power management in the Riot operating system [27]. This integration was realized by Gero muller, here is a summary of the technical tasks and correponding pull request on Riot GitHub:

5.2.2.1. Port RIOT to MSP430+FRAM micro-controllers

- Bring-up the chip against the newer msp430-elf compiler and integrate the toolchain into the RIOT CI infrastructure, cf https://github.com/RIOT-OS/riotdocker/pull/67 , https://github.com/RIOT-OS/riotdocker/pull/82 , https://github.com/RIOT-OS/riotdocker/pull/91
- Implement initial support for the MSP430FR59xx in RIOT, including device drivers for key on-chip peripherals (UART, Timers, GPIO, etc). cf https://github.com/RIOT-OS/RIOT/pull/11012
- Implement a board support package for the MSP-EXP430FR5969 Launchpad Development Kit and the Boost-IR daughter-board (Infrared transceiver + keypad), cf https://github.com/geromueller/RIOT/commit/f13d33
- Participate in IETF hackathon 104 (Prague, March 23–29, 2019) to work on SUIT IoT Firmware Update, cf https://trac.ietf.org/trac/ietf/meeting/wiki/104hackathon

5.2.2.2. Explicit checkpointing in RIOT

- Implement the required low-level code (e.g. DMA driver) for saving/restoring the state of the application to FRAM. cf https://github.com/geromueller/RIOT/commits/checkpoint
- Implement save/restore methods in all relevant device drivers (DMA, GPIO, UART, Timers) and design an API to expose checkpointing as a general system service in RIOT. cf https://github.com/geromueller/RIOT/commit/8b301e

- Participate in the RIOT Summit (Helsinki, September 5–6, 2019) to give a talk about checkpointing and power measurement. cf https://summit.riot-os.org/2019/blog/speakers/gero-muller/ Power measurement

5.2.3. A high-performance ammeter for embedded systems

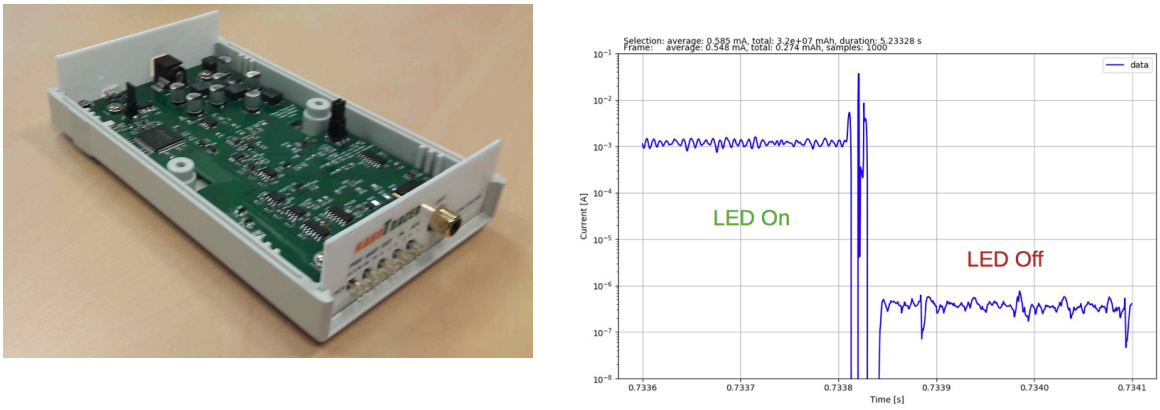


Figure 4. Photo (left) of first packaged nanoTracer prototype and snapshot (right) of a measurement provided by nanoTracer

In embedded low power processing, precise power consumption is a key issue. The Socrate team realized that existing tools could not fullfill the requirements needed for harvesting devices monitoring (measuring from nano-Amperes to milli-Ampere current values at a high sampling rate and continuously).

With the skills of Gero Müller hired on the SytaRiot ADT, the socrate team designed and built a high performance ammeter dedicated to power measurements for small devices. Our prototype measures currents between 100nA and 100mA (gain is audo-adjusted dynamically) with a sampling frequency of 2Msps. Data is streamed to a PC over USB which enables long-running experiments, or just real-time visualization of data (cf screenshot in Fig. 4).

The device, named *nanoTracer*, is referenced in the software section, it is an open project on gitlab (https://gitlab.inria.fr/nanotracer/). A first version is currently tested at Inria (Alexandre Abadie from the IoT SED team) and should soon be available for free for Inria and Academic researcher. We are working on solutions to provide a commercial circuit if requests come from other actors.

5.2.4. Ultra-low latency audio on FPGA

Recently the Socrate team started a collaboration with the researchers of the GRAME group. GRAME is a "Centre National de Création Musicale" (CNCM) organized in three departments: music production, transmission/mediation, and computer music research. Four GRAME researchers have expertise in computer science (compilation), audio DSP, digital lutherie, and human-computer interaction in general. GRAME has been leading the development of the FAUST[0] programming language since its creation in 2004. The GRAME researchers have been associated to CITI as external members in September 2019.

[0] FAUST is a domain specific language for real-time audio signal processing primarily developed at GRAME-CNCM and by a worldwide community. FAUST is based on a compiler "translating" DSP specifications written in FAUST into a wide range of lower-level languages (e.g., C, C++, Rust, Java, WASM, LLVM bitcode, etc.). Thanks to its "architecture" system, generated DSP objects can be embedded into template programs (wrappers) used to turn a FAUST program into a specific ready-to-use object (e.g., standalone, plug-in, smartphone app, webpage, etc.).

Socrate and GRAME have started a collaboration through the Syfala *(synthèse audio faible latence)* project funded by the Fédération Informatique de Lyon. The goal of Syfala is to design an FPGA-based platform for multichannel ultra-low-latency audio Digital Signal Processing (DSP), programmable at high-level with FAUST and using Socrate's software FloPoCo (http://flopoco.gforge.inria.fr). This platform is intended to be usable for various applications ranging from sound synthesis and processing to active sound control and artificial sound field/room acoustics.

Two internships have been working on this project. A first result was a presentation by Florent de Dinechin and Tanguy Risset, introducing the use of HLS and FPGA for audio, at the second *Programmable Audio Workshop* (https://faust.grame.fr/paw/) organized by GRAME.

5.2.5. Evaluation of the posit number system

The posit number system is a very elegant way to represent real numbers in a computers. Its proponents promote it as a better replacement for floating-point arithmetic: posits do indeed improve the application-level accuracy of some applications. However, this also comes with accuracy regressions in other cases. Socrate members, along with members of the AriC project-team, first studied some numerical aspects of posits [18]. Socrate then performed a thorough evaluation of the implementation of the main posit operators, improving the state of the art in hardware posit in the process. Posit operators were then compared to IEEE 754-compliant floating-point operators, and were found to be about twice as slow and twice as expensive [20], [15].

5.2.6. Evaluation of the Unum number system

CEA researcher, in collaboration with Socrate members, designed a complete accelerator for the UNUM number system, including hardware [8] and compiler support [11]. A novelty of this work is the use of a variable-length, self-describing, and memory-oriented floating-point number format [23].

5.2.7. General computer arithmetic

The 10th anniversary of the FloPoCo open-source arithmetic core generator project was the occasion to reflect on the evolutions of the field in a special session about arithmetic generator challenges organized at the ARITH conference [16].

A marked evolution over this period has been the deployment of very good High-Level Synthesis tools, thanks to which hardware is described using a software programming language (usually C++). This comes with many new arithmetic optimization opportunities, some of which have been reviewed in collaboration with Steven Derrien, from Inria Cairn [25]

An issue was the lack in this context of a portable, unified, and hardware-oriented library of arbitrary precision integers. In collaboration with David Thomas from Imperial College, London, we worked on such a library, and demonstrated that it enables a safe description of complex small-grain architectures (such as floating-point or posit operators) with a performance matching traditional hardware description languages [9].

Meanwhile, we keep studying the most basic operators. There has always existed two main methods of implementing mulitplication by a constant in hardware: Table-Based, and Shift-And-Add. This deserved a qualitative and quantitative comparison [17]. This work (with Martin Kumm, from Fulda Technical University, and Silviu Filip, from Inria Cairn) also includes a refined ILP-based algorithm for the problem of multiplying a fixed-point input number by a real constant.

CHROMA Project-Team

7. New Results

7.1. Robust state estimation (Sensor fusion)

This research is the follow up of Agostino Martinelli's investigations carried out during the last five years, which are in the framework of the visual and inertial sensor fusion problem and the unknown input observability problem.

7.1.1. Visual-inertial structure from motion

Participant: Agostino Martinelli.

We have continued our study on the visual inertial sensor fusion problem in the cooperative case, with a special focus on the case of two agents. During this year, we have carried out an exhaustive analysis of all the singularities and minimal cases of this cooperative sensor fusion problem. As in the case of a single agent and in the case of other computer vision problems, the key of the analysis is the establishment of an equivalence between the cooperative visual-inertial sensor fusion problem and a Polynomial Equation System (PES). In the case of a single agent, the PES consists of linear equations and a single polynomial of second degree. In the case of two agents, the number of second degree equations becomes three and, also in this case, a complete analytic solution can be obtained [19], [20]. The power of the analytic solution is twofold. From one side, it allows us to determine the state without the need of an initialization. From another side, it provides fundamental insights into all the structural properties of the problem. The research of this year has focused on this latter issue. Specifically, we have obtained all the minimal cases and singularities depending on the number of camera images and the relative trajectory between the agents. The problem, when non singular, can have up to eight distinct solutions. The usefulness of this analysis has also been illustrated with simulations. In particular, we have quantitatively obtained how the performance of the state estimation worsens near a singularity. The results of this research will be published by the Robotics and Automation Letter (RA-L) journal [18].

7.1.2. Unknown Input Observability

Participant: Agostino Martinelli.

The Unknown Input Observability problem (UIO) in the nonlinear case was an open problem since the sixties years, when it was solved only in the linear case. In the last five years, I have obtained its general analytic solution. So far, I only published the solution for systems characterized by driftless dynamics. In particular, this solution was published as a full paper on the IEEE Transaction on Automatic Control [17]. In December 2018, I was invited by the Society for Industrial and Applied Mathematics (SIAM) to write a book with the general solution. This has been the main work of this year. Since this general solution is based on tensorial calculus (Ricci algebra) and many mathematics procedures and tricks borrowed from theoretical physics, the scope of book has gone much more beyond the presentation of the solution. Basically, by writing this book, I've obtained a new theory of observability.

The current theory of nonlinear observability, does not capture/exploit the key features that are intimately related to the concept of observability. This results in two important limitations:

- The theory, although simple and based on elementary mathematics, can be sometimes burdensome with the risk of easily loosing the meaning of the results and losing the meaning of their assumptions.
- More complex observability problems (e.g., the unknown input observability problem to which this book provides the complete analytic solution) remained unsolved for half a century.

The key to overcome the two above limitations, consists in building a new theory of observability that accounts for the **group of invariance that is inherent to the concept of observability**. This is the typical manner the research in physics has always proceeded. To this regard, I wish to emphasize that the derivation of the basic equations of any physics theory (e.g., the General Relativity, the Yang Mills theory, the Quantum Chromodynamics) starts precisely from the characterization of the group of invariance of the theory.

One of the major novelties introduced by this book is the characterization of the group of invariance of observability and, regarding the case of unknown inputs, the characterization of a subgroup that was called the *Simultaneous Unknown Input Output transformations' group*.

In summary, the book provides several novelties with respect to the existing literature in control theory. Specifically, the reader will learn the following:

- The solution of two open problems in control theory (the book provides separately the solution and the derivation), which are:
  - The extension of the observability rank condition to nonlinear systems driven by also unknown inputs.
  - The extension of the observability rank condition to nonlinear, time-variant systems (both in presence and in absence of unknown inputs)
- A new and more palatable derivation of the existing results in nonlinear observability.
- A new manner of approaching scientific and technological problems, borrowed from theoretical physics (a chapter summarizes in a very intuitive and quick manner the basic mathematics, which includes tensorial calculus).
- A new manner of dealing with the variable *time* in system theory, which is obtained by introducing a new framework, which was called the *chronospace*.

I believe this book could be an opportunity for control and information theory communities to borrow basic mathematics, tricks, types of reasoning from theoretical physics to revisit many aspects of control and information theory.

7.2. Bayesian Perception

Participants: Christian Laugier, Lukas Rummelhard, Jean-Alix David, Jerome Lussereau, Thomas Genevois, Nicolas Turro [SED], Rabbia Asghar, Mario Garzon.

Recognized as one of the core technologies developed within the team over the years (see related sections in previous activity report of Chroma, and previously e-Motion reports), the CMCDOT framework is a generic Bayesian Perception framework, designed to estimate a dense representation of dynamic environments and the associated risks of collision, by fusing and filtering multi-sensor data. This whole perception system has been developed, implemented and tested on embedded devices, incorporating over time new key modules. In 2019, this framework, and the corresponding software, has continued to be the core of many important industrial partnerships and academic contributions, and to be the subject of important developments, both in terms of research and engineering. Some of those recent evolutions are detailed below.

In 2019, the new results have been presented in several invited talks given in some of the major international conferences of the domain [30], [28], [26], [29], [27].

7.2.1. Conditional Monte Carlo Dense Occupancy Tracker (CMCDOT) Framework

Participants: Lukas Rummelhard, Jerome Lussereau, Jean-Alix David, Thomas Genevois, Christian Laugier, Nicolas Turro [SED].

Important developments in the CMCDOT (Fig. 5), in terms of calculation methods and fundamental equations, were introduced and tested. These developments are currently being patented, and will then be used for academic publications. These changes lead to a much higher update frequency, greater flexibility in the management of transitions between states (and therefore a better system reactivity), as well as to the management of a high variability in sensor frequencies (for each sensor over time, and in the set of sensors). The changes include:

- Grid fusion: a new fusion of occupancy grids, enhanced with "unknown" variables, has been developed and implemented. The role of unknown variables has also been enlarged. Currently being patented, it should be the subject of an upcoming paper.
- Ground Estimator: a new method of occupancy grid generation, more accurately taking into account the height of each laser beam, has been developed. Currently being patented, it should be the subject of an upcoming paper.
- Software optimization: the whole CMCDOT framework has been developed on GPUs (implementations in C++/Cuda). An important focus of the engineering has always been, and continued to be in 2019, on the optimization of the software and methods to be embedded on low energy consumption embedded boards (Nvidia Jetson TX1, TX2, AGX Xavier).

Figure 5. CMCDOT results

7.2.2. Multimodal Bayesian perception

Participants: Thomas Genevois, Christian Laugier.

The objective is to extend the concept of Bayesian Perception to the fusion of multiple sensing modalities (including raw data provided by low cost sensors). In 2019, we have developed and implemented a Bayesian model dedicated to ultrasonic range sensors. For any given measurement provided by the sensor, the model computes the occupancy probability in a 2 dimensional grid around the sensor. This computation takes into account the accuracy and the possibility to "miss" an object. Thanks to various parameters, this model has been applied to the sensors of our Renault Zoe demonstrator and to the low cost sensors of our light vehicle demonstrator (flycar).

Fig. 6 .a shows an example, developed and implemented on our light vehicle demonstrator. In this example, the perception is relying on 1 lidar and 5 ultrasonic range sensors. An occupancy grid is generated for each sensor. Then they are fused in a single occupancy grid which is filtered using the CMCDOT approach.

7.2.3. Embedding deep learning for semantics

Participants: Thomas Genevois, Christian Laugier.

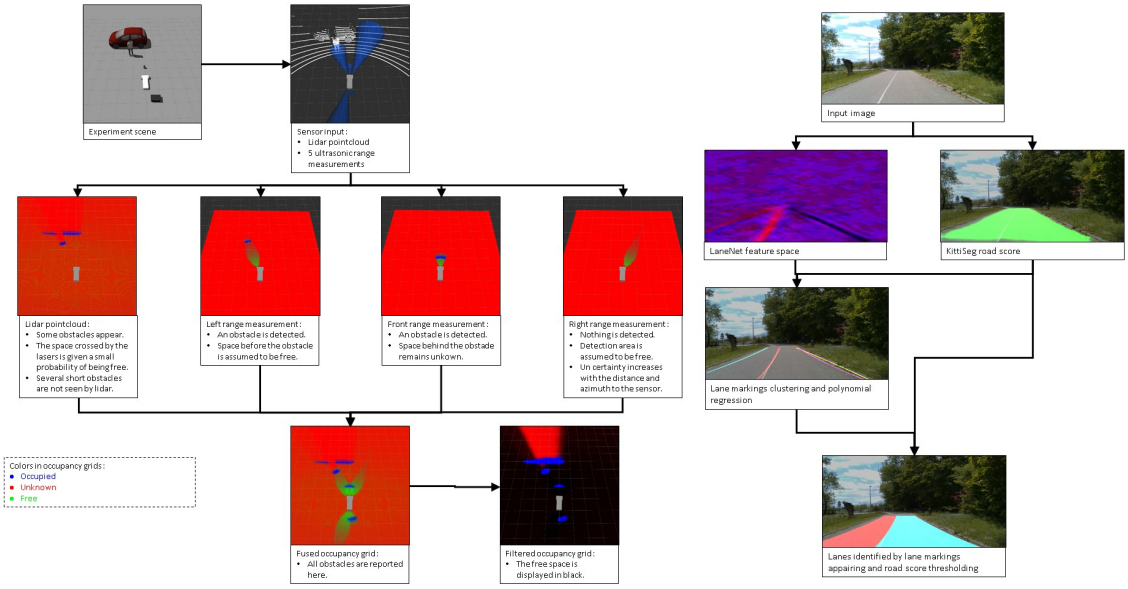


Figure 6. a. Example of multimodal perception, implemented both in simulation and on an actual vehicle demonstrator. b. Combining LaneNet and KittiSeg into a common lane recognition tool.

The objective is to improve embedded Bayesian Perception outputs in our experimental vehicle platforms (Renault Zoe and Flycar), by adding semantics obtained using RGB images and embedded deep learning approaches. In 2019, we have tested several networks for road scene semantic segmentation and implemented two of them in our vehicle platforms:

- LaneNet is a network that provides lane markings detection in road scenarios [83]
- KittiSeg is a network that performs the segmentation of roads [95]

Therefore, Kittiseg is used to identify the shape of the road within an RGB image and LaneNet is used to identify the lane markings that divide the road into lanes. Upon this, we have developed a post-processing technique based on filtering, clustering and regression (Fig. 6 .b). This post-processing technique makes the whole system far more robust and allows to express the lanes in a simple way (polynomial curves in the vehicle's base frame).

Since the objective is to embed semantic segmentation tools on our vehicle platforms, an emphasis has been put on the related embedded constraints (in particular strong real time constraints and appropriate light hardware such as the NVIDIA Jetson TX2). However, the networks LaneNet and KittiSeg have not been optimized neither for real-time inference nor for inference on light hardware. This is why we had to propose an approach for adapting these networks to our strong embedded constraints. This approach relies on the following three main steps: Reducing the resolution of the input image, Removing all computations not needed at inference (some parts of the networks are only needed in the learning phase), Adapting the network's shape to the hardware.

These optimization steps have been followed for KittiSeg and LaneNet networks. The improvement is obvious. Namely, for the network LaneNet the initial inference needed 334 operations while, after optimization, it needs only 10 operations. The inference initially runs at 0.3Hz on our board NVIDIA Jetson TX2 while, after optimization, it runs at 10Hz. Also the memory needed for inference is divided by two due to the optimization.

7.2.4. Online map-relative localization

Participants: Rabbia Asghar, Mario Garzon, Jerome Lussereau, Christian Laugier.

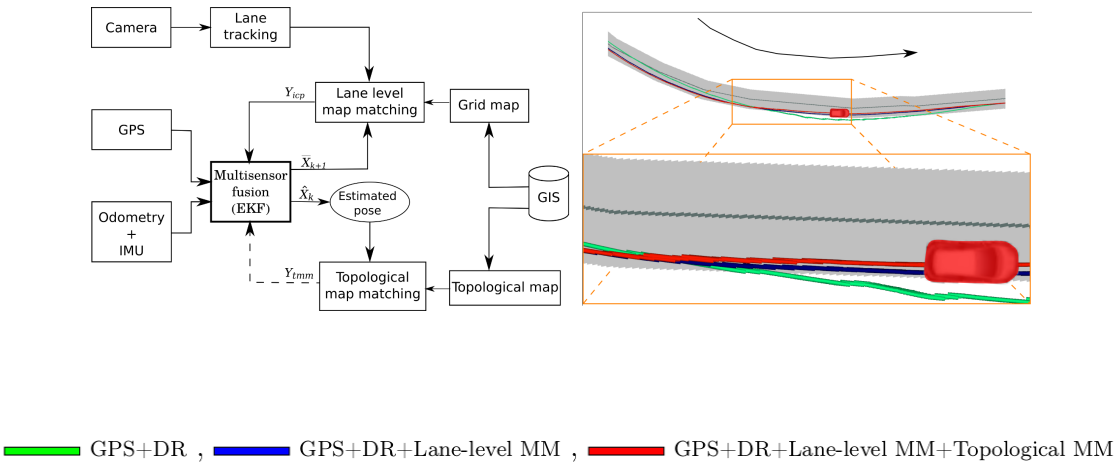


Figure 7. (a) Overview of the map relative localization approach. (b) Estimated pose of the vehicle using three different localization approaches on a curved section of road. The vehicle is provided as a reference where the estimate vehicle pose is just at the curb of the road. Black arrow represents direction of travel.

Localization is one of the key components of the system architecture of autonomous driving and Advanced Driver Assistance Systems (ADAS). Accurate localization is crucial to reliable vehicle navigation and acts as a prerequisite for the planning and control of autonomous vehicles. Offline digital maps are readily available especially in urban scenarios and they play an important role in the field of autonomous vehicles and ADAS. In this framework, we have developed a novel approach for online vehicle localization in a digital map. Two distinct map matching algorithms are proposed:

- Iterative Closest Point (ICP) based lane level map matching (Ll.MM) is performed with visual lane tracker and grid map.
- Decision-rule (DR) based approach is used to perform topological map matching (T. MM).

Results of both map matching algorithms are fused together with GPS and dead reckoning using Extended Kalman Filter to estimate the vehicle's pose relative to the map (see Fig. 7). The approach has been validated on real life conditions on a road-equipped vehicle using a readily available, open source map. Detailed analysis of the experimental results show improved localization using the two aforementioned map matching algorithms (see [50] for more details).
This research work has been carried out in the scope of Project Tornado. A paper on this work was submitted to ICRA2020 and is awaiting review.

7.2.5. System Validation using Simulation and Formal Methods

Participants: Alessandro Renzaglia, Anshul Paigwar, Mathieu Barbier, Philippe Ledent [Chroma/Convecs], Radu Mateescu [Convecs], Christian Laugier, Eduard Baranov [Tamis], Axel Legay [Tamis].

Since 2017, we are working on novel approaches, tools and experimental methodologies with the objective of validating probabilistic perception-based algorithms in the context of autonomous driving. To achieve this goal, a first approach based on Statistical Model Checking (SMC) has been mainly studied in the scope of the European project Enable-S3 and in collaboration with the Inria team Tamis. In this work, we studied the behavior of specifically defined Key Performance Indicators (KPIs), expressed as temporal properties depending on a set of identified metrics, during a large number of simulations via a statistical model checker. As a result, we obtained an evaluation of the probability for the system to meet the KPIs. In particular, we show how this method can be applied to two different subsystems of an autonomous vehicle: a perception system and a decision-making approach for intersection crossing [31]. A more detailed description of the validation scheme for the decision-making approach has been also presented in [49]. This work has been developed in the framework of M. Barbier's PhD thesis, which has been defended in December 2019 [11]. In parallel, in [38], we also proposed a methodology based on a combination of simulation, formal verification, and statistical analysis to validate the collision-risk assessment generated by the Conditional Monte Carlo Dense Occupancy Tracker (CMCDOT), a probabilistic perception system developed in the team. This second work is in collaboration with the Inria team Convecs.

In both cases, the validation methodology relies on the simulation of realistic scenarios generated by using the CARLA simulator [0]. CARLA simulation environment consists of complex urban layouts, buildings and vehicles rendered in high quality, allowing for a realistic representation of real-world scenarios. The ego-vehicle and its sensors, as well as other moving vehicles can be so configured in the simulation to match with the actual system. In order to be able to efficiently generate a large number of execution traces, we have perfected a parameter-based approach which streamlines the process through which the dimensions and initial position and velocity of non-ego vehicles are specified.

We also collected several traces in real experiments by imitating the collision of the ego-vehicle (equipped Renault Zoe) with a pedestrian (by using a mannequin) and with another vehicle (by throwing a big ball). Since it is unfeasible to generate with real experiments a statistically significant number of traces, we focused our analysis on studying how close the simulation traces are to these real experiments by comparing analogous scenarios. These results have been recently submitted to ICRA and are currently under review [0].

7.2.6. Industrial partners and technological transfer

Participants: Christian Laugier, Lukas Rummelhard, Jerome Lussereau, Jean-Alix David, Thomas Genevois.

In 2019, a significant amount of work has been done with the objective to transfer our Bayesian Perception technologies to industrial companies. In a first step, we have developed a new version of CMCDOT based on a clear split of ROS middle-ware code and of GroundEstimator/CMCDOT CUDA code. This allowed us to develop a new version of CMCDOT using the RTMAPS middleware for Toyota Motor Europe. It also allowed us to transfer the CMCDOT technology to some other industrial partners (confidential), in the scope of the project "Security of Autonomous Vehicle" of IRT Nanoelec. Within the IRT Nanoelec framework, we also developed a new "light urban autonomous vehicle" operating using an appropriate version of the CMCDOT and having the capability to navigate with low cost sensors. A first demo of the prototype of this light vehicle has been shown in December 2019, and a start-up project (named Starlink) is currently in incubation.

7.2.7. Autonomous vehicle demonstrations

Participants: Lukas Rummelhard, Jean-Alix David, Thomas Genevois, Jerome Lussereau, Christian Laugier.

In 2019, Chroma has participated to two main public demonstrations:

- **IEEE IV 2019 Conference** (Versailles Satory, June 2019): A one day public demonstration of our Autonomous Vehicle Embedded Perception System has been done using our Renault Zoe platform. Fig. 8 .a and 8 .b show, respectively, the demonstration track (yellow track) and our booth & demonstration vehicle. During the day, we regularly drove people in our Zoe platform for demonstrating how the perception system was working in various situations.

[0] http://carla.org/

[0] A. Paigwar, E. Baranov, A. Renzaglia, C. Laugier and A. Legay, "Probabilistic Collision Risk Estimation for Autonomous Driving: Validation via Statistical Model Checking", *submitted to* IEEE ICRA20.

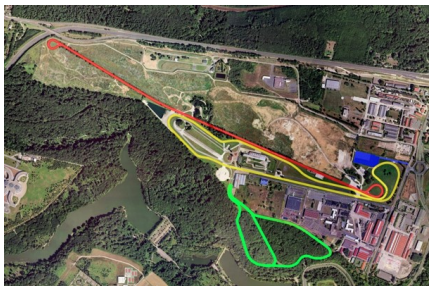


Figure 8. Demonstration at the IV2019 conference : a) track b) demonstration event.

- **FUI Tornado mid-project event** (Rambouillet, September 2019): This one week event included public demonstrations and several open-road tests. During this week, we tested the technologies developed in the scope of the project and we made public and official (for persons from the French Ministries) demonstrations with our Renault Zoe vehicle.

7.3. Situation Awareness & Decision-making for Autonomous Vehicles

Participants: Ozgur Erkent, Christian Wolf, Christian Laugier, Olivier Simonin, Mathieu Barbier, David Sierra-Gonzalez, Jilles Dibangoye, Mario Garzon, Anshul Paigwar, Manuel Alejandro Diaz-Zapata, Victor Romero-Cano [Universidad Autónoma de Occidente, Cali, Colombia], Andrés E. Gómez H., Luiz Serafim-Guardini.

In this section, we include all the novel results in the domains of perception, motion prediction and decision-making for autonomous vehicles. In 2019, these results have also been presented in several invited talks given in some of the major international conferences of the domain [30], [28], [26], [29], [27].

7.3.1. End-to-End Learning of Semantic Grid Estimation Deep Neural Network with Occupancy Grids

Participants: Özgür Erkent, Christian Wolf, Christian Laugier.

Semantic grid is a spatial 2D map of the environment around an autonomous vehicle consisting of cells which represent the semantic information of the corresponding region such as *car, road, vegetation, bikes, etc.*. It consists of an integration of an occupancy grid, which computes the grid states with a Bayesian filter approach, and semantic segmentation information from monocular RGB images, which is obtained with a deep neural network. The network fuses the information and can be trained in an end-to-end manner. The output of the neural network is refined with a conditional random field [15]. The contributions of the study are:

- An end-to-end trainable deep learning method to obtain the semantic grids by integrating the occupancy grids obtained by a Bayesian filter approach and the semantically segmented images by using the monocular RGB images of the environment.
- Grid refinement with conditional random fields (CRFs) on the output of the deep network.
- A comparison of the performances of three different semantic segmentation network architectures in the proposed end-to-end trainable setting.

The proposed method is tested in various datasets (KITTI dataset, Inria-Chroma dataset and SYNTHIA) and different deep neural network architectures are compared (Fig. 9).

7.3.2. Attentional PointNet for 3D object detection in Point Cloud

Participants: Anshul Paigwar, Özgür Erkent, Christian Wolf, Christian Laugier.

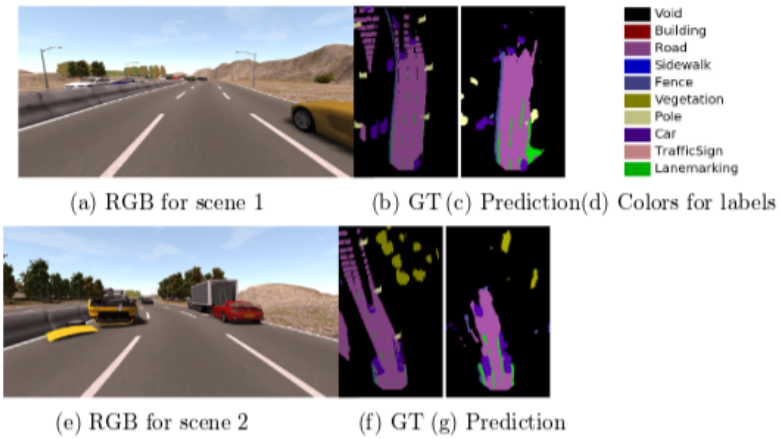


Figure 9. Two scenes with RGB image, ground truth (GT), semantic and segmentation predictions from SYNTHIA dataset.

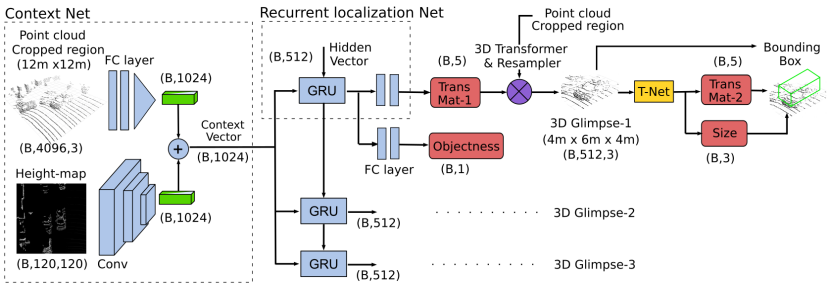


*Figure 10. **Attentional PointNet Architecture**: Given the point cloud and the corresponding height map, network sequentially regress parameters of a 3D Transformation matrix representing pose of a fixed size 3D glimpse. A modified PointNet (T-Net) then estimates another 3D transformation matrix and size representing the 3D bounding box of the object inside the glimpse. Where B is the batch size.*

Accurate detection of objects in 3D point clouds is a central problem for autonomous navigation. Approaches like PointNet [87] that directly operate on sparse point data have shown good accuracy in the classification of single 3D objects. However, LiDAR sensors on Autonomous Vehicles generate a large scale point cloud. Real-time object detection in such a cluttered environment still remains a challenge. In this study, we propose Attentional PointNet, which is a novel end-to-end trainable deep architecture for object detection in point clouds (Fig. 10). We extend the theory of visual attention mechanisms to 3D point clouds and introduce a new recurrent 3D Localization Network module. Rather than processing the whole point cloud, the network learns where to look (finding regions of interest), which significantly reduces the number of points to be processed and inference time. Evaluation on KITTI [72] car detection benchmark shows that our Attentional PointNet achieves comparable results with the *state-of-the-art* LiDAR-based 3D detection methods in detection (Fig. 11) and speed.

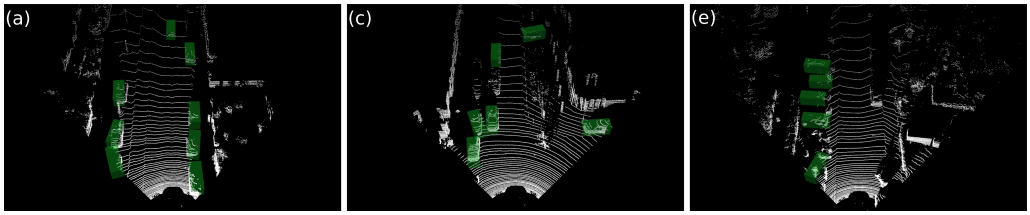


*Figure 11. Visualizations of **Attentional PointNet** results on KITTI dataset for the car category shows model's ability to detect multiple objects in cluttered environments*

This work has been published in CVPR 2019 - Workshop for Autonomous Driving, Long Beach, California, USA [39].

7.3.3. Panoptic Segmentation

Participants: Manuel Alejandro Diaz-Zapata, Victor Romero-Cano [Universidad Autónoma de Occidente, Cali, Colombia], Özgür Erkent, Christian Laugier.

This work has been accomplished during the internship of Manuel Alejandro Diaz Zapata at Inria-Rhone Alpes under supervision of Ozgur Erkent, Victor Romero-Cano and Christian Laugier at Chroma Project Team. Manuel Alejandro Diaz Zapata was a student of Mechatronic Engineering at Universidad Autónoma de Occidente, Colombia during his internship [52].

Semantic segmentation labels an image at the pixel level, where amorphous regions of similar texture or material such as grass, sky or road are given a label depending on the class. Instance segmentation focuses on countable objects such as people, cars or animals by delimiting them in the image using bounding boxes or a segmentation mask. To reduce the gap between the methods used to detect uncountable objects, and things or countable objects, panoptic segmentation has been proposed [75].

We propose a model consisting of three modules: the semantic segmentation module, the instance segmentation module and the panoptic head (Fig.12). Here the semantic segmentation is done by the MobileNetV2 [90] and the instance segmentation is done by Mask R-CNN [73]. The outputs of both networks are joint by the Panoptic Head. The results are provided on two different datasets.

7.3.4. Recognition of dynamic objects for risk assessment

Participants: Andrés E. Gómez H., Özgür Erkent, Christian Laugier.

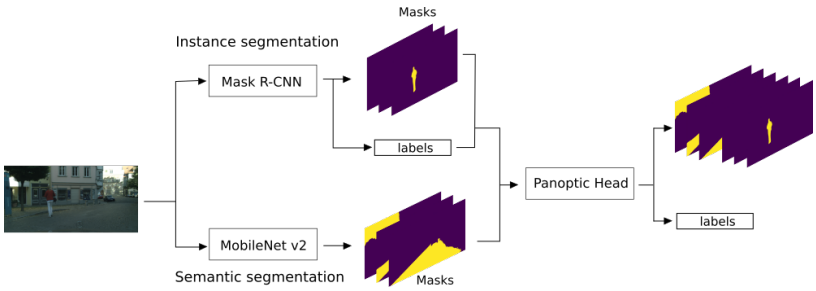


Figure 12. Proposed model for panoptic segmentation.

The Conditional Monte Carlo Dense Occupancy Tracker (*CMCDOT*) framework has proved its accuracy in describing *2D* spatial maps for the Zoe platform. However, this method nowadays cannot recognize the objects in the surrounding. Specifically, the identification of dynamical objects will let us consider different methodologies of risk assessment. This procedure can be possible, through the fusion of RGB and dynamical occupancy grids information.

In the fusion process development, we took into consideration the following steps: *i)* selection of a deep-learning approach, *ii)* development of the projective transformations and *iii)* joining the sub-results. In each step, we used real data from the Zoe platform. In the first step, the *YoloV3* was the deep-learning approach chosen for its accuracy and time performance. In the second step, the projective transformations let us compute the representation of the dynamical points obtained from the occupancy grid plane (i.e., *CMCDOT* framework) in the image plane. Finally, in the third step, we compare the result obtained between the last two-step to identify the dynamic objects around the Zoe platform.

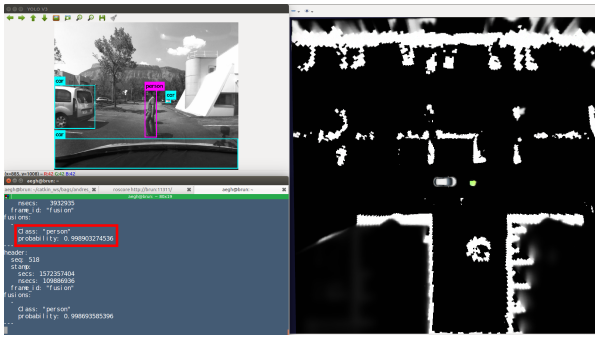


Figure 13. Identification of a pedestrian moving in front of the Zoe platform using the fusion process proposed.

Figure 13 lets us observe the inputs needed for the fusion process and its result.

The work described in this section was done during 2019, inside the activities developed for the Star project. The future work in our project aims to consider the velocity and direction of the dynamic points to define and implement risk behavior functions.

7.3.5. Driving behavior assessment and anomaly detection for intelligent vehicles

Participants: Chule Yang [Nanyang Technological University], Alessandro Renzaglia, Anshul Paigwar, Christian Laugier, Danwei Wang [Nanyang Technological University].

Ensuring safety of both traffic participants and passengers is an important challenge for rapidly growing autonomous vehicle technology. To this purpose, intelligent vehicles not only have to drive safe but must be able to safeguard themselves from other abnormally driving vehicles and avoid potential collisions [56]. Anomaly detection is one of the essential abilities in behavior analysis, which can be used to infer the moving intention of other vehicles and provide evidence for collision risk assessment. In this work, we propose a behavior analysis method based on Hidden Markov Model (HMM) to assess the driving behavior of vehicles on the road and detect anomalous moments. The algorithm uses the real-time velocity and position of the surrounding vehicles provided by the Conditional Monte Carlo Dense Occupancy Tracker (CMCDOT) [89] framework. The movement of each vehicle can be classified into several observation states, namely, Approaching, Braking, Lane Changing, and Lane Keeping. Finally, by chaining these observation states using a Markov model, the abnormality of driving behavior can be inferred into Normal, Attention, and Risk. We perform experiments using CARLA simulator environment to simulate abnormal driving behaviors as shown in Fig. 14 , and we provide results showing the successful detection of abnormal situations.

This work has been published in IEEE CIS-RAM 2019, Bangkok [45].

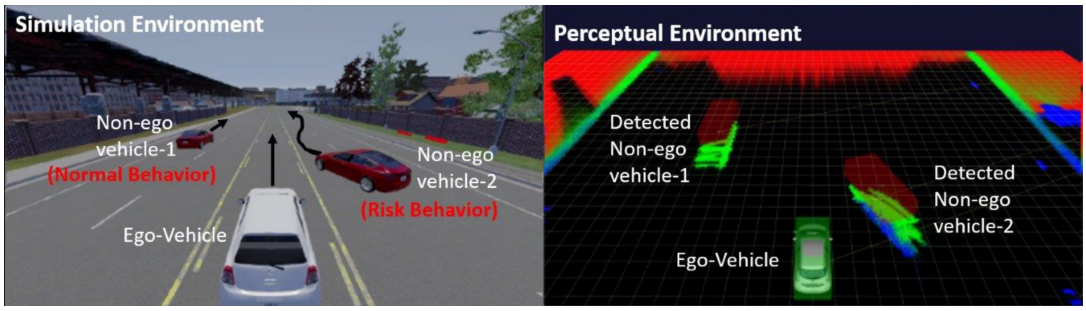


Figure 14. (Left) Simulation environment with CARLA simulator. The white vehicle is the ego-vehicle and two non-ego vehicles are simulated to perform anomaly movements. (Right) Perceptual environment with CMCDOT framework. By analyzing the real-time velocity and position of vehicles, the state and behavior of vehicles can be inferred.

7.3.6. Human-Like Decision-Making for Automated Driving in Highways

Participants: David Sierra-Gonzalez, Mario Garzon, Jilles Dibangoye, Christian Laugier.

Sharing the road with humans constitutes, along with the need for robust perception systems, one of the major challenges holding back the large-scale deployment of automated driving technology. The actions taken by human drivers are determined by a complex set of interdependent factors, which are very hard to model (e.g. intentions, perception, emotions). As a consequence, any prediction of human behavior will always be inherently uncertain, and becomes even more so as the prediction horizon increases. Fully automated vehicles are thus required to make navigation decisions based on the uncertain states and intentions of surrounding vehicles. Building upon previous work, where we showed how to estimate the states and maneuver intentions of surrounding drivers [91], we developed a decision-making system for automated vehicles in highway environments. The task is modeled as a Partially Observable Markov Decision Process and solved in an online fashion using Monte Carlo tree search. At each decision step, a search tree of beliefs is incrementally built and explored in order to find the current best action for the ego-vehicle. The beliefs represent the predicted state of the world as a response to the actions of the ego-vehicle and are updated using an interaction- and

intention-aware probabilistic model. To estimate the long-term consequences of any action, we rely on a lightweight model-based prediction of the scene that assumes risk-averse behavior for all agents. We refer to the proposed decision-making approach as human-like, since it mimics the human abilities of anticipating the intentions of surrounding drivers and of considering the long-term consequences of their actions based on an approximate, common-sense, prediction of the scene. We evaluated the proposed approach in two different simulated navigational tasks: lane change planning and longitudinal control. The results obtained demonstrated the ability of the proposed approach to make foresighted decisions and to leverage the uncertain intention estimations of surrounding drivers.

This work was published in ITSC 2019 [44]. It constitutes the last contribution of the PhD dissertation of David Sierra González, which was defended in April 2019 [12].

7.3.7. *Contextualized Emergency Trajectory Planning using severity curves*

Participants: Luiz Serafim Guardini, Anne Spalanzani, Christian Laugier, Philippe Martinet.

Perception and interpretation of the surroundings is essential for human drivers as well as for (semi-)autonomous vehicles navigation. To improve such interpretation, a lot of effort has been put in place, for example predicting the behavior of pedestrians and other drivers. Nevertheless, to date, cost maps still have considered simple contextualized objects (for instance, binary allowed/forbidden zones or a fixed weight to each type of object). In this work, the risk of injury issued by accidentology is employed to each class of object present in the scene. The scene is analyzed according to dynamic characteristics related to the Ego vehicle and enclosing objects. The aim is to have a better assessment of the surroundings by creating a navigation cost map and to get an improvement on the understanding of the collision severity in the scene. During the first year of his PhD, Luiz Serafim Gaurdini focused on the development of a probabilistic costmap that expresses the Probability of Collision with Injury Risk (PCIR) (see an example on Figure 15). On top of the information gathered by sensors, it includes the severity of injury in the event of a collision between ego and the objects in the scene. This cost map provides enhanced information to perform vehicle motion planning.

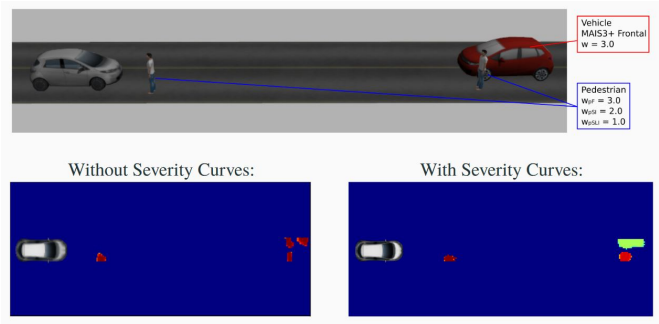


Figure 15. Illustration of the Probabilistic Costmap including the notion of Injury Risk

7.3.8. *Game theoretic decision making for autonomous vehicles' merge manoeuvre in high traffic scenarios*

Participants: Mario Garzon, Anne Spalanzani.

The goal of this work is to provide a solution for a very challenging task: the merge manoeuvre in high traffic scenarios (see Figure 16). Unlike previous approaches, the proposed solution does not rely on vehicle-to-vehicle communication or any specific coordination, moreover, it is capable of anticipating both the actions of other players and their reactions to the autonomous vehicle's movements. The game used is an iterative, multi-player level-k model,which uses cognitive hierarchy reasoning for decision making and has been proved

to correctly model human decisions in uncertain situations. This model uses reinforcement learning to obtain a near-optimal policy, and since it is an iterative model, it is possible to define a goal state so that the policy tries to reach it. To test the decision making process, a kinematic simulation was implemented. The resulting policy was compared with a rule-based approach. The experiments show that the decision making system is capable of correctly performing the merge manoeuvre, by taking actions that require reactions of the other players to be successfully completed. This work was published in [48].

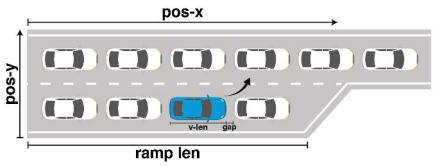


Figure 16. Typical scenario of changing lane in high traffic

7.4. Motion-planning in dense pedestrian environments

We study new motion planning algorithms to allow robots/vehicles to navigate in human populated environment, and to predict human motions. Since 2016, we investigate several directions exploiting vision sensors : prediction of pedestrian behaviors in urban environments (extended GHMM), mapping of human flows (statistical learning), and learning task-based motion planning (RL+Deep-Learning). The works of year 2019 are presented below.

7.4.1. Urban Behavioral Modeling

Participants: Pavan Vasishta, Anne Spalanzani, Dominique Vaufreydaz.

The objective of modeling urban behavior is to predict the trajectories of pedestrians in towns and around cars or platoons (PhD work of P. Vasishta). We first proposed to model pedestrian behaviour in urban scenes by combining the principles of urban planning and the sociological concept of Natural Vision. This model assumes that the environment perceived by pedestrians is composed of multiple potential fields that influence their behaviour. These fields are derived from static scene elements like side-walks, cross-walks, buildings, shops entrances and dynamic obstacles like cars and buses for instance. This work was published in [98]. We then developed an extension to the Growing Hidden Markov Model (GHMM) method that has been proposed to model behavior of pedestrian without observed data or with very few of them. This is achieved by building on existing work using potential cost maps and the principle of Natural Vision. As a consequence, the proposed model is able to predict pedestrian positions more precisely over a longer horizon compared to the state of the art. The method is tested over legal and illegal behavior of pedestrians, having trained the model with sparse observations and partial trajectories. The method, with no training data (see Fig. 17 .a), is compared against a trained state of the art model. It is observed that the proposed method is robust even in new, previously unseen areas. This work was published in [99] and won the **best student paper** of the conference. In 2019, Pavan Vasishta defended his PhD on this topic.

7.4.2. Proactive Navigation for navigating dense human populated environments

Participants: Maria Kabtoul, Anne Spalanzani, Philippe Martinet.

Figure 17. a. Prior Topological Map of the dataset from the Traffic Anomaly Dataset : first figure shows the generated potential cost map and second figure the "Prior Topology" of the image from scene. b. Illustration of the Principle of Proactive Navigation.

Developing autonomous vehicles capable of navigating safely and socially around pedestrians is a major challenge in intelligent transportation. This challenge cannot be met without understanding pedestrians' behavioral response to an autonomous vehicle, and the task of building a clear and quantitative description of the pedestrian to vehicle interaction remains a key milestone in autonomous navigation research. As a step towards safe proactive navigation in a spaceshared with pedestrians, we start to introduce in 2018 a pedestrian-vehicle interaction behavioral model. The model estimates the pedestrian's cooperation with the vehicle in an interaction scenario by a quantitative time-varying function. Using this cooperation estimation the pedestrian's trajectory is predicted by a cooperation-based trajectory planning model (see Figure 17 .b). Both parts of the model are tested and validated using real-life recorded scenarios of pedestrian-vehicle interaction. The model is capable of describing and predicting agents' behaviors when interacting with a vehicle in both lateral and frontal crossing scenarios.

7.4.3. Modelling crowds and autonomous vehicles using Extended Social Force Models

Participants: Manon Predhumeau, Anne Spalanzani, Julie Dugdale.

The focus of this work has been on the realistic simulation of crowds in shared spaces. We have developed a simulator, based on empirical studies and the state of the art, using PED-SIM software. The simulator takes into account the density of crowds, different social group structures in different contexts, inter and intra group forces, and collision avoidance strategies of pedestrians. The Social Force Model (SFM) successfully reproduces many collective phenomena in evacuations or dense crowds. However, pedestrians behaviour is context dependent and the SFM has some limitations when simulating crowds in an open environment under normal conditions. Specifically, in an urban public square pedestrians tend to expand their personal space and try to avoid dense areas to reduce the risk of collision. Based on the SFM, the proposed model splits the perception of pedestrians into a large perception zone and a restricted frontal zone to which they pay more attention. Through their perceptions, the agents estimate the crowd density and dynamically adapt their personal space. Finally, the original social force is tuned to reflect pedestrians preference of avoiding dense areas by turning rather than slowing down as long as there is enough space. Simulation results show that in the considered context the proposed approach produces more realistic behaviours than the original SFM. The simulated crowd is less dense with the same number of pedestrians and less collisions occur, which better fits the observations of sparse crowds in an open place under normal condition [40].

7.4.4. Deep Reinforcement Learning based Vehicle Navigation amongst pedestrians

Participants: Niranjan Deshpande, Anne Spalanzani, Dominique Vaufreydaz.

The objective of this work is to develop a navigation system for an autonomous vehicle in urban environments. The urban environment would consist of other road users as well including other vehicles and pedestrians. Specifically, the focus is on the decision making (behaviour planning) aspect of navigation. In this work, we propose to use Deep Reinforcement Learning as a method to learn decision making. We have developed a Deep Q-Network based agent for decision making amongst pedestrians using the SUMO simulator. This Deep Q-Network based agent is trained for a typical intersection crossing setup amongst pedestrians (see Figure 18). We propose a grid based representation as a state space input to the learning agent. With this grid based representation and our reward function the agent learns a policy capable of driving safely around pedestrians and also follow the traffic rule. This work was published in [35].

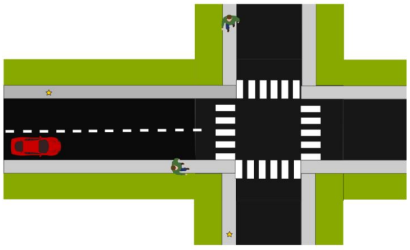


Figure 18. Typical intersection crossing used for training the behavior of the autonomous vehicle

7.5. Learning robot high-level behaviors

7.5.1. Learning task-based motion planning

Participants: Christian Wolf, Jilles Dibangoye, Laetitia Matignon, Olivier Simonin, Edward Beeching.

Our goal is the automatic learning of robot navigation in complex environments based on specific tasks and from visual input. The robot automatically navigates in the environment in order to solve a specific problem, which can be posed explicitly and be encoded in the algorithm (e.g. find all occurences of a given object in the environment, or recognize the current activities of all the actors in this environment) or which can be given in an encoded form as additional input, like text. Addressing these problems requires competences in computer vision, machine learning and AI, and robotics (navigation and paths planning).

A critical part for solving these kind of problems involving autonomous agents is handling memory and planning. An example can be derived from biology, where an animal that is able to store and recall pertinent information about their environment is likely to exceed the performance of an animal whose behavior is purely reactive. Many control problems in partially observed 3D environments involve long term dependencies and planning. Solving these problems requires agents to learn several key capacities: *spatial reasoning* — to explore the environment in an efficient manner and to learn spatio-temporal regularities and affordances. The agent needs to discover relevant objects, store their positions for later use, their possible interactions and the eventual relationships between the objects and the task at hand. Semantic mapping is a key feature in these tasks. A second feature is *discovering semantics from interactions* — while solutions exist for semantic mapping and semantic SLAM [64], [94], a more interesting problem arises when the semantics of objects and their affordances are not supervised, but defined through the task and thus learned from reward.

We started this work in the end of 2017, following the arrival of C. Wolf and his 2 year delegation in the team between Sept 2017. to Sept. 2019, through combinations of reinforcement learning and deep learning. The underlying scientific challenge here is to automatically learn representations which allow the agent to solve multiple sub problems required for the task. In particular, the robot needs to learn a metric representation (a map) of its environment based from a sequence of ego-centric observations. Secondly, to solve the problem,

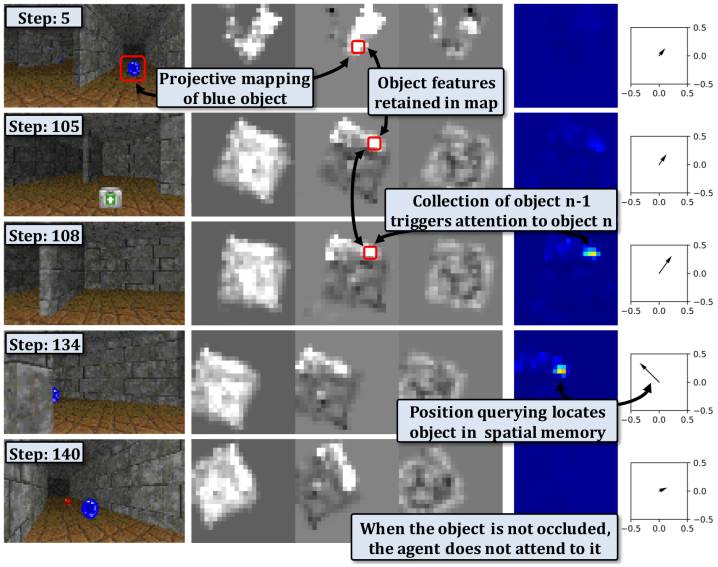


Figure 19. Analysis of the EgoMap for key steps (different rows) during an episode. Left column - RGB observations, central column - the three largest PCA components of features mapped in the spatially structured memory, right - attention heat map (result of the query) and x,y query position vector.

it needs to create a representation which encodes the history of ego-centric observations which are relevant to the recognition problem. Both representations need to be connected, in order for the robot to learn to navigate to solve the problem. Learning these representations from limited information is a challenging goal. This is the subject of the PhD thesis of Edward Beeching, which started on October 2018.

First work proposed a new 3D benchmark for Reinforcement learning, which requires high-level reasoning through the automatic discovery of object affordances [58]. Follow-up work proposed EgoMap, a spatially structured metric neural memory architecture integrating projective geometry in deep reinforcement learning, which we show to outperform classical recurrent baselines. In particular, we show that through visualizations that the agents learns to map relevant objects in its spatial memory without any supervision purely from reward (see Fig. 19). Ongoing work aims to propose a fully differentiable topological memory for Deep-RL.

Creating agents capable of high-level reasoning based on structured memory is main topic of the AI Chair "REMEMBER" obtained by C.Wolf in late 2019 and which involves O. Simonin and J. Dibangoye (Inria Chroma) as well as Laetitia Matignon (LIRIS/Univ Lyon 1). The chair is co-financed by ANR, Naver Labs Europe and INSA-Lyon.

7.5.2. Social robot : NAMO extension and RoboCup@home competition

Participants: Jacques Saraydaryan, Fabrice Jumel, Olivier Simonin, Benoit Renault, Laetitia Matignon, Christian Wolf.

Since 3 years, we investigate robot/humanoid navigation and complex tasks in populated environments such as homes :

- In 2018 we started to study NAMO problems (Navigation Among Movable Obstacles). In his PhD work, Benoit Renault is extending NAMO to Social-NAMO by modeling obstacle hindrance in regards to space access. Defining new spatial cost functions, we extend NAMO algorithms with the ability to maintain area accesses (connectivity) for humans and robots [41]. We also developed a simulator of NAMO problems and algorithms, called S-NAMO-SIM.

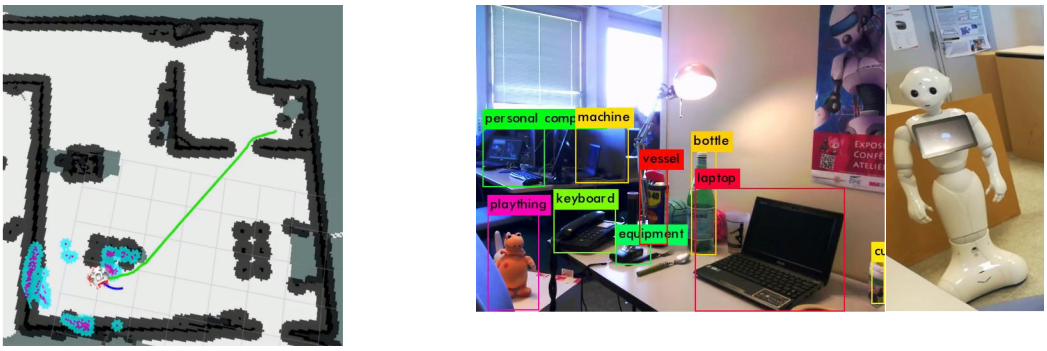


Figure 20. (a) Pepper's navigation and mapping (b) Object detection with Pepper based on vision/deep learning techniques.

- In the context of the **RoboCup** international competition, we created in 2017 the'LyonTech' team, gathering members from Chroma (INSA/CPE/UCBL). We investigated several issues to make humanoid robots able to evolve in a populated indoor environment : decision making and navigation (Fig. 20 .a), human and object recognition based on deep learning techniques (Fig. 20 .b) and human-robot interaction. In July 2018, we participated for the first time to the RoboCup and we reached the 5th place of the SSL league (Robocup@home with Pepper). In July 2019, we participated to the RoboCup organized in Sydney and we obtained the 3rd place of the SSL league. We also awarded the scientific Best Paper of the RoboCup conference [43].

7.6. Sequential decision-making

This research is the follow up of a subgroup led by Jilles S. Dibangoye carried out during the last four years, which include foundations of sequential decision making by a group of cooperative or competitive robots or more generally artificial agents. To this end, we explore combinatorial, convex optimization and reinforcement learning methods.

7.6.1. Optimally solving zero-sum games using centralized planning for decentralized control theory

Participants: Jilles S. Dibangoye, Olivier Buffet [Inria Nancy], Vincent Thomas [Inria Nancy], Abdallah Saffidine [Univ. New South Whales], Christopher Amato [Univ. New Hampshire], François Charpillet [Inria Nancy, Larsen team].

During the last two years, we investigated deep and standard reinforcement learning for solving systems with multiple agents and different information structures. Our preliminary results include:

1. (Theoretical) – As an extension of [68] in the competitive cases, we characterize the optimal solution of two-player fully and partially observable stochastic games.
2. (Theoretical) – We further exhibit new underlying structures of the optimal solution for both non-cooperative two-player settings with information asymmetry, one agent sees what the other does and sees.
3. (Algorithmic) – We extend a non-trivial procedure for computing such optimal solutions.

This work aims at reinforcing a recent theory and algorithms to optimally solving a two-person zero-sum POSGs (zs-POSGs). That is, a general framework for modeling and solving two-person zero-sum games (zs-Games) with imperfect information. Our theory builds upon a proof that the original problem is reducible to a zs-Game—but now with perfect information. In this form, we show that the dynamic programming theory applies. In particular, we extended Bellman equations [59] for zs-POSGs, and coined them maximin (resp. minimax) equations. Even more importantly, we demonstrated Von Neumann & Morgenstern's minimax theorem [102] [103] holds in zs-POSGs. We further proved that value functions—solutions of maximin (resp. minimax) equations—yield special structures. More specifically, the optimal value functions are Lipchitz-continuous. Together these findings allow us to extend planning techniques from simpler settings to zs-POSGs. To cope with high-dimensional settings, we also investigated low-dimensional (possibly non-convex) representations of the approximations of the optimal value function. In that direction, we extended algorithms that apply for convex value functions to lipschitz value functions.

7.6.2. Learning 3D Navigation Protocols on Touch Interfaces with Cooperative Multi-Agent Reinforcement Learning

Participants: Jilles S. Dibangoye, Christian Wolf [INSA Lyon], Quentin Debard [INSA Lyon], Stephane Canu [INSA Rouen].

During the last year, we investigated a number of real-life applications of deep multi-agent reinforcement learning techniques [34]. In particular, we propose to automatically learn a new interaction protocol allowing to map a 2D user input to 3D actions in virtual environments using reinforcement learning (RL). A fundamental problem of RL methods is the vast amount of inter- actions often required, which are difficult to come by when humans are involved. To overcome this limitation, we make use of two collaborative agents. The first agent models the human by learning to perform the 2D finger trajectories. The second agent acts as the interaction protocol, interpreting and translating to 3D operations the 2D finger trajectories from the first agent. We restrict the learned 2D trajectories to be similar to a training set of collected human gestures by first performing state representation learning, prior to reinforcement learning. This state representation learning is addressed by projecting the gestures into a latent space learned by a variational auto encoder (VAE).

7.7. Multi-Robot Routing

7.7.1. Global-local optimization in autonomous multi-vehicle systems

Participants: Guillaume Bono, Jilles Dibangoye, Laetitia Matignon, Olivier Simonin, Florian Peyreron [VOLVO Group, Lyon].

This work is part of the PhD thesis in progress of Guillaume Bono, with the VOLVO Group, in the context of the INSA-VOLVO Chair. The goal of this project is to plan and learn at both global and local levels how to act when facing a vehicle routing problem (VRP). We started with a state-of-the-art paper on vehicle routing problems as it currently stands in the literature [62]. We were surprised to notice that few attention has been devoted to deep reinforcement learning approaches to solving VRP instances. Hence, we investigated our own deep reinforcement learning approach that can help one vehicle to learn how to generalize strategies from solved instances of travelling salesman problems (an instance of VRPs) to unsolved ones.

The difficulty of this problem lies in the fact that its Markov decision process' formulation is intractable, i.e., the number of states grows doubly exponentially with the number of cities to be visited by the salesman. To gain in scalability, we build inspiration on a recent work by DeepMind, which suggests using pointer-net, i.e., a novel deep neural network architecture, to address learning problems in which entries are sequences (here cities to be visited) and output are also sequences (here order in which cities should be visited). Preliminary results are encouraging and we are extending this work to the multi-agent setting.

7.7.2. Towards efficient algorithms for two-echelon vehicle routing problems

Participants: Mohamad Hobballah, Jilles S. Dibangoye, Olivier Simonin, Elie Garcia [VOLVO Group, Lyon], Florian Peyreron [VOLVO Group, Lyon].

During the last year, Mohamad Hobballah (post-doc INSA VOLVO Chair) investigated efficient meta-heuristics for solving two-echelon vehicle routing problems (2E-VRPs) along with realistic logistic constraints. Algorithms for this problem are of interest in many real-world applications. Our short-term application targets goods delivery by a fleet of autonomous vehicles from a depot to the clients through an urban consolidation center using bikers. Preliminary results include:

1. (Methodological) Design of a novel meta-heuristic based on differential evolution algorithm [66] and iterative local search [101]. The former permits us to avoid being attracted by poor local optima whereas the latter performs the local solution improvement.
2. (Empirical) Empirical results on standard benchmarks available at http://www.vrp-rep.org/datasets.html show state-of-the-art performances on most VRP, MDVRP and 2E-VRP instances.

7.7.3. Multi-Robot Routing (MRR) for evolving missions

Participants: Mihai Popescu, Olivier Simonin, Anne Spalanzani, Fabrice Valois [INSA/Inria, Agora team].

After considering Multi-Robot Patrolling of known targets [86], we generalized to MRR (multi-robot routing) and to DMRR (Dynamic MRR) in the work of the PhD of M. Popescu. Target allocation problems have been frequently treated in contexts such as multi-robot rescue operations, exploration, or patrolling, being often formalized as multi-robot routing problems. There are few works addressing dynamic target allocation, such as allocation of previously unknown targets. We recently developed different solutions to variants of this problem :

- MRR-Sat : Multi-robot routing decentralized solutions consist in auction-based methods. Our work addresses the MRR problem and proposes MRR with saturation constraints (MRR-Sat), where the cost of each robot treating its allocated targets cannot exceed a bound (called saturation). We provided a NP-Complete proof for the problem of MRR-Sat. Then, we proposed a new auction-based algorithm for MRR-Sat and MRR, which combines ideas of parallel allocations with target-oriented heuristics. An empirical analysis of the experimental results shows that the proposed algorithm outperforms state-of-the art methods, obtaining not only better team costs, but also a much lower running time. Results are under review.
- DMRR : we defined the Dynamic-MRR problem as the continuous adaptation of the ongoing robot missions to new targets. We proposed a framework for dynamically adapting the existent robot missions to new discovered targets. Dynamic saturation-based auctioning (DSAT) is proposed for adapting the execution of robots to the new targets. Comparison was made with algorithms ranging from greedy to auction-based methods with provable sub-optimality. The results for DSAT shows it outperforms state-of-the-art methods.
- Synchronization : When patrolling targets along bounded cycles, robots have to meet periodically to exchange information, data (e.g. results of their tasks). Data will finally reach a delivery point. Hence, patrolling cycles sometimes have common points (rendezvous points), where the information needs to be exchanged between different cycles (robots). We investigated this problem by defining the following first solutions : random-wait, speed adaptation (first-multiple), primality of periods, greedy interval overlapping. In the context of the PHC 'DRONEM' project [0] we also developed a flow-based approach to the synchronization problem with the team of Prof. Gabriela Czibula from Babes-Bolyai University in Cluj-Napoca, Romania, see [37].

7.8. Multi-UAV exploration and communication

7.8.1. Multi-UAV Exploration and Visual Coverage of 3D Environments

Participants: Alessandro Renzaglia, Olivier Simonin, Jilles Dibangoye, Vincent Le Doze.

[0] Hubert Curien Partnership

Figure 21. (a) UAVs Chroma simulator (b) Intel Aero quadrotors platform (c) Crazyflie micro-UAV platform extended with UWB decawave chip.

Multi-robot teams, especially when involving aerial vehicles (UAVs [0]), are extremely efficient systems to help humans in acquiring information on large and complex environments. In these scenarios, two fundamental tasks are static coverage and exploration. In both cases, the robots' goal is to navigate through the environment and cooperate to maximize the observed area, either by finding the optimal static configuration which provides the best global view in the case of the coverage or by maximizing the new observed areas at every step until the environment becomes completely known in the case of the exploration.

Although these tasks are usually considered separately in the literature, we proposed a common framework where both problems are formulated as the maximization of online acquired information via the definition of single-robot optimization functions, which differs only slightly in the two cases to take into account the static and dynamic nature of coverage and exploration respectively [0]. A common derivative-free approach based on a stochastic approximation of these functions and their successive optimization is proposed, resulting in a fast and decentralized solution. The locality of this methodology limits however this solution to have local optimality guarantees and specific additional layers are proposed for the two problems to improve the final performance.

For the exploration problem, this resulted in a novel decentralized approach which alternates gradient-free stochastic optimization and a frontier-based approach [42] (IROS'19), [47]. Our method allows each robot to generate its own trajectory based on the collected data and the local map built integrating the information shared by its teammates. Whenever a local optimum is reached, which corresponds to a location surrounded by already explored areas, the algorithm identifies the closest frontier to get over it and restarts the local optimization. Its low computational cost, the capability to deal with constraints and the decentralized decision-making make it particularly suitable for multi-robot applications in complex 3D environments.

In the case of visual coverage, we studied how suitable initializations for the UAVs' positions can be computed offline based on a partial knowledge on the environment and how they can affect the final performance of the online measurements-based optimization. The main contribution of this work was thus to add another layer, based on the concept of Centroidal Voronoi Tessellation, to the optimization scheme in order to exploit an a priori sparse information on the environment to cover. The resulting method, taking advantages of the complementary properties of geometric and stochastic optimization, significantly improves the result of the uninitialized solution and notably reduces the probability of a far-to-optimal final configuration. Moreover, the number of iterations necessary for the convergence of the on-line algorithm is also reduced [88].

[0] Unmanned Aerial Vehicles

[0] A. Renzaglia, J. Dibangoye, V. Le Doze and O. Simonin, "A Common Optimization Framework for Multi-Robot Exploration and Coverage in 3D Environments," *sumbitted to* Journal of Intelligent & Robotic Systems, *under review*.

Both previous approaches have been tested in realistic simulations based on our extension of Gazebo, called SimuDronesGR (see Fig. 21 .a). The development of this UAVs simulator, which includes realistic models of both the environment and the aerial vehicle's dynamics and sensors, is an important current activity in Chroma. Such a simulator has the fundamental role of allowing for realistic tests to validate the developed algorithms and to better prepare the implementation of these solutions on the robotic platform of the team (Intel Aero quadrotors, Fig. 21 .b) for real experiments.

7.8.2. Communication-based control of swarm of UAVs

Participants: Remy Grunblatt, Olivier Simonin, Isabelle Guerin-Lassous [Inria/Lyon 1 Dante team], Alexandre Bonnefond.

Intel WiFi controllers are used in many common devices, such as laptops, but also in the Intel Aero Ready-to-Fly UAVs (Unmanned Aerial Vehicle). The mobility capabilities of these devices lead to greater dynamics in radio conditions, and therefore introduce a need for a suitable and efficient rate adaptation algorithm. In the context of the PhD of Remy Grunblatt, we have reverse-engineered the Intel rate adaptation mechanism from the source code of the IwlWifi Linux driver, and we have given, in a comprehensive form, the underlying rate adaptation algorithm named Iwl-Mvm-Rs. We have also implemented the Iwl-Mvm-Rs algorithm in the NS-3 simulator. Thanks to this implementation, we can evaluate the performance of Iwl-Mvm-Rs in different scenarios (static and with mobility, with and without fast fading). We also compared the performances of Iwl-Mvm-Rs with the ones of Minstrel-HT and IdealWifi, also implemented in the NS-3 simulator. This work has been published in ACM MSWiM conference (A) [36].

In the end of 2019, we obtained a DGA/Inria AI project, called "Dynaflock", aiming to extend the flocking approach to control swarm of communicating UAVs. Alexandre Bonnefond started a PhD to elaborate dynamic flocking models based on the link quality, which can be measured online.

7.8.3. Ultra-WideBand based localization & control of micro-UAVs fleets

Participants: Stephane d'Alu, Olivier Simonin, Oana Iova [Inria/INSA Agora team], Hervé Rivano [Inria/INSA Agora team].

The literature on autonomous flight of swarm of UAVs in indoor environments shows it requires the use of an external camera-based localization, i.e. a motion capture system. Indoor flying without such an expensive equipment installed in the infrastructure remains a challenge. To tackle this challenge, we investigate the Ultra-WideBand technology which can be embedded on micro UAVs as a way to estimate inter-drone distances (see Fig. 21 .c Crazyflie micro-UAV). In our approach, the distance information is a fundamental building block to perform a self-maintaining formation flight. We defined and experimented a time-of-flight distance computation, using UWB decawave chips. We showed a Crazyflie flying and computing its position in function of three fixed anchors. We also tested a two-UAV flight where inter-distance is measured to avoid collisions. See first results in [33].

IMAGINE Project-Team

7. New Results

7.1. Star-Shaped Metrics for Mechanical Metamaterial Design

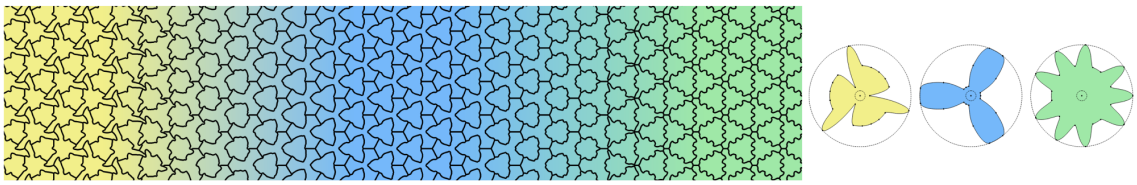


Figure 1. Our method generates a smoothly-graded pattern (left) when interpolating between three star-shaped distance functions (right) on a regular honeycomb lattice. Each distance function is compactly parameterized with polar coordinates, allowing for simple interpolation in metric space as indicated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

We present a method for designing mechanical metamaterials 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, as depicted in Figure 1 . This capability opens up a rich space of structures with interesting aesthetics and a wide range of mechanical properties, including isotropic, tetragonal, orthotropic, as well as smoothly graded materials. We evaluate our method by creating large sets of example structures, provided as accompanying material. We validate the mechanical properties predicted by simulation through tensile tests on a set of physical prototypes.

7.2. Computational Design of Fabric Formwork

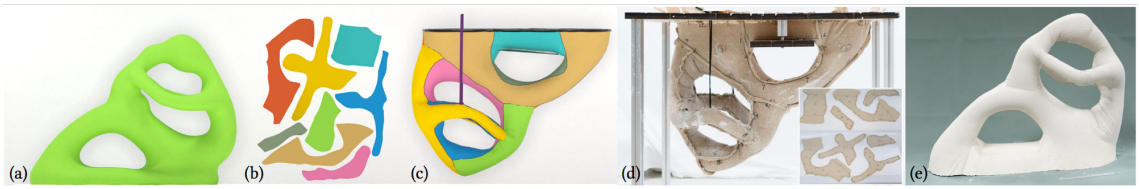


Figure 2. A fertility model designed and fabricated using our computational approach. For a target 3D model (a), our system can automatically compute a set of flat panels (b) that can be sewn together to serve as fabric containers to form a target shape by pressure of liquid plaster poured in – see (c) for the simulation under force equilibrium of membrane tension, liquid pressure and external supports. The generated flat panels are used to conduct the physical fabrication of fabric formwork (d). After drying and unwrapping the fabric container, a sculpture with the designed target shape has been fabricated (e).

This work (illustrated in Figure 2) presents an inverse design tool for fabric formwork - a process where flat panels are sewn together to form a fabric container for casting a plaster sculpture. Compared to 3D printing techniques, the benefit of fabric formwork is its properties of low-cost and easy transport. The process of fabric formwork is akin to molding and casting but having a soft boundary. Deformation of the fabric container is governed by force equilibrium between the pressure forces from liquid fill and tension in the stretched fabric. The final result of fabrication depends on the shapes of the flat panels, the fabrication orientation and the placement of external supports. Our computational framework generates optimized flat panels and fabrication orientation with reference to a target shape, and determines effective locations for external supports. We demonstrate the function of this design tool on a variety of models with different shapes and topology. Physical fabrication is also demonstrated to validate our approach.

7.3. Spatial Motion Doodles

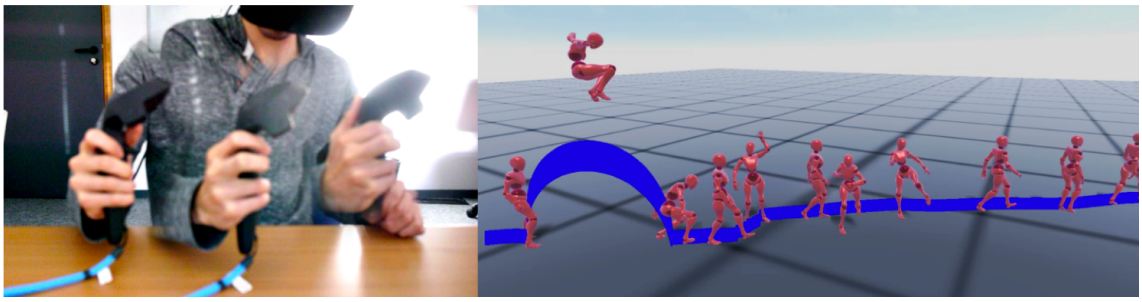


Figure 3. Left: A user drawing a spatial motion doodle (SMD) which is the six-dimensional trajectory of a moving frame (position and orientation), here attached to the HTC Vive controller. Right: The SMD is parsed into a string of motion tokens,allowing to recognize actions and extract the associated motion qualities. This information is transferred to an articulated character to generate an expressive 3D animation sequence.

We present a method for easily drafting expressive character animation by playing with instrumented rigid objects (see Figure 3). We parse the input 6D trajectories (position and orientation over time) – called spatial motion doodles – into sequences of actions and convert them into detailed character animations using a dataset of parameterized motion clips which are automatically fitted to the doodles in terms of global trajectory and timing. Moreover, we capture the expressiveness of user-manipulation by analyzing Laban effort qualities in the input spatial motion doodles and transferring them to the synthetic motions we generate. We validate the ease of use of our system and the expressiveness of the resulting animations through a series of user studies, showing the interest of our approach for interactive digital storytelling applications dedicated to children and non-expert users, as well as for providing fast drafting tools for animators.

7.4. Text-to-Movie Authoring of Anatomy Lessons

With popular use of multimedia and 3D content in anatomy teaching there is a need for a simple yet comprehensive tool to create and edit pedagogical anatomy video lessons. This work introduces an automated video authoring tool (shown in Figure 4) created for teachers. It takes text written in a novel domain specific language (DSL) called the Anatomy Storyboard Language (ASL) as input and translates it to real time 3D animation. Preliminary results demonstrates the ease of use and effectiveness of the tool for quickly drafting video lessons in realistic medical anatomy teaching scenarios.

7.5. Approximate Reconstruction of 3D Scenes From Bas-Reliefs

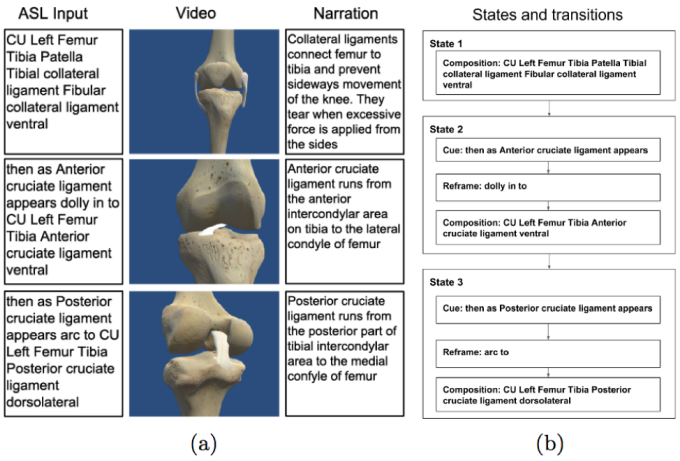


Figure 4. Text-to-movie generation example with hierarchical finite state machines representation.

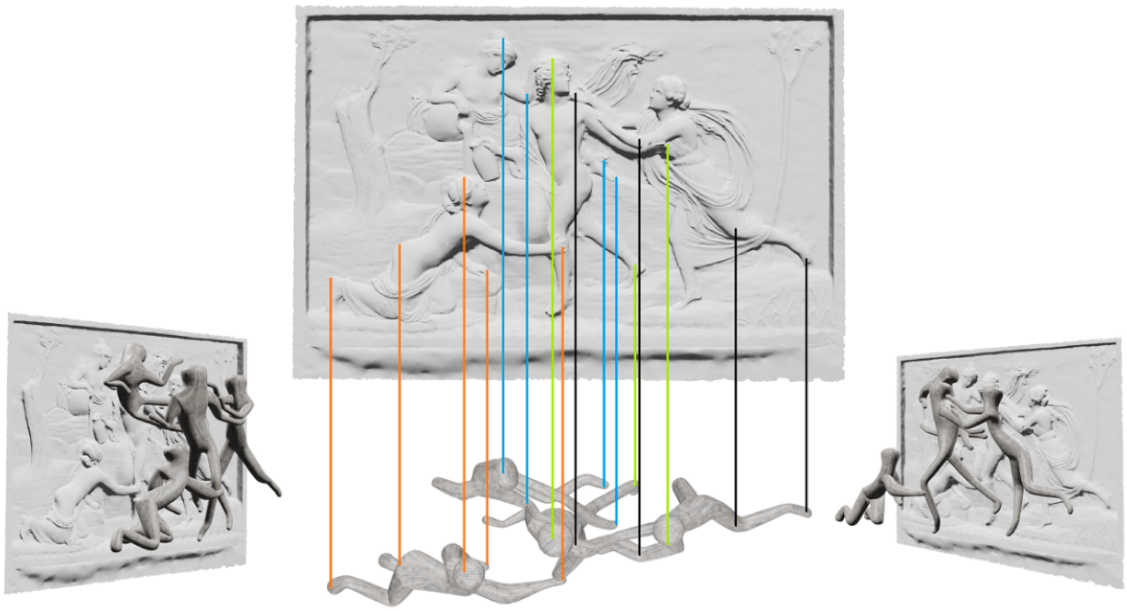


Figure 5. 3D interpretation of the mythological story of Hylas and the Water Nymphs, after a bas-relief marble by Bertel Thorvaldsen (1833). Hylas was sent to fetch water for the camp. Finding a pool in a clearing, he was encircled by water nymphs reaching up to kiss him and there disappeared with them forever. Using hand-drawn silhouette shapes and 2D skeletons of the four characters, we compute a plausible 3D reconstruction of the scene with rigged and skinned models suitable for 3D animation.

For thousands of years, bas-reliefs such as the one depicted in Figure 5 have been used to depict scenes of everyday life, mythology and historic events. Yet, the precise geometry of those scenes remains difficult to interpret and reconstruct. Over the past decade, methods have been developed for generating bas-reliefs from 3D scenes. With this work, we investigate the inverse problem of interpreting and reconstructing 3D scenes from their bas-relief depictions. Even approximate reconstructions can be useful for art historians and museum exhibit designers, as a first entry to the complete interpretation of the narratives told in stone or marble. To create such approximate reconstructions, we present methods for extracting 3D base mesh models of all characters depicted in a bas-relief. We take advantages of the bas-relief geometry and high-level knowledge of human body proportions to recover body parts and their three-dimensional structure, even in severe cases of contact and occlusion. We present experimental results for 6 bas-relief depictions of Greek mythological and historical scenes involving 18 characters and draw conclusions for future work.

MAVERICK Project-Team

6. New Results

6.1. Texture synthesis

6.1.1. 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, ubiquitous in games and special effects. By calling a single procedure in every pixel – or voxel – large quantities of details are generated at low cost, enhancing textures, producing complex structures within and along surfaces. Such procedures are typically implemented as pixel shaders. We propose a novel procedural pattern synthesis technique that exhibits desirable properties for modeling highly contrasted patterns, that are especially well suited to produce surface and microstructure details. In particular, our synthesizer affords for a precise control over the profile, orientation and distribution of the produced stochastic patterns, while allowing to grade all these parameters spatially. Our technique defines a stochastic smooth phase field – a phasor noise – that is then fed into a periodic function (e.g. a sine wave), producing an oscillating field with prescribed main frequencies and preserved contrast oscillations. In addition, the profile of each oscillation is directly controllable as shown Figure 2 . Our technique builds upon a reformulation of Gabor noise in terms of a phasor field that affords for a clear separation between local intensity and phase. Applications range from texturing to modeling surface displacements, as well as multi-material microstructures in the context of additive manufacturing.

This paper was published in ACM TOG [6] and presented at Siggraph 2019.

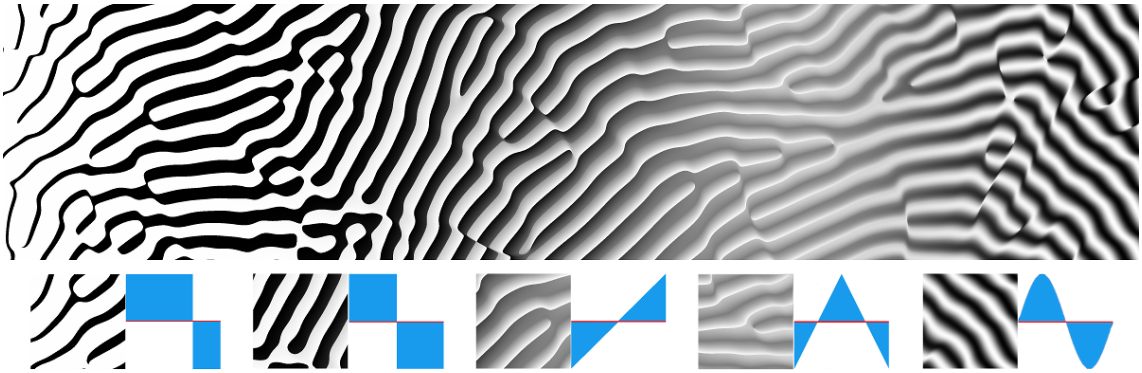


Figure 2. High-contrast patterns produced by our approach. Note how the profile of the oscillations smoothly transition from a rectangular wave (20% black), to a square wave, to a triangular profile and finally a sine wave. At the same time, the orientation of the waves changes from left to right. The field visualized here is purely procedural. It is obtained by feeding our phasor noise into periodic profile functions (shown in blue), that are interpolated from left to right.

6.1.2. Making Gabor Noise Fast and Normalized

Participants: Vincent Tavernier, Fabrice Neyret, Romain Vergne, Joëlle Thollot.

Gabor Noise is a powerful procedural texture synthesis technique, but it has two major drawbacks: It is costly due to the high required splat density and not always predictable because properties of instances can differ from those of the process. We bench performance and quality using alternatives for each Gabor Noise ingredient: point distribution, kernel weighting and kernel shape. For this, we introduce 3 objective criteria to measure process convergence, process stationarity, and instance stationarity. We show that minor implementation changes allow for 17-24× speed-up with same or better quality.

This paper has been presented at Eurographics-short 2019 [11].

6.2. Illumination simulation and materials

6.2.1. Harmonic Analysis of the Light Transport Operator

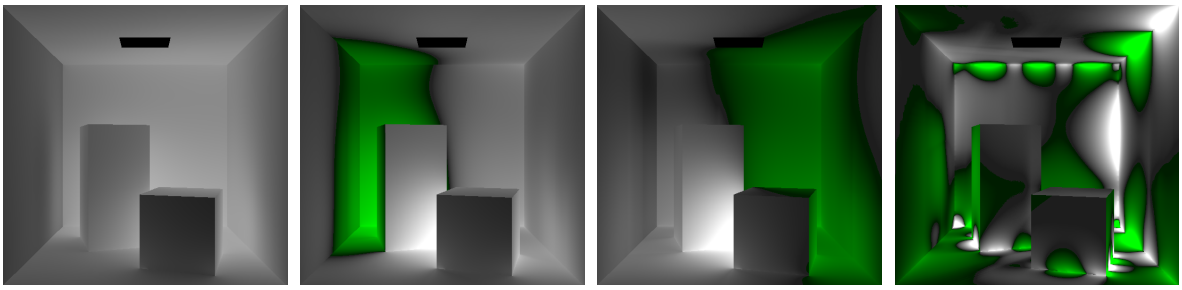
Participants: Ronak Molazem, Cyril Soler.

In this work we study the eigenvalues and eigenfunctions of the light transport operator. While computing the spectrum of the light transport operator is a simple task in Lambertian scenes by applying a traditional eigensolver to the linear system obtained from discretized geometry, it becomes a real challenge in general environments where discretizing the geometry is not possible anymore. “Diagonalizing” light transport however can be a very effective way to perform re-lighting and rapidely compute light transport solutions.

In this work we propose an analysis of the properties of the spectrum of the light transport operator, connecting the calculation of eigenvalues to resolvent theory. We show that the eigenfunctions are generally not orthogonal nor positive, but they can still be used to efficiently represent light distributions.

We analyse the performance of different methods to compute eigenvalues and images of their eigenfunctions using path tracing. We prove in particular that it is possible to compute the eigenfunctions of the light transport operator by integrating "circular" light paths of various lengths accross the scene.

This work is part of the PhD of Ronak Molazem and is funded by the ANR project "CaLiTrOp". At the time of writing this (Dec. 2019), we’re about to submit a paper to ACM Transactions on Graphics.



Λ_0 Λ_1 Λ_2 Λ_{144}

Figure 3. Path-traced images of four eigenfunctions of the light transport operator in the Cornell Box. A green scale is used to represent negative values.

6.2.2. Low Dimension Approximations of Light Transport

Participants: Ronak Molazem, Cyril Soler.

Light transport is known to be a low rank linear operator: the vector space formed by solutions of a light transport problem for different initial conditions is of low dimension. Approximating this space using appropriate bases is therefore of primordial help to efficiently compute solutions to light transport problems.

In this work, we're interested into generating such approximations using *ad-hoc* methods that rely on deep learning. The goal is to be able to efficiently generate a sensible basis for light transport solutions on which we can efficiently project a noisy image. Other applications of this work include relighting pictures, in which an approximate geometry is used to project the illumination in the image, that can further be manipulated while staying in the space of expected light transport solutions.

This work is an ongoing collaboration with Unity Research Grenoble, and part of the PhD of Ronak Molazem, currently in her second year of PhD, and is funded by the ANR project "CaLiTrOp".

6.2.3. Precomputed Multiple Scattering for Rapid Light Simulation in Participating Media

Participants: Nicolas Holzschuch, Liangsheng Ge, Beibei Wang.

Rendering translucent materials is costly: light transport algorithms need to simulate a large number of scattering events inside the material before reaching convergence. The cost is especially high for materials with a large albedo or a small mean-free-path, where higher-order scattering effects dominate. In [7], we present a new method for fast computation of global illumination with participating media. Our method uses precomputed multiple scattering effects, stored in two compact tables. These precomputed multiple scattering tables are easy to integrate with any illumination simulation algorithm. We give examples for virtual ray lights (VRL), photon mapping with beams and paths (UPBP), Metropolis Light Transport with Manifold Exploration (MEMLT). The original algorithms are in charge of low-order scattering, combined with multiple scattering computed using our table. Our results show significant improvements in convergence speed and memory costs, with negligible impact on accuracy.

6.2.4. Fast Computation of Single Scattering in Participating Media with Refractive Boundaries using Frequency Analysis

Participants: Nicolas Holzschuch, Yulin Liang, Lu Wang, Beibei Wang.

Many materials combine a refractive boundary and a participating media on the interior. If the material has a low opacity, single scattering effects dominate in its appearance. Refraction at the boundary concentrates the incoming light, resulting in an important phenomenon called volume caustics. This phenomenon is hard to simulate. Previous methods used point-based light transport, but attributed point samples inefficiently, resulting in long computation time. In [3], we use frequency analysis of light transport to allocate point samples efficiently. Our method works in two steps: in the first step, we compute volume samples along with their covariance matrices, encoding the illumination frequency content in a compact way. In the rendering step, we use the covariance matrices to compute the kernel size for each volume sample: small kernel for high-frequency single scattering, large kernel for lower frequencies. Our algorithm computes volume caustics with fewer volume samples, with no loss of quality. Our method is both faster and uses less memory than the original method. It is roughly twice as fast and uses one fifth of the memory. The extra cost of computing covariance matrices for frequency information is negligible.

6.2.5. Reparameterizing discontinuous integrands for differentiable rendering

Participants: Nicolas Holzschuch, Wenzel Jakob, Guillaume Loubet.

Differentiable rendering has recently opened the door to a number of challenging inverse problems involving photorealistic images, such as computational material design and scattering-aware reconstruction of geometry and materials from photographs. Differentiable rendering algorithms strive to estimate partial derivatives of pixels in a rendered image with respect to scene parameters, which is difficult because visibility changes are inherently non-differentiable.

We propose [5] a new technique for differentiating path-traced images with respect to scene parameters that affect visibility, including the position of cameras, light sources, and vertices in triangle meshes. Our algorithm computes the gradients of illumination integrals by applying changes of variables that remove or strongly reduce the dependence of the position of discontinuities on differentiable scene parameters. The underlying parameterization is created on the fly for each integral and enables accurate gradient estimates using standard Monte Carlo sampling in conjunction with automatic differentiation. Importantly, our approach does not rely

on sampling silhouette edges, which has been a bottleneck in previous work and tends to produce high-variance gradients when important edges are found with insufficient probability in scenes with complex visibility and high-resolution geometry. We show that our method only requires a few samples to produce gradients with low bias and variance for challenging cases such as glossy reflections and shadows. Finally, we use our differentiable path tracer to reconstruct the 3D geometry and materials of several real-world objects from a set of reference photographs.

6.3. Expressive rendering

6.3.1. Procedural Stylization

Participants: Maxime Isnel, Mohamed Amine Farhat, Romain Vergne, Joëlle Thollot.

Stylizing 3D scenes is a long term goal for the expressive rendering community. During the master internship of Maxime Isnel we have worked on a procedurel approach based on a procedural solid noise used in image space to generate brush strokes or 2.5D visual primitives, such as fur. The overview of the approach is shown Figure 4 . This project is still in progress and will continue with a post-doc in 2020.

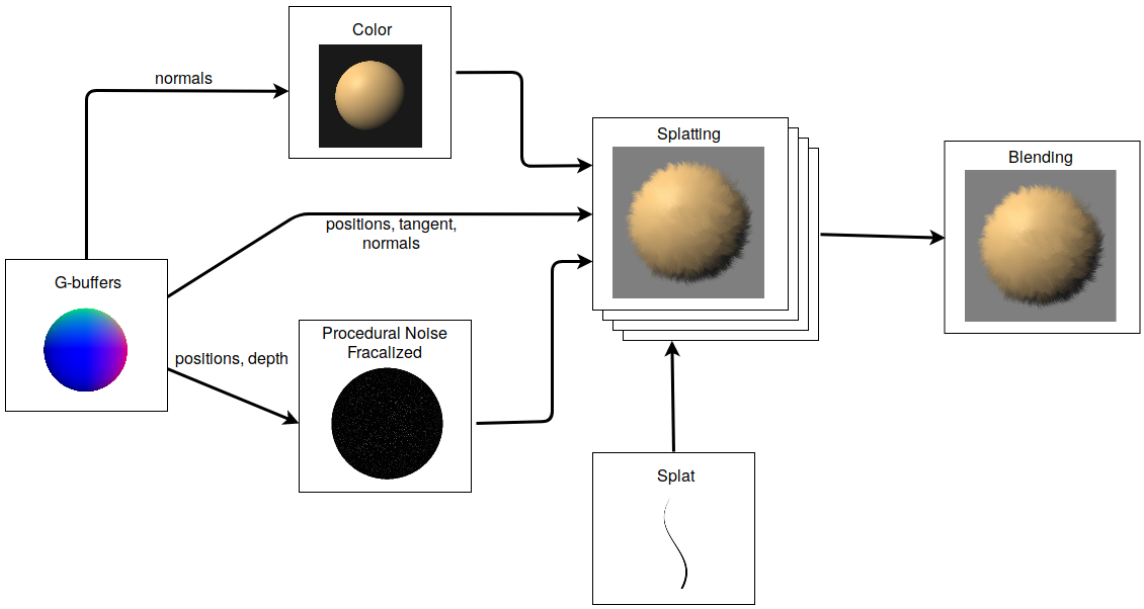


Figure 4. Based on a procedural solid noise and the use of geometry buffers, we propose an image-space approach to stylize a 3D object on the GPU.

MOEX Project-Team

6. New Results

6.1. Cultural knowledge evolution

Our cultural knowledge evolution work currently focusses on alignment evolution.

Agents may use ontology alignments to communicate when they represent knowledge with different ontologies: alignments help reclassifying objects from one ontology to the other. Such alignments may be provided by dedicated algorithms [4], but their accuracy is far from satisfying. Yet agents have to proceed. They can take advantage of their experience in order to evolve alignments: upon communication failure, they will adapt the alignments to avoid reproducing the same mistake.

We performed such repair experiments [3] and revealed that, by playing simple interaction games, agents can effectively repair random networks of ontologies or even create new alignments.

6.1.1. Modelling in dynamic epistemic logic

Participants: Manuel Atencia, Jérôme Euzenat, Line Van Den Berg [Correspondent].

We explored how closely these operators resemble logical dynamics. We developed a variant of Dynamic Epistemic Logic to capture the dynamics of the cultural alignment repair game. The ontologies are modelled as knowledge and alignments as beliefs in a variant of plausibility-based dynamic epistemic logic. The dynamics of the game is achieved through (public) announcement of the game issue and the adaptation operators are defined through conservative upgrades, i.e. modalities that transform models by reordering world-plausibility. This allowed us to formally establish some limitations and redundancy of the operators [9]. More precisely, for a complete logical reasoner, the operators are redundant and some may be inconsistent with the agent knowledge.

These results hold for one agent in the game but not necessarily for the other that may not know the classes by which the alignment is repaired, nor the relations between them. The former can be dealt with by declaring that agents are aware of the signature of both ontologies (public signature assumption) but this does not allow ontologies to evolve. We are currently investigating partial semantics as a more dynamic alternative solution to this problem.

This work is part of the PhD thesis of Line van den Berg.

6.1.2. Populations

Participants: Manuel Atencia, Fatima Danash, Jérôme Euzenat [Correspondent].

We started taking the population standpoint on experimental cultural evolution. For that purpose we introduced the concept of population within the experiments. So far, a population is characterised as a set of agents sharing the same ontology. Such agents play the same alignment repair games as before with agents of other populations.

The notion of population enables to experiment with different transmission mechanisms found in cultural evolution: vertical transmission, in which culture spreads, like genes, from parents to sibblings, and horizontal transmission, in which it spreads among all members of a population. We implemented explicit horizontal transmission through a synchronisation procedure in which, at a given interval, agents of the same population exchange their knowledge, i.e. alignments.

6.1.3. Link with interactor-replicator

Participant: Jérôme Euzenat [Correspondent].

Cultural evolution may be studied at a 'macro' level, inspired from population dynamics, or at a 'micro' level, inspired from genetics. The replicator-interactor model generalises the genotype-phenotype distinction of genetic evolution. We considered how it can be applied to cultural knowledge evolution experiments [8]. More specifically, we consider knowledge as the replicator and the behaviour it induces as the interactor. We showed that this requires to address problems concerning transmission. We discussed the introduction of horizontal transmission within the replicator-interactor model and/or differential reproduction within cultural evolution experiments.

6.1.4. Experiment reproducibility

Participants: Jimmy Avae, Robin Couret, Jérôme Euzenat [Correspondent].

Experiments are described and performed in our *Lazy lavender* platform which offers scripts to specify, run, and analyse experiments. This year, we investigated expressing experiment descriptions, i.e. design, results and analysis, in RDF. This facilitates the search of experiments based on structured queries that can be expressed in SPARQL: "which experiments have been performed but not analysed?", "which experiments are derived from another specific experiment?", "which hypotheses have not been confirmed since a precise release?", "which experiments test F-measure increase?". This also suggest a better organisation of our experiment reports.

6.2. Link keys

Link keys (§3.2) are explored following two directions:

- Extracting link keys;
- Reasoning with link keys.

6.2.1. Link key extraction with relational concept analysis

Participants: Manuel Atencia, Jérôme David [Correspondent], Jérôme Euzenat.

We first described our extraction approach [1] in the framework of formal context analysis (FCA, [20]). We recently showed that link keys extracted by formal concept analysis are equivalent to an extension of those which were extracted by our former algorithm [15]. We also used pattern structures, an extension of FCA with ordered structures, to reformulate this problem [6].

Furthermore, we used relational concept analysis (RCA, [22]), an extension of FCA taking relations between concepts into account. We showed that it is possible to encode the link key extraction problem in RCA to extract the optimal link keys even in the presence of cyclic dependencies [5]. Moreover, the proposed process does not require information about the alignments between the ontologies to find out from which pairs of classes to extract link keys.

We implemented these methods and evaluated them by reproducing the experiments made in previous studies. This shows that the method extracts the expected results as well as (also expected) scalability issues.

6.2.2. Combining link keys

Participants: Manuel Atencia, Alice Caporali, Jérôme David [Correspondent], Jérôme Euzenat, Basile Legal.

For certain data sets, it may be necessary to use several link keys, even on the same pair of classes, for retrieving a more complete link set. We introduced operators to combine link keys over the same pair of classes, investigated their relations and extended measures to evaluate their quality.

We specifically proposed strategies to extract disjunctions from RDF data and apply existing quality measures to evaluate them. We experimented with these strategies showing their benefits [7].

6.2.3. Tableau method for $\mathcal{ALC}+$Link key reasoning

Participants: Manuel Atencia [Correspondent], Jérôme Euzenat, Khadija Jradeh.

Link keys can also be thought of as axioms in a description logic. As such, they can contribute to infer ABox axioms, such as links, terminological axioms, or other link keys. This has important practical applications, such as link key inference, link key consistency and link key redundancy checking. Yet, no reasoning support existed for link keys.

We previously extended the tableau method designed for the $\mathcal{ALC}$ description logic to support reasoning with link keys in $\mathcal{ALC}$. We showed how this extension enables combining link keys with classical terminological reasoning with and without ABox and TBox and generating non-trivial link keys. We further extended the method and have proven that this extended method terminates, is sound, complete, and that its complexity is 2EXPTIME [11].

This work is part of the PhD thesis of Khadija Jradeh, co-supervised with Chan Le Duc (LIASD).

MORPHEO Project-Team

7. New Results

7.1. Surface Motion Capture Animation Synthesis

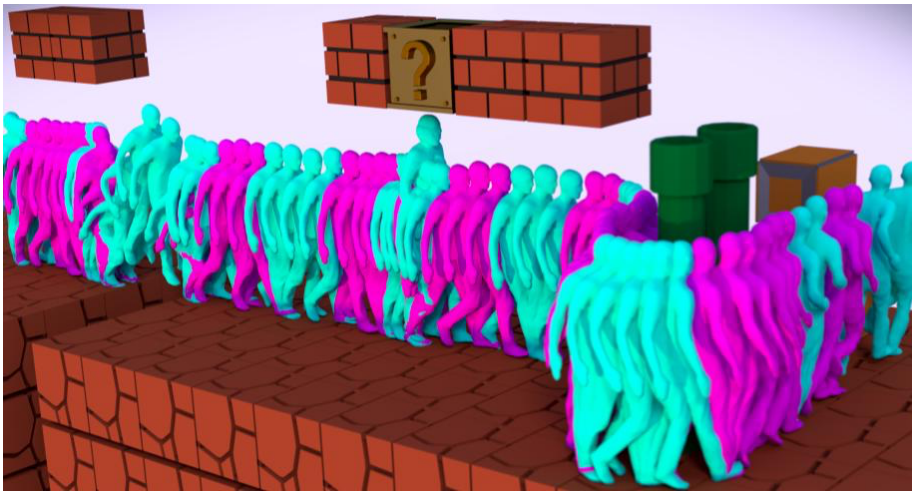


Figure 2. Animation Synthesis

We propose to generate novel animations from a set of elementary examples of video-based surface motion capture, under user-specified constraints. 4D surface capture animation is motivated by the increasing demand from media production for highly realistic 3D content. To this aim, data driven strategies that consider video-based information can produce animation with real shapes, kinematics and appearances. Our animations rely on the combination and the interpolation of textured 3D mesh data, which requires examining two aspects: (1) Shape geometry and (2) appearance. First, we propose an animation synthesis structure for the shape geometry, the Essential graph, that outperforms standard Motion graphs in optimality with respect to quantitative criteria, and we extend optimized interpolated transition algorithms to mesh data. Second, we propose a compact view-independent representation for the shape appearance. This representation encodes subject appearance changes due to viewpoint and illumination, and due to inaccuracies in geometric modelling independently. Besides providing compact representations, such decompositions allow for additional applications such as interpolation for animation (see figure 2).

This result was published in a prominent computer graphics journal, IEEE Transactions on Visualization and Computer Graphics [7].

7.2. CBCT of a Moving Sample from X-rays and Multiple Videos

We consider dense volumetric modeling of moving samples such as body parts. Most dense modeling methods consider samples observed with a moving X-ray device and cannot easily handle moving samples. We propose instead a novel method to observe shape motion from a fixed X-ray device and to build dense in-depth attenuation information. This yields a low-cost, low-dose 3D imaging solution, taking benefit of equipment widely available in clinical environments. Our first innovation is to combine a video-based surface motion

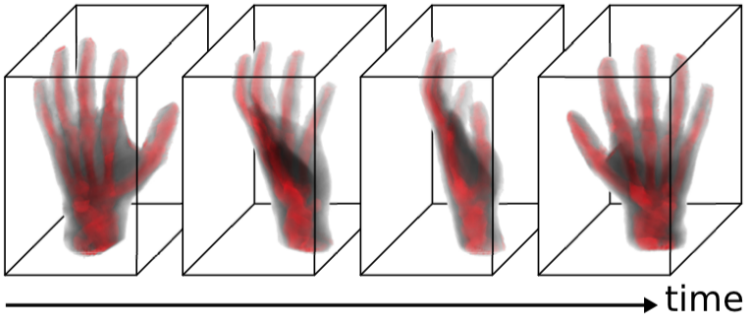


Figure 3. Dense volumetric attenuation reconstruction from a rigidly moving sample captured by a single planar X-ray imaging device and a surface motion capture system. Higher attenuation (here bone structure) is highlighted in red.

capture system with a single low-cost/low-dose fixed planar X-ray device, in order to retrieve the sample motion and attenuation information with minimal radiation exposure. Our second innovation is to rely on Bayesian inference to solve for a dense attenuation volume given planar radioscopic images of a moving sample. This approach enables multiple sources of noise to be considered and takes advantage of very limited prior information to solve an otherwise ill-posed problem. Results show that the proposed strategy is able to reconstruct dense volumetric attenuation models from a very limited number of radiographic views over time on synthetic and in-situ data, as illustrated in Figure 3 .

This result was published in a prominent medical journal, IEEE Transactions on Medical Imaging [9].

7.3. Learning and Tracking the 3D Body Shape of Freely Moving Infants from RGB-D sequences

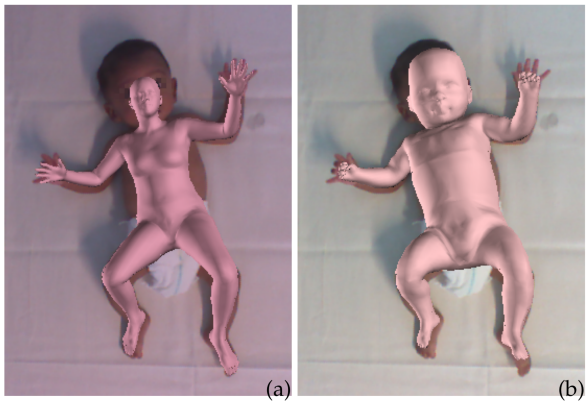


Figure 4. (a) Simply scaling a generic adult body model and fitting it to an infant does not work as body proportions significantly differ. (b) The proposed SMIL model properly captures the infants' shape and pose

Statistical models of the human body surface are generally learned from thousands of high-quality 3D scans in predefined poses to cover the wide variety of human body shapes and articulations. Acquisition of such data requires expensive equipment, calibration procedures, and is limited to cooperative subjects who can understand and follow instructions, such as adults. We presented a method for learning a statistical 3D Skinned Multi-Infant Linear body model (SMIL) from incomplete, low-quality RGB-D sequences of freely moving infants. Quantitative experiments show that SMIL faithfully represents the RGB-D data and properly factorizes the shape and pose of the infants. To demonstrate the applicability of SMIL, we fitted the model to RGB-D sequences of freely moving infants and show, with a case study, that our method captures enough motion detail for General Movements Assessment (GMA), a method used in clinical practice for early detection of neurodevelopmental disorders in infants. SMIL provides a new tool for analyzing infant shape and movement and is a step towards an automated system for GMA. This result was published in a prominent computer vision journal, IEEE Transactions on PAMI [8].

7.4. The Virtual Caliper: Rapid Creation of Metrically Accurate Avatars from 3D Measurements

Figure 5. Using the wand controllers of the HTC Vive, the Virtual Caliper produces a rigged 3D model with exactly the dimensions of the measured person.

Creating metrically accurate avatars is important for many applications such as virtual clothing try-on, ergonomics, medicine, immersive social media, telepresence, and gaming. Creating avatars that precisely represent a particular individual is challenging however, due to the need for expensive 3D scanners, privacy issues with photographs or videos, and difficulty in making accurate tailoring measurements. We overcome these challenges by creating "The Virtual Caliper", which uses VR game controllers to make simple measurements. First, we establish what body measurements users can reliably make on their own body. We find several distance measurements to be good candidates and then verify that these are linearly related to 3D body shape as represented by the SMPL body model. The Virtual Caliper enables novice users to accurately measure themselves and create an avatar with their own body shape. We evaluate the metric accuracy relative to ground truth 3D body scan data, compare the method quantitatively to other avatar creation tools, and perform extensive perceptual studies. We also provide a software application to the community that enables novices to rapidly create avatars in fewer than five minutes. Not only is our approach more rapid than existing methods, it exports a metrically accurate 3D avatar model that is rigged and skinned.

This result was published in a prominent computer graphics journal, IEEE Transactions on Visualization and Computer Graphics [10].

7.5. Adaptive Mesh Texture for Multi-View Appearance Modeling

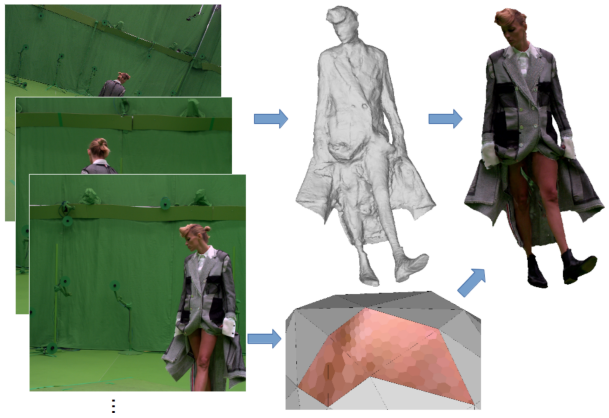


Figure 6. Texturing 3D models: given a set of input photographs (left), a geometric mesh is computed (top), along with an appearance function stored within the surface mesh structure (bottom).

Most applications in image based 3D modeling resort to texture maps, a 2D mapping of shape color information into image files. Despite their unquestionable merits, in particular the ability to apply standard image tools, including compression, image textures still suffer from limitations that result from the 2D mapping of information that originally belongs to a 3D structure. This is especially true with 2D texture atlases, a generic 2D mapping for 3D mesh models that introduces discontinuities in the texture space and plagues many 3D appearance algorithms. Moreover, the per-triangle texel density of 2D image textures cannot be individually adjusted to the corresponding pixel observation density without a global change in the atlas mapping function. To address these issues, we have proposed a new appearance representation for image-based 3D shape modeling, which stores appearance information directly on 3D meshes, rather than a texture atlas. We have shown this representation to allow for input-adaptive sampling and compression support. Our experiments demonstrated that it outperforms traditional image textures, in multi-view reconstruction contexts, with better visual quality and memory foot- print, which makes it a suitable tool when dealing with large amounts of data as with dynamic scene 3D models.

This result was published in the international conference on 3D Vision (3DV'19) [11].

7.6. Contact Preserving Shape Transfer for Motion Retargeting

Retargeting a motion from a source to a target character is an important problem in computer animation, as it allows to reuse existing rigged databases or transfer motion capture to virtual characters. Surface based pose transfer is a promising approach to avoid the trial-and-error process when controlling the joint angles. In this work we investigated whether shape transfer instead of pose transfer would better preserve the original contextual meaning of the source pose. To this end, we proposed an optimization-based method to deform the source shape+pose using three main energy functions: similarity to the target shape, body part volume preservation, and collision management (preserve existing contacts and prevent penetrations). The results show that this strategy is able to retarget complex poses, including several contacts, to very different morphologies. In particular, we introduced new contacts that are linked to the change in morphology, and which would be difficult to obtain with previous works based on pose transfer that aim at distance preservation between body parts.

This result was published in the ACM SIGGRAPH Conference on Motion Interaction and Games (MIG'19) [12].

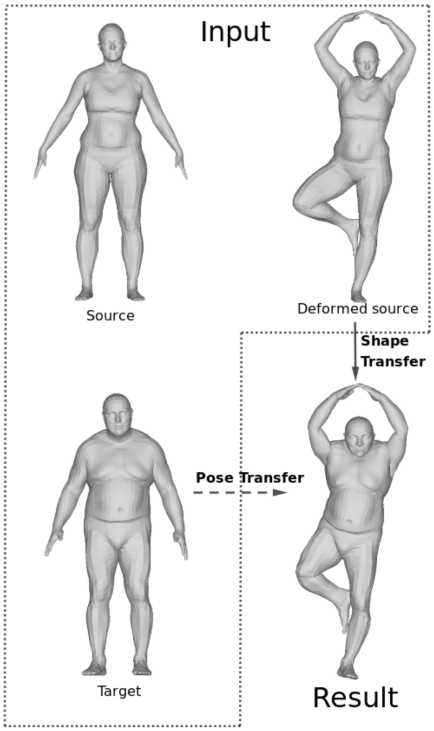


Figure 7. Motion Retargeting: Instead of transferring the pose from a source to a target shape, we propose to transfer the shape of the target to the deformed source character.

7.7. A Decoupled 3D Facial Shape Model by Adversarial Training

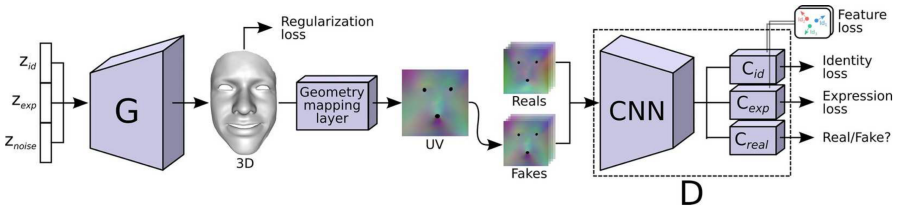


Figure 8. The face generator. Identity and expression codes zid , zexp are used to control the generator, and classification losses are added to decouple between the two. A feature loss is introduced to ensure consistency over features with fixed identities or expressions

Data-driven generative 3D face models are used to compactly encode facial shape data into meaningful parametric representations. A desirable property of these models is their ability to effectively decouple natural sources of variation, in particular identity and expression. While factorized representations have been proposed for that purpose, they are still limited in the variability they can capture and may present modeling artifacts when applied to tasks such as expression transfer. In this work, we explored a new direction with Generative Adversarial Networks and showed that they contribute to better face modeling performances, especially in decoupling natural factors, while also achieving more diverse samples. To train the model we introduced a novel architecture that combines a 3D generator with a 2D discriminator that leverages conventional CNNs, where the two components are bridged by a geometry mapping layer. We further presented a training scheme, based on auxiliary classifiers, to explicitly disentangle identity and expression attributes. Through quantitative and qualitative results on standard face datasets, we illustrated the benefits of our model and demonstrate that it outperforms competing state of the art methods in terms of decoupling and diversity.

This result was published in the international conference on computer vision (ICCV'19) [13]

7.8. Non-parametric 3D Human Shape Estimation from Single Images

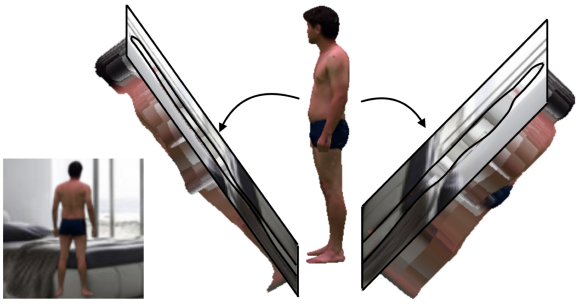


Figure 9. Given a single image, we estimate the "visible" and the "hidden" depth maps from the camera point of view. The two depth maps can be seen as the two halves of a virtual "mould".

In this work, we tackle the problem of 3D human shape estimation from single RGB images. While the recent progress in convolutional neural networks has allowed impressive results for 3D human pose estimation, estimating the full 3D shape of a person is still an open issue. Model-based approaches can output precise meshes of naked under-cloth human bodies but fail to estimate details and un-modelled elements such as hair or clothing. On the other hand, non-parametric volumetric approaches can potentially estimate complete shapes but, in practice, they are limited by the resolution of the output grid and cannot produce detailed estimates. In this work, we propose a non-parametric approach that employs a double depth map to represent the 3D shape of a person: a visible depth map and a "hidden" depth map are estimated and combined, to reconstruct the human 3D shape as done with a "mould". This representation through 2D depth maps allows a higher resolution output with a much lower dimension than voxel-based volumetric representations. Additionally, our fully derivable depth-based model allows us to efficiently incorporate a discriminator in an adversarial fashion to improve the accuracy and "humanness" of the 3D output. We train and quantitatively validate our approach on SURREAL and on 3D-HUMANS, a new photorealistic dataset made of semi-synthetic in-house images annotated with 3D ground truth surfaces.

This work was published in the international conference on computer vision (ICCV'19) [14]

7.9. Probabilistic Reconstruction Networks

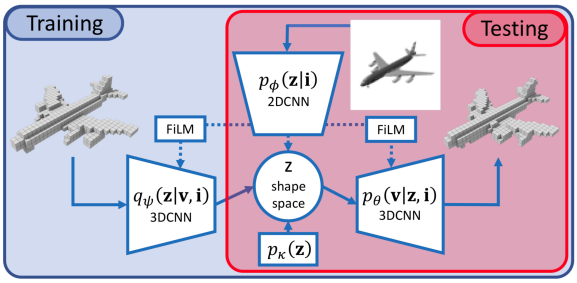


Figure 10. Probabilistic Reconstruction Networks for 3D shape inference from a single image. Arrows show the computational flow through the model, dotted arrows show optional image conditioning. The inference network q_Ψ is only used during training for variational inference

We study end-to-end learning strategies for 3D shape inference from images, in particular from a single image. Several approaches in this direction have been investigated that explore different shape representations and suitable learning architectures. We focus instead on the underlying probabilistic mechanisms involved and contribute a more principled probabilistic inference-based reconstruction framework, which we coin Probabilistic Reconstruction Networks. This framework expresses image conditioned 3D shape inference through a family of latent variable models, and naturally decouples the choice of shape representations from the inference itself. Moreover, it suggests different options for the image conditioning and allows training in two regimes, using either Monte Carlo or variational approximation of the marginal likelihood. Using our Probabilistic Reconstruction Networks we obtain single image 3D reconstruction results that set a new state of the art on the ShapeNet dataset in terms of the intersection over union and earth mover's distance evaluation metrics. Interestingly, we obtain these results using a basic voxel grid representation, improving over recent work based on finer point cloud or mesh based representations.

This work was published in the British machine vision conference (BMVC'19) [15] where it won the runner-up best paper award.

PERCEPTION Project-Team

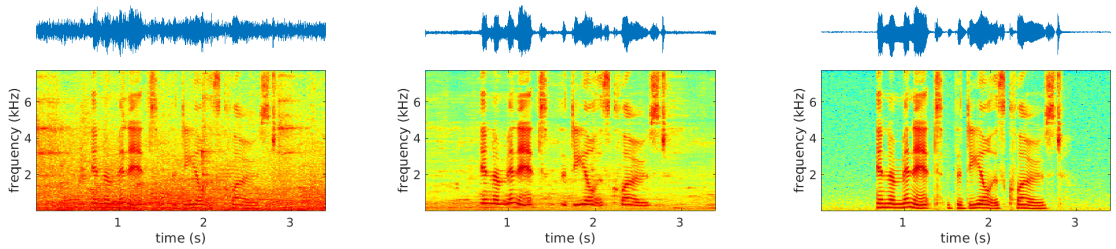
6. New Results

6.1. Multichannel Speech Separation and Enhancement Using the Convolutive Transfer Function

We addressed the problem of speech separation and enhancement from multichannel convolutive and noisy mixtures, *assuming known mixing filters*. We proposed to perform the speech separation and enhancement tasks in the short-time Fourier transform domain, using the convolutive transfer function (CTF) approximation [43], [44]. Compared to time-domain filters, CTF has much less taps, consequently it has less near-common zeros among channels and less computational complexity. The work proposes three speech-source recovery methods, namely: (i) the multichannel inverse filtering method, i.e. the multiple input/output inverse theorem (MINT), is exploited in the CTF domain, and for the multi-source case, (ii) a beamforming-like multichannel inverse filtering method applying single source MINT and using power minimization, which is suitable whenever the source CTFs are not all known, and (iii) a constrained Lasso method, where the sources are recovered by minimizing the ℓ_1-norm to impose their spectral sparsity, with the constraint that the ℓ_2-norm fitting cost, between the microphone signals and the mixing model involving the unknown source signals, is less than a tolerance. The noise can be reduced by setting a tolerance onto the noise power. Experiments under various acoustic conditions are carried out to evaluate the three proposed methods. The comparison between them as well as with the baseline methods is presented.

6.2. Speech Denoising and Enhancement with LTSMs

We have started to address the problems of multichannel speech denoising [45] and enhancement [51] in the short-time Fourier transform (STFT) domain and in the framework of sequence-to-sequence deep learning. In the case of denoising, the magnitude of noisy speech is mapped onto the noise power spectral density. In the case of speech enhancement, the noisy speech is mapped onto clean speech. A long short-time memory (LSTM) network takes as input a sequence of STFT coefficients associated with a frequency bin of multichannel noisy-speech signals. The network's output is a sequence of single-channel cleaned speech at the same frequency bin. We propose several clean-speech network targets, namely, the magnitude ratio mask, the complex ideal ratio mask, the STFT coefficients and spatial filtering [54]. A prominent feature of the proposed model is that the same LSTM architecture, with identical parameters, is trained across frequency bins. The proposed method is referred to as narrow-band deep filtering. This choice stays in contrast with traditional wide-band speech enhancement methods. The proposed deep filter is able to discriminate between speech and noise by exploiting their different temporal and spatial characteristics: speech is non-stationary and spatially coherent while noise is relatively stationary and weakly correlated across channels. This is similar in spirit with unsupervised techniques, such as spectral subtraction and beamforming. We describe extensive experiments with both mixed signals (noise is added to clean speech) and real signals (live recordings). We empirically evaluate the proposed architecture variants using speech enhancement and speech recognition metrics, and we compare our results with the results obtained with several state of the art methods. In the light of these experiments we conclude that narrow-band deep filtering has very good performance, and excellent generalization capabilities in terms of speaker variability and noise type, e.g. Figure 2 .
Website: https://team.inria.fr/perception/research/mse-lstm/.



Noisy speech input Filtered output Ground-truth speech signal

Figure 2. An example of narrow-band deep filtering for speech enhancement [54]. Waveforms and spectrograms of the noisy (unprocessed) input, the filtered output and the ground-truth clean-speech. Four microphones were used in this example. The signal-to-noise ratio in this example is 0 dB.

6.3. Multichannel Speech Enhancement with Variational Auto-Encoder

We addressed speaker-independent multichannel speech enhancement in unknown noisy environments. Our work is based on a well-established multichannel local Gaussian modeling framework. We propose to use a neural network for modeling the speech spectro-temporal content. The parameters of this supervised model are learned using the framework of variational autoencoders. The noisy recording environment is supposed to be unknown, so the noise spectro-temporal modeling remains unsupervised and is based on non-negative matrix factorization (NMF). We develop a Monte Carlo expectation-maximization algorithm and we experimentally show that the proposed approach outperforms its NMF-based counterpart, where speech is modeled using supervised NMF [49].
Website: https://team.inria.fr/perception/research/icassp-2019-mvae/

6.4. Audio-visual Speech Enhancement with Conditional Variational Auto-Encoder

Variational auto-encoders (VAEs) are deep generative latent variable models that can be used for learning the distribution of complex data. VAEs have been successfully used to learn a probabilistic prior over speech signals, which is then used to perform speech enhancement. One advantage of this generative approach is that it does not require pairs of clean and noisy speech signals at training. In this work, we propose audio-visual variants of VAEs for single-channel and speaker-independent speech enhancement. We developed a conditional VAE (CVAE) where the audio speech generative process is conditioned on visual information of the lip region, e.g. Figure 3 . At test time, the audio-visual speech generative model is combined with a noise model, based on nonnegative matrix factorization, and speech enhancement relies on a Monte Carlo expectation-maximization algorithm. Experiments were conducted with the recently published NTCD-TIMIT dataset. The results confirm that the proposed audio-visual CVAE effectively fuse audio and visual information, and it improves the speech enhancement performance compared with the audio-only VAE model, especially when the speech signal is highly corrupted by noise. We also showed that the proposed unsupervised audio-visual speech enhancement approach outperforms a state-of-the-art supervised deep learning method [55].
Website: https://team.inria.fr/perception/research/av-vae-se/

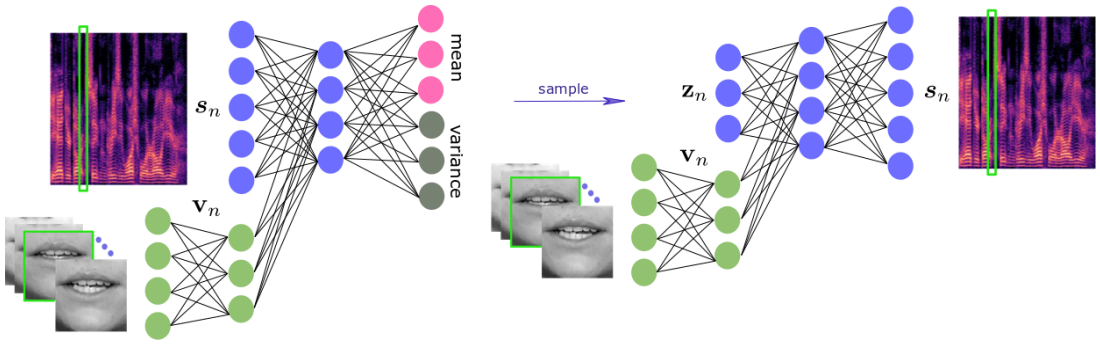


Figure 3. We proposed a conditional variational auto-encoder architecture for fusing audio and visual data for speech enhancement [55].

6.5. Variational Bayesian Inference of Audio-visual Speaker Tracking

We addressed the problem of tracking multiple speakers via the fusion of visual and auditory information [36]. We proposed to exploit the complementary nature of these two modalities in order to accurately estimate smooth trajectories of the tracked persons, to deal with the partial or total absence of one of the modalities over short periods of time, and to estimate the acoustic status – either speaking or silent – of each tracked person along time, e.g. Figure 1 . We proposed to cast the problem at hand into a generative audio-visual fusion (or association) model formulated as a latent-variable temporal graphical model. This may well be viewed as the problem of maximizing the posterior joint distribution of a set of continuous and discrete latent variables given the past and current observations, which is intractable. We proposed a variational inference model which amounts to approximate the joint distribution with a factorized distribution. The solution takes the form of closed-form expectation maximization procedures using Gaussian distributions [38]. We described in detail the inference algorithm, we evaluated its performance and we compared the results with several baseline methods. These experiments show that the proposed audio-visual tracker performs well in informal meetings involving a time-varying number of people. Real-time versions of the algorithm were implemented on our robotic platform [47].
Website: https://team.inria.fr/perception/research/var-av-track/.

6.6. Detection, Localization and Tracking of Multiple Audio Sources

We addressed the problem of online detection, localization and tracking of multiple moving speakers in reverberant environments [36]. The work has the following contributions. We used the direct-path relative transfer function (DP-RTF), an inter-channel feature that encodes acoustic information robust against reverberation, and we proposed an online algorithm well suited for estimating DP-RTFs associated with moving audio sources. Another crucial ingredient of the proposed method is its ability to properly assign DP-RTFs to audio-source directions. Towards this goal, we adopted a maximum-likelihood formulation and we proposed to use the exponentiated gradient (EG) to efficiently update source-direction estimates starting from their currently available values. The problem of multiple-speaker tracking is computationally intractable because the number of possible associations between observed source directions and physical speakers grows exponentially with time. We adopt a Bayesian framework and we proposed two variational approximations of the posterior filtering distributions associated with multiple speaker tracking, as well as two efficient variational expectation maximization (VEM) solvers [41], [37]. The proposed online localization and tracking methods were thoroughly evaluated using two datasets that contain recordings performed in real environments.
Websites:

https://team.inria.fr/perception/research/audiotrack-vonm/
https://team.inria.fr/perception/research/multi-speaker-tracking/.

6.7. The Kinovis Multiple-Speaker Tracking Datasets

The Kinovis multiple speaker tracking (Kinovis-MST) datasets contain live acoustic recordings of multiple moving speakers in a reverberant environment. The data were recorded in the Kinovis multiple-camera laboratory at Inria Grenoble Rhône-Alpes. The room size is $10.2 \times 9.9 \times 5.6$ meters with T60 = 0.53 seconds. The data were recorded with four microphones embedded into the head of a NAO robot. Because there is a fan located inside the robot head nearby the microphones, there is a fair amount of stationary and spatially correlated microphone noise. The signal-to-noise ratio of the microphone signals is of approximately 2.7 dB. The recordings contain between one and three moving participants that speak naturally, hence the number of active speech sources varies over time. The robot-to-speaker distance ranges between 1.5 and 3.5 meters. Ground-truth trajectories and speech activity information were obtained in the following way. Participants were wearing optical markers placed on their heads such that the Kinovis motion capture system provides accurate 3D trajectories for each participant. Moreover, an infrared marker is placed on the participants' foreheads. This enables the identification of each participant over time. Whenever time a participant is silent, he/she hides his/her infrared marker, thus allowing speaking/silent annotations of the recordings.
Website: https://team.inria.fr/perception/the-kinovis-mst-dataset/.

6.8. Deep Regression

Deep learning revolutionized data science, and recently its popularity has grown exponentially, as did the amount of papers employing deep networks. Vision tasks, such as human pose estimation, did not escape from this trend. There is a large number of deep models, where small changes in the network architecture, or in the data pre-processing, together with the stochastic nature of the optimization procedures, produce notably different results, making extremely difficult to sift methods that significantly outperform others. This situation motivates the current study, in which we perform a systematic evaluation and statistical analysis of vanilla deep regression, i.e. convolutional neural networks with a linear regression top layer. This is the first comprehensive analysis of deep regression techniques. We perform experiments on four vision problems, and report confidence intervals for the median performance as well as the statistical significance of the results, if any. Surprisingly, the variability due to different data pre-processing procedures generally eclipses the variability due to modifications in the network architecture. Our results reinforce the hypothesis according to which, in general, a general-purpose network (e.g. VGG-16 or ResNet-50) adequately tuned can yield results close to the state-of-the-art without having to resort to more complex and ad-hoc regression models, [40].
Website: https://team.inria.fr/perception/research/deep-regression/.

6.9. Deep Reinforcement Learning for Audio-Visual Robot Control

More recently, we investigated the use of reinforcement learning (RL) as an alternative to sensor-based robot control. The robotic task consists of turning the robot head (gaze control) towards speaking people. The method is more general in spirit than visual (or audio) servoing because it can handle an arbitrary number of speaking or non speaking persons and it can improve its behavior online, as the robot experiences new situations. An overview of the proposed method is shown in Fig. 4 . The reinforcement learning formulation enables a robot to learn where to look for people and to favor speaking people via a trial-and-error strategy.

Past, present and future HRI developments require datasets for training, validation, test as well as for benchmarking. HRI datasets are challenging because it is not easy to record realistic interactions between a robot and users. RL avoids systematic recourse to annotated datasets for training. In [39] we proposed the use of a simulated environment for pre-training the RL parameters, thus avoiding spending hours of tedious interaction.
Website: https://team.inria.fr/perception/research/deep-rl-for-gaze-control/.

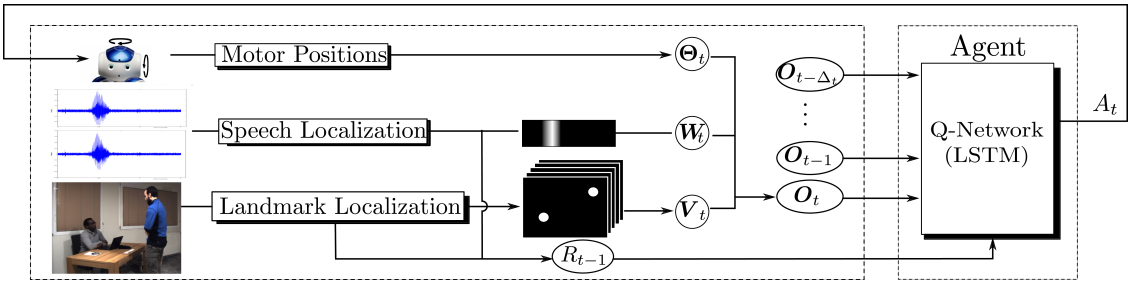


Figure 4. Overview of the proposed deep RL method for controlling the gaze of a robot. At each time index t, audio and visual data are represented as binary maps which, together with motor positions, form the set of observations $\mathbf{O}_t$. A motor action A_t (rotate the head left, right, up, down, or stay still) is selected based on past and present observations via maximization of current and future rewards. The rewards R are based on the number of visible persons as well as on the presence of speech sources in the camera field of view. We use a deep Q-network (DQN) model that can be learned both off-line and on-line. Please consult [39] for further details.

PERVASIVE Project-Team

6. New Results

6.1. Observing and Modelling Expertise and Awareness from Eye-gaze and Emotion

Participants: Thomas Guntz, James Crowley, Dominique Vaufreydaz, Philippe Dessus, Raffaella Balzarini.

We have constructed an instrument for capturing and interpreting multimodal signals of humans engaged in solving challenging problems. Our instrument captures eye gaze, fixations, body postures, and facial expressions signals from humans engaged in interactive tasks on a touch screen. We use a 23 inch Touch-Screen computer, a Kinect 2.0 mounted 35 cm above the screen to observe the subject, a 1080p Webcam for a frontal view, a Tobii Eye-Tracking bar (Pro X2-60 screen-based) and two adjustable USB-LED for lighting condition control. A wooden structure is used to rigidly mount the measuring equipment in order to assure identical sensor placement and orientation for all recordings.

As a pilot study, we observed expert chess players engaged in solving problems of increasing difficulty]. Our initial hypothesis was that we could directly detect awareness of significant configurations of chess pieces (chunks) from eye-scan and physiological measurements of emotion in reaction to game situation. The pilot experiment demonstrated that this initial hypothesis was overly simplistic.

In order to better understand the phenomena observed in our pilot experiment, we have constructed a model of the cognitive processes involved, using theories from cognitive science and classic (symbolic) artificial intelligence. This model is a very partial description that allows us to ask questions and make predictions to guide future experiments. Our model posits that experts reason with a situation model that is strongly constrained by limits to the number of entities and relations that may be considered at a time. This limitation forces subjects to construct abstract concepts (chunks) to describe game play, in order to explore alternative moves. Expert players retain associations of situations with emotions in long-term memory. The rapid changes in emotion correspond to recognition of previously encountered situations during exploration of the game tree. Recalled emotions guide selection of situation models for reasoning. This hypothesis is in accordance with Damasio's Somatic Marker hypothesis, which posits that emotions guide behavior, particularly when cognitive processes are overloaded.

Our hypothesis is that the subject uses the evoked emotions to select from the many possible situations for reasoning about moves during orientation and exploration. With this interpretation, the player rapidly considers partial descriptions as situations composed of a limited number of perceived chunks. Recognition of situations from experience evokes emotions that are displayed as face expressions and body posture.

With this hypothesis, valence, arousal and dominance are learned from experience and associated with chess situations in long-term memory to guide reasoning in chess. Dominance corresponds to the degree of experience with the recognized situation. As players gain experience with alternate outcomes for a situation, they become more assured in their ability to spot opportunities and avoid dangers. Valence corresponds to whether the situation is recognized as favorable (providing opportunities) or unfavorable (creating threats). Arousal corresponds to the imminence of a threat or opportunity. A defensive player will give priority to reasoning about unfavorable situations and associated dangers. An aggressive player will seek out high valence situations. All players will give priority to situations that evoke strong arousal. The amount of effort that player will expend exploring a situation can determined by dominance.

6.2. Recognition, Modelling and Description of Manipulation Actions

Participants: Nachwa Abou Bakr, James Crowley.

A full understanding of human actions requires: recognizing what action has been performed, predicting how it will affect the surrounding environment, explaining why this action has been performed, and who is performing it. Classic approaches to action recognition interpret a spatio-temporal pattern in a video sequence to tell what action has been performed, and perhaps how and where it was performed. A more complete understanding requires information about why the action was performed, and how it affects the environment. This face of understanding can be provided by explaining the action as part of a narrative.

We have addressed the problem of recognition, modelling and description of human activities, with results on three problems: (1) the use of transfer learning for simultaneous visual recognition of objects and object states, (2) the recognition of manipulation actions from state transitions, and (3) the interpretation of a series of actions and states as events in a predefined story to construct a narrative description.

These results have been developed using food preparation activities as an experimental domain. We start by recognizing food classes such as tomatoes and lettuce and food states, such as sliced and diced, during meal preparation. We adapt the VGG network architecture to jointly learn the representations of food items and food states using transfer learning. We model actions as the transformation of object states. We use recognised object properties (state and type) to detect corresponding manipulation actions by tracking object transformations in the video. Experimental performance evaluation for this approach is provided using the 50 salads and EPIC-Kitchen datasets. We use the resulting action descriptions to construct narrative descriptions for complex activities observed in videos of 50 salads dataset.

THOTH Project-Team

7. New Results

7.1. Visual Recognition and Robotics

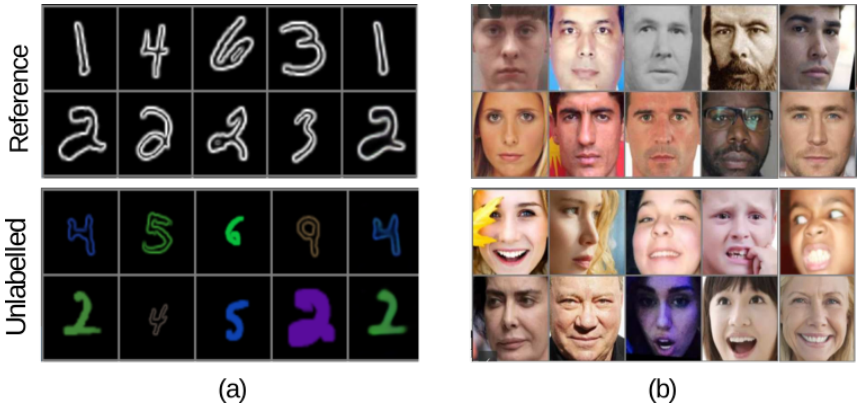


Figure 1. Illustration of different reference-based disentangling problems. (a) Disentangling style from digits. The reference distribution is composed by numbers with a fixed style (b) Disentangling factors of variations related with facial expressions. Reference images correspond to neutral faces. Note that pairing information between unlabelled and reference images is not available during training.

7.1.1. Learning Disentangled Representations with Reference-Based Variational Autoencoders

Participants: Adria Ruiz, Oriol Martinez, Xavier Binefa, Jakob Verbeek.

Learning disentangled representations from visual data, where different high-level generative factors are independently encoded, is of importance for many computer vision tasks. Supervised approaches, however, require a significant annotation effort in order to label the factors of interest in a training set. To alleviate the annotation cost, in [32] we introduce a learning setting which we refer to as "reference-based disentangling". Given a pool of unlabelled images, the goal is to learn a representation where a set of target factors are disentangled from others. The only supervision comes from an auxiliary "reference set" that contains images where the factors of interest are constant. See Fig. 1 for illustrative examples. In order to address this problem, we propose reference-based variational autoencoders, a novel deep generative model designed to exploit the weak supervisory signal provided by the reference set. During training, we use the variational inference framework where adversarial learning is used to minimize the objective function. By addressing tasks such as feature learning, conditional image generation or attribute transfer, we validate the ability of the proposed model to learn disentangled representations from minimal supervision.

7.1.2. Tensor Decomposition and Non-linear Manifold Modeling for 3D Head Pose Estimation

Participants: Dmytro Derkach, Adria Ruiz, Federico M. Sukno.

Head pose estimation is a challenging computer vision problem with important applications in different scenarios such as human-computer interaction or face recognition. In [5], we present a 3D head pose estimation algorithm based on non-linear manifold learning. A key feature of the proposed approach is that it allows modeling the underlying 3D manifold that results from the combination of rotation angles. To do so, we use tensor decomposition to generate separate subspaces for each variation factor and show that each of them has a clear structure that can be modeled with cosine functions from a unique shared parameter per angle (see Fig. 2). Such representation provides a deep understanding of data behavior. We show that the proposed framework can be applied to a wide variety of input features and can be used for different purposes. Firstly, we test our system on a publicly available database, which consists of 2D images and we show that the cosine functions can be used to synthesize rotated versions from an object from which we see only a 2D image at a specific angle. Further, we perform 3D head pose estimation experiments using other two types of features: automatic landmarks and histogram-based 3D descriptors. We evaluate our approach on two publicly available databases, and demonstrate that angle estimations can be performed by optimizing the combination of these cosine functions to achieve state-of-the-art performance.

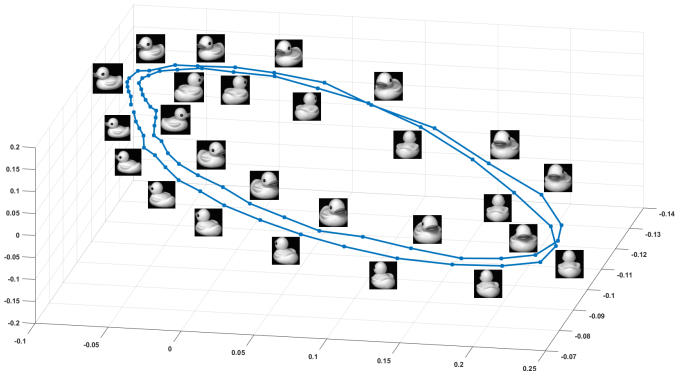


Figure 2. Visualization of the first three coefficients of the pose variation subspace for a dataset of single object rotated about the vertical axis.

7.1.3. Spreading vectors for similarity search

Participants: Alexandre Sablayrolles, Matthijs Douze, Cordelia Schmid, Hervé Jégou.

Discretizing multi-dimensional data distributions is a fundamental step of modern indexing methods. State-of-the-art techniques learn parameters of quantizers on training data for optimal performance, thus adapting quantizers to the data. In this work [29], we propose to reverse this paradigm and adapt the data to the quantizer: we train a neural net which last layer forms a fixed parameter-free quantizer, such as pre-defined points of a hyper-sphere. As a proxy objective, we design and train a neural network that favors uniformity in the spherical latent space, while preserving the neighborhood structure after the mapping. We propose a new regularizer derived from the Kozachenko–Leonenko differential entropy estimator to enforce uniformity and combine it with a locality-aware triplet loss. Experiments show that our end-to-end approach outperforms most learned quantization methods, and is competitive with the state of the art on widely adopted benchmarks. Furthermore, we show that training without the quantization step results in almost no difference in accuracy, but yields a generic catalyzer 3 that can be applied with any subsequent quantizer. The code is available online.

7.1.4. Diversity with Cooperation: Ensemble Methods for Few-Shot Classification

Participants: Nikita Dvornik, Cordelia Schmid, Julien Mairal.

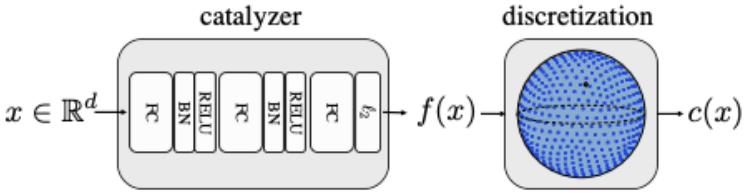


Figure 3. Our method learns a network that encodes the input space $\mathbb{R}^d$ into a code $c(x)$. It is learned end-to-end, yet the part of the network in charge of the discretization operation is fixed in advance, thereby avoiding optimization problems. The learnable function f, namely the "catalyzer", is optimized to increase the quality of the subsequent coding stage.

Few-shot classification consists of learning a predictive model that is able to effectively adapt to a new class, given only a few annotated samples. To solve this challenging problem, meta-learning has become a popular paradigm that advocates the ability to "learn to adapt". Recent works have shown, however, that simple learning strategies without meta-learning could be competitive. In our ICCV'19 paper [17], we go a step further and show that by addressing the fundamental high-variance issue of few-shot learning classifiers, it is possible to significantly outperform current meta-learning techniques. Our approach consists of designing an ensemble of deep networks to leverage the variance of the classifiers, and introducing new strategies to encourage the networks to cooperate, while encouraging prediction diversity, as illustrated in Figure 4 . Evaluation is conducted on the mini-ImageNet and CUB datasets, where we show that even a single network obtained by distillation yields state-of-the-art results.

7.1.5. Unsupervised Pre-Training of Image Features on Non-Curated Data

Participants: Mathilde Caron, Piotr Bojanowski [Facebook AI], Julien Mairal, Armand Joulin [Facebook AI].

Pre-training general-purpose visual features with convolutional neural networks without relying on annotations is a challenging and important task. Most recent efforts in unsupervised feature learning have focused on either small or highly curated datasets like ImageNet, whereas using non-curated raw datasets was found to decrease the feature quality when evaluated on a transfer task. Our goal is to bridge the performance gap between unsupervised methods trained on curated data, which are costly to obtain, and massive raw datasets that are easily available. To that effect, we propose a new unsupervised approach, DeeperCluster [13], described in Figure 5 which leverages self-supervision and clustering to capture complementary statistics from large-scale data. We validate our approach on 96 million images from YFCC100M, achieving state-of-the-art results among unsupervised methods on standard benchmarks, which confirms the potential of unsupervised learning when only non-curated raw data are available. We also show that pre-training a supervised VGG-16 with our method achieves 74.9% top-1 classification accuracy on the validation set of ImageNet, which is an improvement of $+0.8\%$ over the same network trained from scratch.

7.1.6. Learning to Augment Synthetic Images for Sim2Real Policy Transfer

Participants: Alexander Pashevich, Robin Strudel [Inria WILLOW], Igor Kalevatykh [Inria WILLOW], Ivan Laptev [Inria WILLOW], Cordelia Schmid.

Vision and learning have made significant progress that could improve robotics policies for complex tasks and environments. Learning deep neural networks for image understanding, however, requires large amounts of domain-specific visual data. While collecting such data from real robots is possible, such an approach

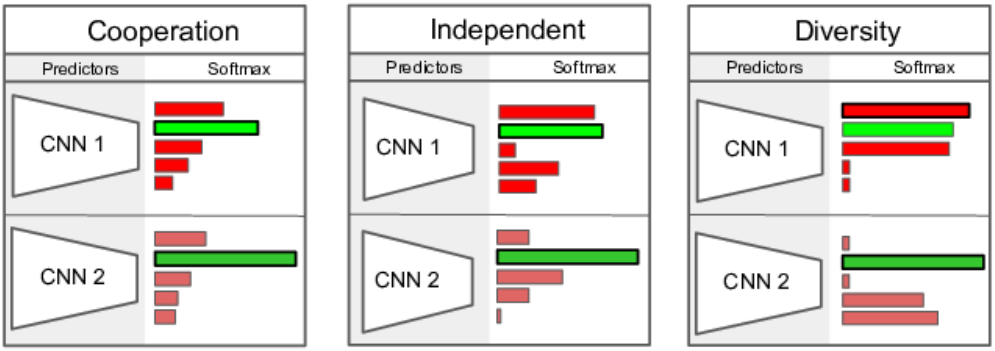


Figure 4. ***Illustration of the cooperation and diversity strategies on two networks.*** *All networks receive the same image as input and compute corresponding class probabilities with softmax. Cooperation encourages the non-ground truth probabilities (in red) to be similar, after normalization, whereas diversity encourages orthogonality.*

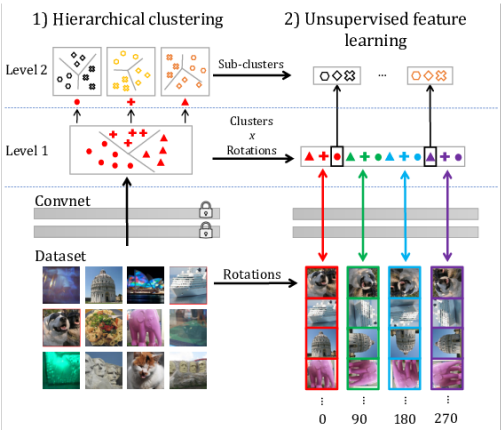


Figure 5. DeeperCluster alternates between a hierachical clustering of the features and learning the parameters of a convnet by predicting both the rotation angle and the cluster assignments in a single hierachical loss.

limits the scalability as learning policies typically requires thousands of trials. In this work [25] we attempt to learn manipulation policies in simulated environments. Simulators enable scalability and provide access to the underlying world state during training. Policies learned in simulators, however, do not transfer well to real scenes given the domain gap between real and synthetic data. We follow recent work on domain randomization and augment synthetic images with sequences of random transformations. Our main contribution is to optimize the augmentation strategy for sim2real transfer and to enable domain-independent policy learning, as illustrated in Figure 6 . We design an efficient search for depth image augmentations using object localization as a proxy task. Given the resulting sequence of random transformations, we use it to augment synthetic depth images during policy learning. Our augmentation strategy is policy-independent and enables policy learning with no real images. We demonstrate our approach to significantly improve accuracy on three manipulation tasks evaluated on a real robot.

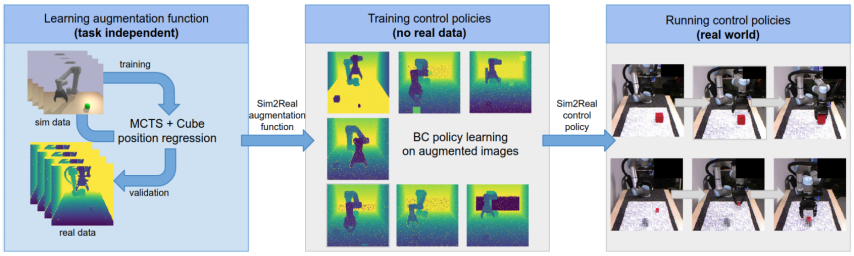


Figure 6. Overview of the method. Our contribution is the policy-independent learning of depth image augmentations (left). The resulting sequence of augmentations is applied to synthetic depth images while learning manipulation policies in a simulator (middle). The learned policies are directly applied to real robot scenes without finetuning on real images.

7.1.7. Learning to combine primitive skills: A step towards versatile robotic manipulation

Participants: Robin Strudel [Inria WILLOW], Alexander Pashevich, Igor Kalevatykh [Inria WILLOW], Ivan Laptev [Inria WILLOW], Josef Sivic [Inria WILLOW], Cordelia Schmid.

Manipulation tasks such as preparing a meal or assembling furniture remain highly challenging for robotics and vision. Traditional task and motion planning (TAMP) methods can solve complex tasks but require full state observability and are not adapted to dynamic scene changes. Recent learning methods can operate directly on visual inputs but typically require many demonstrations and/or task-specific reward engineering. In this work [40] we aim to overcome previous limitations and propose a reinforcement learning (RL) approach to task planning that learns to combine primitive skills illustrated in Figure 7 . First, compared to previous learning methods, our approach requires neither intermediate rewards nor complete task demonstrations during training. Second, we demonstrate the versatility of our vision-based task planning in challenging settings with temporary occlusions and dynamic scene changes. Third, we propose an efficient training of basic skills from few synthetic demonstrations by exploring recent CNN architectures and data augmentation. Notably, while all of our policies are learned on visual inputs in simulated environments, we demonstrate the successful transfer and high success rates when applying such policies to manipulation tasks on a real UR5 robotic arm.

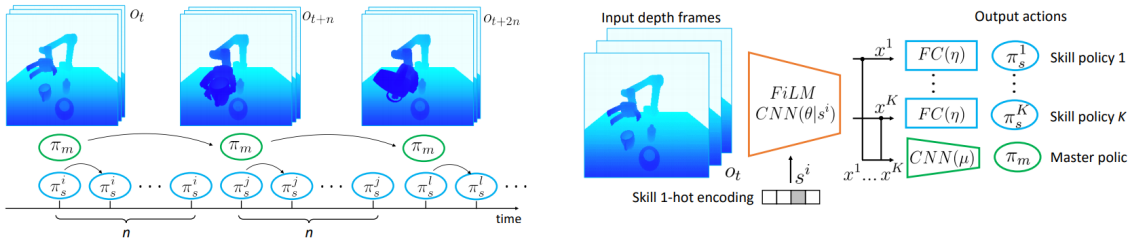


Figure 7. Illustration of our approach. (Left): Temporal hierarchy of master and skill policies. The master policy π_m is executed at a coarse interval of n time-steps to select among K skill policies $\pi_s^1 ... \pi_s^K$. Each skill policy generates control for a primitive action such as grasping or pouring. (Right): CNN architecture used for the skill and master policies.

7.1.8. Probabilistic Reconstruction Networks for 3D Shape Inference from a Single Image

Participants: Roman Klokov, Jakob Verbeek, Edmond Boyer [Inria Morpheo].

In our BMVC'19 paper [21], we study end-to-end learning strategies for 3D shape inference from images, in particular from a single image. Several approaches in this direction have been investigated that explore different shape representations and suitable learning architectures. We focus instead on the underlying probabilistic mechanisms involved and contribute a more principled probabilistic inference-based reconstruction framework, which we coin Probabilistic Reconstruction Networks. This framework expresses image conditioned 3D shape inference through a family of latent variable models, and naturally decouples the choice of shape representations from the inference itself. Moreover, it suggests different options for the image conditioning and allows training in two regimes, using either Monte Carlo or variational approximation of the marginal likelihood. Using our Probabilistic Reconstruction Networks we obtain single image 3D reconstruction results that set a new state of the art on the ShapeNet dataset in terms of the intersection over union and earth mover's distance evaluation metrics. Interestingly, we obtain these results using a basic voxel grid representation, improving over recent work based on finer point cloud or mesh based representations. In Figure 8 we show a schematic overview of our model.

7.1.9. Hierarchical Scene Coordinate Classification and Regression for Visual Localization

Participants: Xiaotian Li [Aalto Univ., Finland], Shuzhe Wang [Aalto Univ., Finland], Li Zhao [Aalto Univ., Finland], Jakob Verbeek, Juho Kannala [Aalto Univ., Finland].

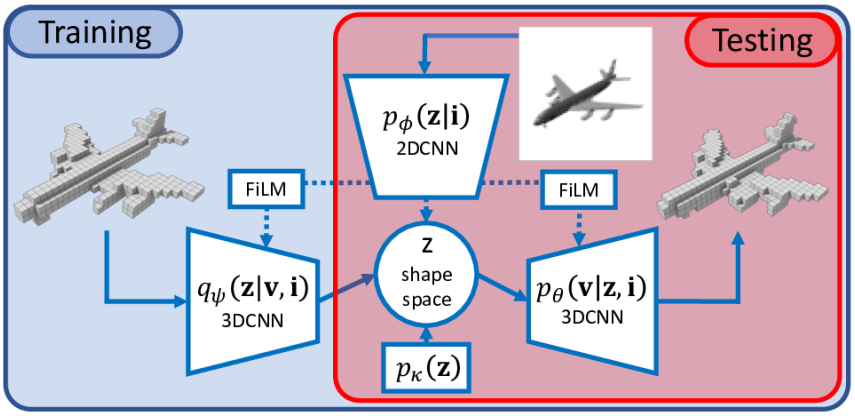


Figure 8. Probabilistic Reconstruction Networks for 3D shape inference from a single image. Arrows show the computational flow through the model, dotted arrows show optional image conditioning. Conditioning between 2D and 3D tensors is achieved by means of FiLM layers. The inference network q_ψ is only used during training for variational inference.

Visual localization is critical to many applications in computer vision and robotics. To address single-image RGB localization, state-of-the-art feature-based methods match local descriptors between a query image and a pre-built 3D model. Recently, deep neural networks have been exploited to regress the mapping between raw pixels and 3D coordinates in the scene, and thus the matching is implicitly performed by the forward pass through the network. However, in a large and ambiguous environment, learning such a regression task directly can be difficult for a single network. In our paper [37], we present a new hierarchical scene coordinate network to predict pixel scene coordinates in a coarse-to-fine manner from a single RGB image. The network consists of a series of output layers with each of them conditioned on the previous ones. The final output layer predicts the 3D coordinates and the others produce progressively finer discrete location labels. The proposed method outperforms the baseline regression-only network and allows us to train single compact models which scale robustly to large environments. It sets a new state-of-the-art for single-image RGB localization performance on the 7-Scenes, 12-Scenes, Cambridge Landmarks datasets, and three combined scenes. Moreover, for large-scale outdoor localization on the Aachen Day-Night dataset, our approach is much more accurate than existing scene coordinate regression approaches, and reduces significantly the performance gap w.r.t. explicit feature matching approaches. In Figure 9 we illustrate the scene coordinate predictions for the Aachen dataset experiments.

7.1.10. Moulding Humans: Non-parametric 3D Human Shape Estimation from Single Images

Participants: Valentin Gabeur, Jean-Sébastien Franco [Inria Morpheo], Xavier Martin, Cordelia Schmid, Gregory Rogez [NAVER LABS Europe].

While the recent progress in convolutional neural networks has allowed impressive results for 3D human pose estimation, estimating the full 3D shape of a person is still an open issue. Model-based approaches can output precise meshes of naked under-cloth human bodies but fail to estimate details and un-modelled elements such as hair or clothing. On the other hand, non-parametric volumetric approaches can potentially estimate complete shapes but, in practice, they are limited by the resolution of the output grid and cannot produce detailed estimates. In this paper [19], we propose a non-parametric approach that employs a double depth map 10 to represent the 3D shape of a person: a visible depth map and a "hidden" depth map are estimated and combined, to reconstruct the human 3D shape as done with a "mould". This representation through 2D

Figure 9. The scene coordinate predictions are visualized as 2D-2D matches between the query (left) and database (right) images. For each pair, the retrieved database image with the largest number of inliers is selected, and only the inlier matches are visualized. We show that our method is able to produce accurate correspondences for challenging queries.

depth maps allows a higher resolution output with a much lower dimension than voxel-based volumetric representations.

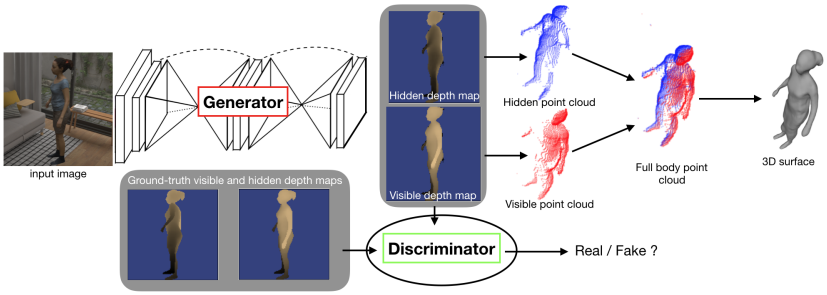


Figure 10. Given a single image, we estimate the "visible" and the "hidden" depth maps. The 3D point clouds of these 2 depth maps are combined to form a full-body 3D point cloud, as if lining up the 2 halves of a "mould". The 3D shape is then reconstructed using Poisson reconstruction. An adversarial training with a discriminator is employed to increase the humanness of the estimation.

7.1.11. Focused Attention for Action Recognition

Participants: Vladyslav Sydorov, Karteek Alahari.

In this paper [30], we introduce an attention model for video action recognition that allows processing video in higher resolution, by focusing on the relevant regions first. The network-specific saliency is utilized to guide the cropping, we illustrate the procedure in Figure 11 . We show performance improvement on the Charades dataset with this strategy.

7.2. Statistical Machine Learning

7.2.1. A Contextual Bandit Bake-off

Participants: Alberto Bietti, Alekh Agarwal [Microsoft Research], John Langford [Microsoft Research].

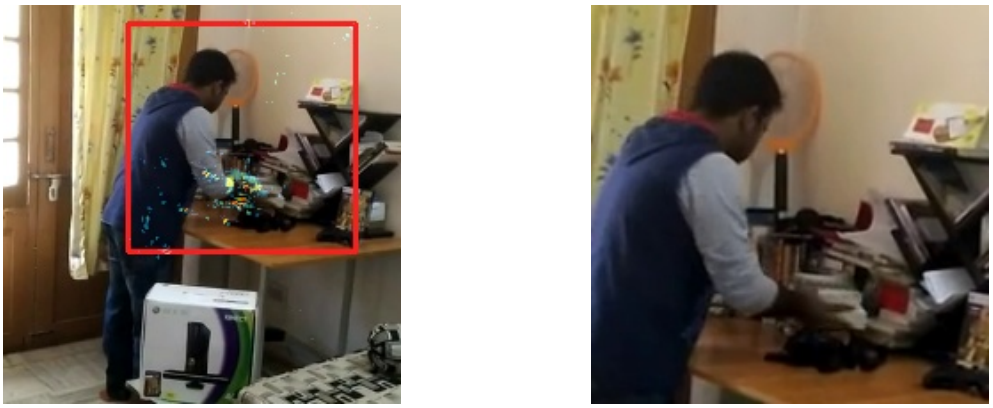


Figure 11. Example of attention on Charades action recognition dataset. (Left) Saliency scores (displayed as a heatmap) are localized around the object, a box maximizing the saliency measure within is selected. (Right) The network is provided with the relevant crop of the video, and can process it at a higher resolution.

Contextual bandit algorithms are essential for solving many real-world interactive machine learning problems. Despite multiple recent successes on statistically and computationally efficient methods, the practical behavior of these algorithms is still poorly understood. In , we leverage the availability of large numbers of supervised learning datasets to compare and empirically optimize contextual bandit algorithms, focusing on practical methods that learn by relying on optimization oracles from supervised learning. We find that a recent method using optimism under uncertainty works the best overall. A surprisingly close second is a simple greedy baseline that only explores implicitly through the diversity of contexts, followed by a variant of Online Cover which tends to be more conservative but robust to problem specification by design. Along the way, we also evaluate and improve several internal components of contextual bandit algorithm design. Overall, this is a thorough study and review of contextual bandit methodology.

7.2.2. A Generic Acceleration Framework for Stochastic Composite Optimization

Participants: Andrei Kulunchakov, Julien Mairal.

In [35], we introduce various mechanisms to obtain accelerated first-order stochasticoptimization algorithms when the objective function is convex or stronglyconvex. Specifically, we extend the Catalyst approach originally designed fordeterministic objectives to the stochastic setting. Given an optimizationmethod with mild convergence guarantees for strongly convex problems,the challenge is to accelerate convergence to a noise-dominated region, andthen achieve convergence with an optimal worst-case complexity depending on thenoise variance of the gradients.A side contribution of our work is also a generic analysis that canhandle inexact proximal operators, providing new insights about the robustness of stochastic algorithms when the proximal operator cannot be exactly computed. An illustration from this work is explained in Figure 12 .

7.2.3. Estimate Sequences for Variance-Reduced Stochastic Composite Optimization

Participants: Andrei Kulunchakov, Julien Mairal.

In [23], we propose a unified view of gradient-based algorithms for stochastic convex composite optimization. By extending the concept of estimate sequence introduced by Nesterov, we interpret a large class of stochastic optimization methods as procedures that iteratively minimize a surrogate of the objective. This point of view covers stochastic gradient descent (SGD), the variance-reduction approaches SAGA, SVRG, MISO, their proximal variants, and has several advantages: (i) we provide a simple generic proof of convergence for all of the aforementioned methods; (ii) we naturally obtain new algorithms with the same guarantees; (iii) we derive

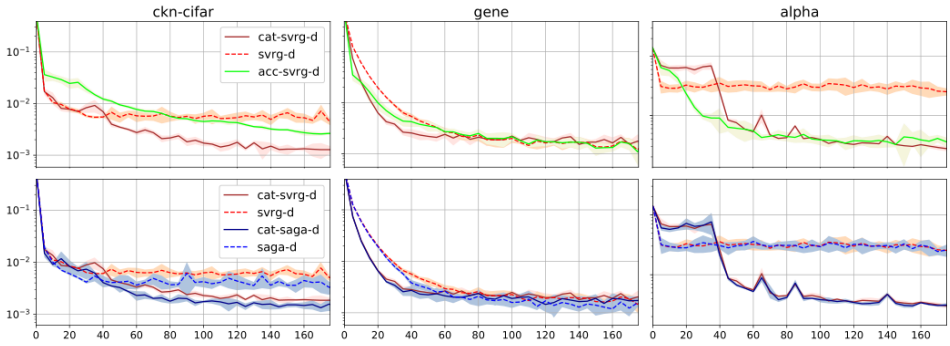


Figure 12. Accelerating SVRG-like (top) and SAGA (bottom) methods for ℓ_2-logistic regression with $\mu = 1/(100n)$ (bottom) for mild dropout, which imitates stochasticity in the gradients. All plots are on a logarithmic scale for the objective function value, and the x-axis denotes the number of epochs. The colored tubes around each curve denote a standard deviations across 5 runs. The curves show that acceleration may be useful even in the stochastic optimization regime.

generic strategies to make these algorithms robust to stochastic noise, which is useful when data is corrupted by small random perturbations. Finally, we show that this viewpoint is useful to obtain accelerated algorithms. A comparison with different approaches is shown in Figure 13 .

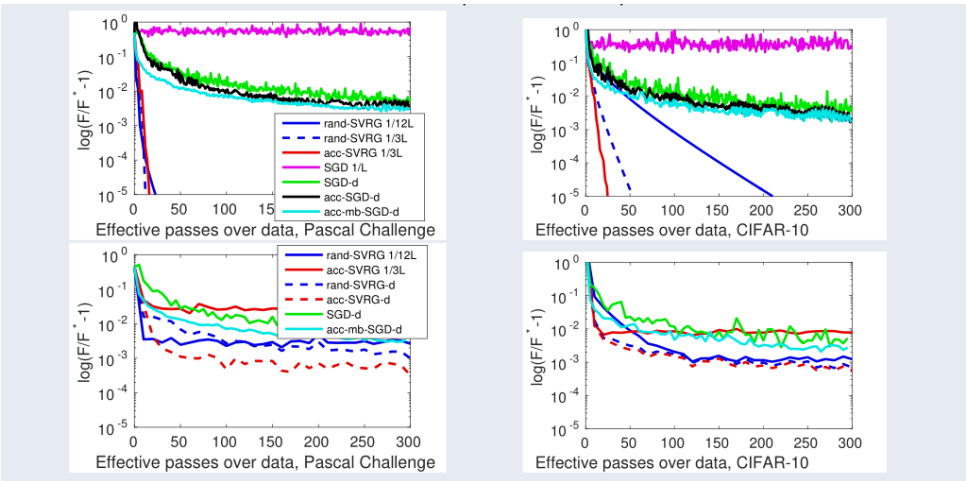


Figure 13. Comparison of different standard approaches with our developed method on two datasets for ℓ_2-logistic regression with mild dropout (bottom) and deterministic case (above). The case of exact gradient computations clearly shows benefits from acceleration, which consist in fast linear convergence. In the stochastic case, we demonstrate either superiority or high competitiveness of the developed method along with its unbiased convergence to the optimum. In both cases, we show that acceleration is able to generically comprise strengths of standard methods and even outperform them.

7.2.4. White-box vs Black-box: Bayes Optimal Strategies for Membership Inference

Participants: Alexandre Sablayrolles, Matthijs Douze, Yann Ollivier, Cordelia Schmid, Hervé Jégou.

Membership inference determines, given a sample and trained parameters of a machine learning model, whether the sample was part of the training set. In this paper [28], we derive the optimal strategy for membership inference with a few assumptions on the distribution of the parameters. We show that optimal attacks only depend on the loss function, and thus black-box attacks are as good as white-box attacks. As the optimal strategy is not tractable, we provide approximations of it leading to several inference methods 14 , and show that existing membership inference methods are coarser approximations of this optimal strategy. Our membership attacks outperform the state of the art in various settings, ranging from a simple logistic regression to more complex architectures and datasets, such as ResNet-101 and Imagenet.

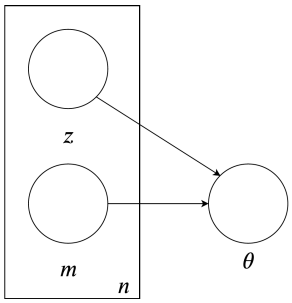


Figure 14. Plate notation of the membership inference problem: for each data point z_i, a binary membership variable m_i is sampled, and z_i belongs to the training set iff $m_i = 1$. Given the trained parameters θ and a sample z_i, we want to infer the value of m_i.

7.3. Theory and Methods for Deep Neural Networks

7.3.1. Group Invariance, Stability to Deformations, and Complexity of Deep Convolutional Representations

Participants: Alberto Bietti, Julien Mairal.

The success of deep convolutional architectures is often attributed in part to their ability to learn multiscale and invariant representations of natural signals. However, a precise study of these properties and how they affect learning guarantees is still missing. In the paper [3], we consider deep convolutional representations of signals; we study their invariance to translations and to more general groups of transformations, their stability to the action of diffeomorphisms, and their ability to preserve signal information. This analysis is carried by introducing a multilayer kernel based on convolutional kernel networks and by studying the geometry induced by the kernel mapping. We then characterize the corresponding reproducing kernel Hilbert space (RKHS), showing that it contains a large class of convolutional neural networks with homogeneous activation functions. This analysis allows us to separate data representation from learning, and to provide a canonical measure of model complexity, the RKHS norm, which controls both stability and generalization of any learned model. In addition to models in the constructed RKHS, our stability analysis also applies to convolutional networks with generic activations such as rectified linear units, and we discuss its relationship with recent generalization bounds based on spectral norms.

7.3.2. A Kernel Perspective for Regularizing Deep Neural Networks

Participants: Alberto Bietti, Grégoire Mialon, Dexiong Chen, Julien Mairal.

We propose a new point of view for regularizing deep neural networks by using the norm of a reproducing kernel Hilbert space (RKHS) [12]. Even though this norm cannot be computed, it admits upper and lower approximations leading to various practical strategies. Specifically, this perspective (i) provides a common umbrella for many existing regularization principles, including spectral norm and gradient penalties, or adversarial training, (ii) leads to new effective regularization penalties, and (iii) suggests hybrid strategies combining lower and upper bounds to get better approximations of the RKHS norm. We experimentally show this approach to be effective when learning on small datasets, or to obtain adversarially robust models.

7.3.3. On the Inductive Bias of Neural Tangent Kernels

Participants: Alberto Bietti, Julien Mairal.

State-of-the-art neural networks are heavily over-parameterized, making the optimization algorithm a crucial ingredient for learning predictive models with good generalization properties. A recent line of work has shown that in a certain over-parameterized regime, the learning dynamics of gradient descent are governed by a certain kernel obtained at initialization, called the neural tangent kernel. In [12], we study the inductive bias of learning in such a regime by analyzing this kernel and the corresponding function space (RKHS). In particular, we study smoothness, approximation, and stability properties of functions with finite norm, including stability to image deformations in the case of convolutional networks, and compare to other known kernels for similar architectures.

7.3.4. Large Memory Layers with Product Keys

Participants: Guillaume Lample, Alexandre Sablayrolles, Marc'Aurelio Ranzato, Ludovic Denoyer, Hervé Jégou.

This paper introduces a structured memory which can be easily integrated into a neural network. The memory is very large by design and significantly increases the capacity of the architecture, by up to a billion parameters with a negligible computational overhead. Its design and access pattern is based on product keys, which enable fast and exact nearest neighbor search. The ability to increase the number of parameters while keeping the same computational budget lets the overall system strike a better trade-off between prediction accuracy and computation efficiency both at training and test time. This memory layer, shown in Figure 15 , allows us to tackle very large scale language modeling tasks. In our experiments we consider a dataset with up to 30 billion words, and we plug our memory layer in a state-of-the-art transformer-based architecture. In particular, we found that a memory augmented model with only 12 layers outperforms a baseline transformer model with 24 layers, while being twice faster at inference time. We release our code for reproducibility purposes.

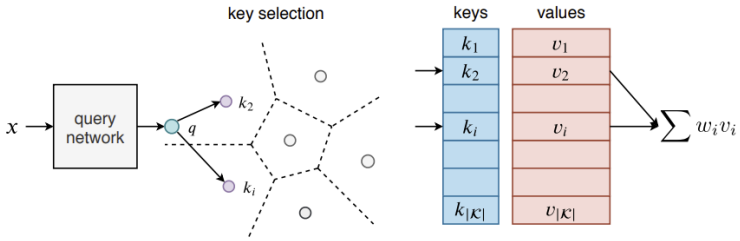


Figure 15. Overview of a key-value memory layer: The input x is processed through a query network that produces a query vector q, which is compared to all the keys. The output is the sparse weighted sum over the memories associated with the selected keys. For a large number of keys $|\mathcal{K}|$, the key selection procedure becomes too expensive in practice. Our product key method is exact and makes this search process very fast.

7.3.5. Understanding Priors in Bayesian Neural Networks at the Unit Level

Participants: Mariia Vladimirova, Jakob Verbeek, Pablo Mesejo [Univ. Granada, Spain], Julyan Arbel [Inria MISTIS].

In our ICML'19 paper [31], we investigate deep Bayesian neural networks with Gaussian weight priors and a class of ReLUlike nonlinearities. Bayesian neural networks with Gaussian priors are well known to induce an L2, "weight decay", regularization. Our results characterize a more intricate regularization effect at the level of the unit activations. Our main result establishes that the induced prior distribution on the units before and after activation becomes increasingly heavy-tailed with the depth of the layer. We show that first layer units are Gaussian, second layer units are sub-exponential, and units in deeper layers are characterized by sub-Weibull distributions. Our results provide new theoretical insight on deep Bayesian neural networks, which we corroborate with experimental simulation results.

7.3.6. Adaptative Inference Cost With Convolutional Neural Mixture Models

Participants: Adria Ruiz, Jakob Verbeek.

Despite the outstanding performance of convolutional neural networks (CNNs) for many vision tasks, the required computational cost during inference is problematic when resources are limited. In this paper [27], we propose Convolutional Neural Mixture Models (CNMMs), a probabilistic model embedding a large number of CNNs that can be jointly trained and evaluated in an efficient manner. Within the proposed framework, we present different mechanisms to prune subsets of CNNs from the mixture, allowing to easily adapt the computational cost required for inference (see Fig. 16). Image classification and semantic segmentation experiments show that our method achieve excellent accuracy-compute trade-offs. Moreover, unlike most of previous approaches, a single CNMM provides a large range of operating points along this trade-off, without any re-training.

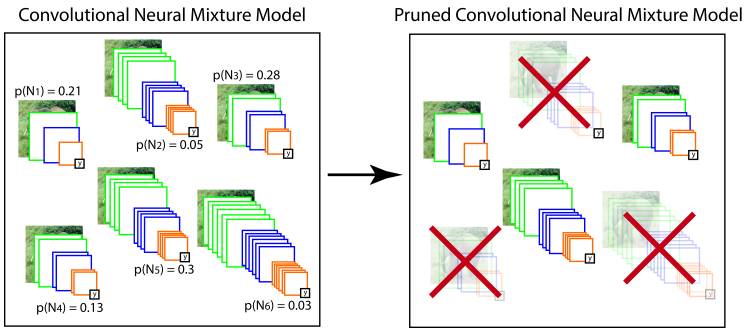


Figure 16. A Convolutional Neural Mixture Model embeds a large number of CNNs. Weight sharing enables efficient joint training of all networks and computation of the mixture output. The learned mixing weights can be used to remove networks from the mixture, and thus reduce the computational cost of inference.

7.4. Pluri-disciplinary Research

7.4.1. Biological Sequence Modeling with Convolutional Kernel Networks

Participants: Dexiong Chen, Laurent Jacob, Julien Mairal.

The growing number of annotated biological sequences available makes it possible to learn genotype-phenotype relationships from data with increasingly high accuracy. When large quan- tities of labeled samples are available for training a model, convolutional neural networks can be used to predict the phenotype of unannotated sequences with good accuracy. Unfortunately, their performance with medium- or small-scale datasets is mitigated, which requires inventing new data-efficient approaches. In this paper [4], [14], we introduce a hybrid approach between convolutional neural networks and kernel methods to model biological sequences. Our method, shown in Figure 17 , enjoys the ability of convolutional neural networks to learn data representations that are adapted to a specific task, while the kernel point of view yields algorithms that perform significantly better when the amount of training data is small. We illustrate these advantages for transcription factor binding prediction and protein homology detection, and we demonstrate that our model is also simple to interpret, which is crucial for discovering predictive motifs in sequences. The source code is freely available at https://gitlab.inria.fr/dchen/CKN-seq.

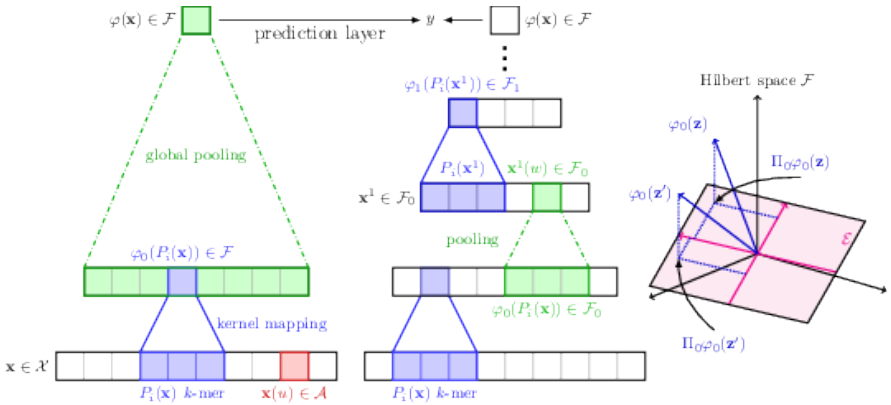


Figure 17. Construction of single-layer (left) and multilayer (middle) CKN-seq and the approximation of one layer (right). For a single-layer model, each k-mer $P_i(\mathbf{x})$ is mapped to $\varphi_0(P_i(\mathbf{x}))$ in $\mathcal{F}$ and projected to $\Pi\varphi_0(P_i(\mathbf{x}))$ parametrized by $\psi_0(P_i(\mathbf{x}))$. Then, the final finite-dimensional sequence is obtained by the global pooling, $\psi(\mathbf{x}) = \frac{1}{m}\sum_{i=0}^{m}\psi_0(P_i(\mathbf{x}))$. The multilayer construction is similar, but relies on intermediate maps, obtained by local pooling.

7.4.2. Recurrent Kernel Networks

Participants: Dexiong Chen, Laurent Jacob [CNRS, LBBE Laboratory], Julien Mairal.

Substring kernels are classical tools for representing biological sequences or text. However, when large amounts of annotated data are available, models that allow end-to-end training such as neural networks are often preferred. Links between recurrent neural networks (RNNs) and substring kernels have recently been drawn, by formally showing that RNNs with specific activation functions were points in a reproducing kernel Hilbert space (RKHS). In this paper [15], we revisit this link by generalizing convolutional kernel networks——originally related to a relaxation of the mismatch kernel——to model gaps in sequences. It results in a new type of recurrent neural network (Figure 18), which can be trained end-to-end with backpropagation, or without supervision by using kernel approximation techniques. We experimentally show that our approach is well suited to biological sequences, where it outperforms existing methods for protein classification tasks.

7.4.3. Depth-adaptive Transformer

Participants: Maha Elbayad, Jiatao Gu [Facebook AI], Edouard Grave [Facebook AI], Michael Auli [Facebook AI].

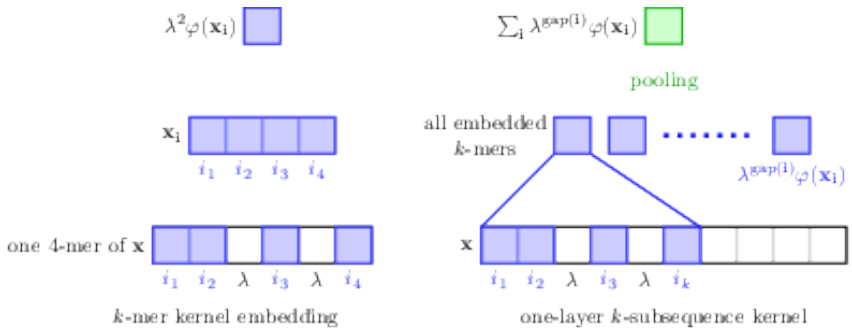


Figure 18. Representation of a sequence in a RKHS based on our kernel.

State of the art sequence-to-sequence models for large scale tasks perform a fixed number of computations for each input sequence regardless of whether it is easy or hard to process. In our ICLR'2020 paper [18], we train Transformer models which can make output predictions at different stages of the network and we investigate different ways to predict how much computation is required for a particular sequence. Unlike dynamic computation in Universal Transformers, which applies the same set of layers iteratively, we apply different layers at every step to adjust both the amount of computation as well as the model capacity. On IWSLT German-English translation our approach matches the accuracy of a well tuned baseline Transformer while using less than a quarter of the decoder layers. Figure 19 illustrates the different halting mechanisms investigated in this work. Namely, a sequence-level approach where we assume all the sequence's tokens are equally difficult and a token-level approach where tokens exit at varying depths.

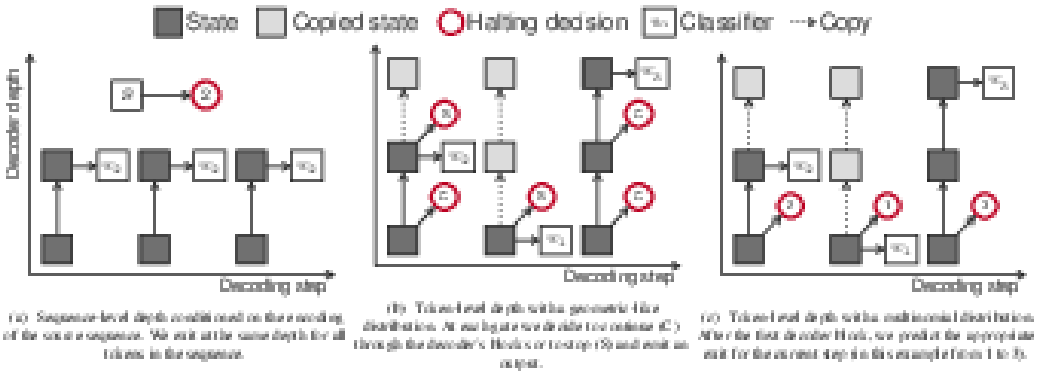


Figure 19. Illustration of the variant adaptive depth predictors: (a) the sequence-level and (b, c) at the token-level.

TYREX Project-Team

6. New Results

6.1. On the Optimization of Recursive Relational Queries: Application to Graph Queries

Graph databases have received a lot of attention as they are particularly useful in many applications such as social networks, life sciences and the semantic web. Various languages have emerged to query graph databases, many of which embed forms of recursion which reveal essential for navigating in graphs. The relational model has benefited from a huge body of research in the last half century and that is why many graph databases rely on techniques of relational query engines. Since its introduction, the relational model has seen various attempts to extend it with recursion and it is now possible to use recursion in several SQL or Datalog based database systems. The optimization of recursive queries remains, however, a challenge. We propose μ-RA, a variation of the Relational Algebra equipped with a fixpoint operator for expressing recursive relational queries. μ-RA can notably express unions of conjunctive regular path queries. Leveraging the fact that this fixpoint operator makes recursive terms more amenable to algebraic transformations, we propose new rewrite rules. These rules make it possible to generate new query execution plans, that cannot be obtained with previous approaches. We have defined the syntax and semantics of μ-RA, together with the rewriting rules that we specifically devised to tackle the optimization of recursive queries. We have also conducted practical experiments that show that the newly generated plans can provide significant performance improvements for evaluating recursive queries over graphs.

These results will be presented at the SIGMOD 2020 conference [9].

6.2. An Algebra with a Fixpoint Operator for Distributed Data Collections

We propose an algebra with a fixpoint operator which is suitable for modeling recursive computations with distributed data collections. We show that under reasonable conditions this fixpoint can be evaluated by parallel loops with one final merge rather than by a global loop requiring network overhead after each iteration. We also propose rewrite rules, showing when and how filters can be pushed through recursive terms, and how to filter inside a fixpoint before a join. Experiments with the Spark platform illustrate performance gains brought by these systematic optimizations [10].

6.3. Backward Type Inference for XML Queries

Although XQuery is a statically typed, functional query language for XML data, some of its features such as upward and horizontal XPath axes are typed imprecisely. The main reason is that while the XQuery data model allows to navigate upwards and between siblings from a given XML node, the type model, e.g., regular tree types, can only describe the subtree structure of the given node. In 2015, Giuseppe Castagna and our team independently proposed a precise forward type inference system for XQuery using an extended type language that can describe not only a given XML node but also its context. Recently, as a complementary method to such forward type inference systems, we propose an enhanced backward type inference system for XQuery, based on an extended type language. Results include an exact type system for XPath axes and a sound type system for XQuery expressions.

6.4. Scalable and Interpretable Predictive Models for Electronic Health Records

Early identification of patients at risk of developing complications during their hospital stay is currently one of the most challenging issues in healthcare. Complications include hospital-acquired infections, admissions to intensive care units, and in-hospital mortality. Being able to accurately predict the patients' outcomes is a crucial prerequisite for tailoring the care that certain patients receive, if it is believed that they will do poorly without additional intervention. We consider the problem of complication risk prediction, such as patient mortality, from the electronic health records of the patients. We study the question of making predictions on the first day at the hospital, and of making updated mortality predictions day after day during the patient's stay. We are developping distributed models that are scalable and interpretable. Key insights include analyzing diagnoses known at admission and drugs served, which evolve during the hospital stay. We leverage a distributed architecture to learn interpretable models from training datasets of gigantic size. We test our analyses with more than one million of patients from hundreds of hospitals, and report on the lessons learned from these experiments.

Preliminary results were presented at the 2018 International Conference on Data Science and Applications, and extended results have been submitted for publication consideration.

6.5. What can millions of laboratory test results tell us about the temporal aspect of data quality? Study of data spanning 17 years in a clinical data warehouse.

In this work, our objective is to identify common temporal evolution profiles in biological data and to propose a semi-automated method to these patterns in a clinical data warehouse (CDW). We leveraged the CDW of the European Hospital Georges Pompidou and tracked the evolution of 192 biological parameters over a period of 17 years (for 445,0 0 0 + patients, and 131 million laboratory test results). We have identified three common profiles of evolution: discretization, breakpoints, and trends. We developed computational and statistical methods to identify these profiles in the CDW. Overall, of the 192 observed biological parameters (87,814,136 values), 135 presented at least one evolution. We identified breakpoints in 30 distinct parameters, discretizations in 32, and trends in 79. As a conclusion, we can say that our method allows the identification of several temporal events in the data. Considering the distribution over time of these events, we identified probable causes for the observed profiles: instruments or software upgrades and changes in computation formulas. We evaluated the potential impact for data reuse. Finally, we formulated recommendations to enable safe use and sharing of biological data collection to limit the impact of data evolution in retrospective and federated studies (e.g. the annotation of laboratory parameters presenting breakpoints or trends) [4].

6.6. Interactive Mapping Specification with Exemplar Tuples

While schema mapping specification is a cumbersome task for data curation specialists, it becomes unfeasible for non-expert users, who are unacquainted with the semantics and languages of the involved transformations.

In this work, we propose an interactive framework for schema mapping specification suited for non-expert users. The underlying key intuition is to leverage a few exemplar tuples to infer the underlying mappings and iterate the inference process via simple user interactions under the form of Boolean queries on the validity of the initial exemplar tuples. The approaches available so far are mainly assuming pairs of complete universal data examples, which can be solely provided by data curation experts, or are limited to poorly expressive mappings.

We present a quasi-lattice-based exploration of the space of all possible mappings that satisfy arbitrary user exemplar tuples. Along the exploration, we challenge the user to retain the mappings that fit the user's requirements at best and to dynamically prune the exploration space, thus reducing the number of user interactions. We prove that after the refinement process, the obtained mappings are correct and complete. We present an extensive experimental analysis devoted to measure the feasibility of our interactive mapping strategies and the inherent quality of the obtained mappings [2].

6.7. Schema Validation and Evolution for Graph Databases

Despite the maturity of commercial graph databases, little consensus has been reached so far on the standardization of data definition languages (DDLs) for property graphs (PG). Discussion on the characteristics of PG schemas is ongoing in many standardization and community groups. Although some basic aspects of a schema are already present in most commercial graph databases, full support is missing allowing to constraint property graphs with more or less flexibility. In this work, we show how schema validation can be enforced through homomorphisms between PG schemas and PG instances by leveraging a concise schema DDL inspired by Cypher syntax. We also briefly discuss PG schema evolution that relies on graph rewriting operations allowing to consider both prescriptive and descriptive schemas [6].

6.8. MapRepair: Mapping and Repairing under Policy Views

Mapping design is overwhelming for end users, who have to check at par the correctness of the mappings and the possible information disclosure over the exported source instance. In our tool MapRepair, we focus on the latter problem by proposing a novel practical solution to ensure that a mapping faithfully complies with a set of privacy restrictions specified as source policy views. We showcase MapRepair, that guides the user through the tasks of visualizing the results of the data exchange process with and without the privacy restrictions. MapRepair leverages formal privacy guarantees and is inherently data-independent, i.e. if a set of criteria are satisfied by the mapping statement, then it guarantees that both the mapping and the underlying instances do not leak sensitive information. Furthermore, MapRepair also allows to automatically repair an input mapping w.r.t. a set of policy views in case of information leakage. We build on various demonstration scenarios, including synthetic and real-world instances and mappings [5].

6.9. Approximate Querying on Property Graphs

Property graphs are becoming widespread when modeling data with complex structural characteristics and enhancing edges and nodes with a list of properties. We worked on the approximate evaluation of counting queries involving recursive paths on property graphs. As such queries are already difficult to evaluate over pure RDF graphs, they require an ad-hoc graph summary for their approximate evaluation on property graphs. We prove the intractability of the optimal graph summarization problem, under our algorithm's conditions. We design and implement a novel property graph summary suitable for the above queries, along with an approximate query evaluation module. Finally, we show the compactness of the obtained summaries as well as the accuracy of answering counting recursive queries on them [8].

6.10. RDF Graph Anonymization Robust to Data Linkage

Privacy is a major concern when publishing new datasets in the context of Linked Open Data (LOD). A new dataset published in the LOD is indeed exposed to privacy breaches due to the linkage to objects already present in the other datasets of the LOD. In this work, we focus on the problem of building safe anonymizations of an RDF graph to guarantee that linking the anonymized graph with any external RDF graph will not cause privacy breaches. Given a set of privacy queries as input, we study the data-independent safety problem and the sequence of anonymization operations necessary to enforce it. We provide sufficient conditions under which an anonymization instance is safe given a set of privacy queries. Additionally, we show that our algorithms for RDF data anonymization are robust in the presence of sameAs links that can be explicit or inferred by additional knowledge.

6.11. Navigating the Maze of Wikidata Query Logs

We propose an in-depth and diversified analysis of the Wikidata query logs, recently made publicly available. Although the usage of Wikidata queries has been the object of recent studies, our analysis of the query traffic reveals interesting and unforeseen findings concerning the usage, types of recursion, and the shape classification of complex recursive queries. Wikidata specific features combined with recursion let us identify a significant subset of the entire corpus that can be used by the community for further assessment. We

consider and analyze the queries across many different dimensions, such as the robotic and organic queries, the presence/absence of constants along with the correctly executed and timed out queries. A further investigation that we pursue is to find, given a query, a number of queries structurally similar to the given query. We provide a thorough characterization of the queries in terms of their expressive power, their topological structure and shape, along with a deeper understanding of the usage of recursion in these logs. We make the code for the analysis available as open source [7].

6.12. Graph Generators: State of the Art and Open Challenges

The abundance of interconnected data has fueled the design and implementation of graph generators reproducing real-world linking properties, or gauging the effectiveness of graph algorithms, techniques and applications manipulating these data. We consider graph generation across multiple subfields, such as Semantic Web, graph databases, social networks, and community detection, along with general graphs. Despite the disparate requirements of modern graph generators throughout these communities, we analyze them under a common umbrella, reaching out the functionalities, the practical usage, and their supported operations. We argue that this classification is serving the need of providing scientists, researchers and practitioners with the right data generator at hand for their work. This survey provides a comprehensive overview of the state-of-the-art graph generators by focusing on those that are pertinent and suitable for several data-intensive tasks. Finally, we discuss open challenges and missing requirements of current graph generators along with their future extensions to new emerging fields [3].

6.13. A trichotomy for regular simple path queries on graphs

We focus on the computational complexity of regular simple path queries (RSPQs). We consider the following problem RSPQ(L) for a regular language L: given an edge-labeled digraph Gand two nodes xand y, is there a simple path from x to y that forms a word belonging to L? We fully characterize the frontier between tractability and intractability for RSPQ(L). More precisely, we prove RSPQ(L)is either AC0, NL-complete or NP-complete depending on the language L. We also provide a simple characterization of the tractable fragment in terms of regular expressions. Finally, we also discuss the complexity of deciding whether a language L belongs to the fragment above. We consider several alternative representations of L: DFAs, NFAs or regular expressions, and prove that this problem is NL-complete for the first representation and PSpace-complete for the other two [1].