# Activity Report 2015

# Section New Results

<span style="color:red">ARIC Project-Team</span>

# 7. New Results

## 7.1. Floating-point Arithmetic

### 7.1.1. *On the maximum relative error when computing integer powers by iterated multiplications in floating-point arithmetic*

We improve the usual relative error bound for the computation of $x^n$ through iterated multiplications by $x$ in binary floating-point arithmetic. The obtained error bound is only slightly better than the usual one, but it is simpler. We also discuss the more general problem of computing the product of $n$ terms. [5]

### 7.1.2. *Formally verified certificate checkers for hardest-to-round computation*

In order to derive efficient and robust floating-point implementations of a given function $f$, it is crucial to compute its hardest-to-round points, i.e. the floating-point numbers $x$ such that $f(x)$ is closest to the midpoint of two consecutive floating-point numbers. Depending on the floating-point format one is aiming at, this can be highly computationally intensive. In this paper, we show how certificates based on Hensel's lemma can be added to an algorithm using lattice basis reduction so that the result of a computation can be formally checked in the Coq proof assistant. [7]

### 7.1.3. *On the error of Computing $ab + cd$ using Cornea, Harrison and Tang's method*

In their book, Scientific Computing on the Itanium, Cornea et al. (2002) introduce an accurate algorithm for evaluating expressions of the form $ab + cd$ in binary floating-point arithmetic, assuming an FMA instruction is available. They show that if $p$ is the precision of the floating-point format and if $u = 2^{-p}$, the relative error of the result is of order $u$. We improve their proof to show that the relative error is bounded by $2u + 7u^2 + 6u^3$. Furthermore, by building an example for which the relative error is asymptotically (as $p \to \infty$ or, equivalently, as $u \to 0$) equivalent to $2u$, we show that our error bound is asymptotically optimal. [8]

### 7.1.4. *Improved error bounds for floating-point products and Horner's scheme*

Let $u$ denote the relative rounding error of some floating-point format. Recently it has been shown that for a number of standard Wilkinson-type bounds the typical factors $\gamma_k := ku/(1-ku)$ can be improved into $ku$, and that the bounds are valid without restriction on $k$. Problems include summation, dot products and thus matrix multiplication, residual bounds for LU- and Cholesky-decomposition, and triangular system solving by substitution. In this note we show a similar result for the product $\prod_{i=0}^{k} x_i$ of real and/or floating-point numbers $x_i$, for computation in any order, and for any base $\beta \geq 2$. The derived error bounds are valid under a mandatory restriction of $k$. Moreover, we prove a similar bound for Horner's polynomial evaluation scheme. [9]

### 7.1.5. *Comparison between binary and decimal floating-point numbers*

In collaboration with Christoph Lauter and Marc Mezzarobba (LIP6 laboratory, Paris), Nicolas Brisebarre and Jean-Michel Muller introduce an algorithm to compare a binary floating-point (FP) number and a decimal FP number, assuming the "binary encoding" of the decimal formats is used, and with a special emphasis on the basic interchange formats specified by the IEEE 754-2008 standard for FP arithmetic. It is a two-step algorithm: a first pass, based on the exponents only, quickly eliminates most cases, then, when the first pass does not suffice, a more accurate second pass is performed. They provide an implementation of several variants of our algorithm, and compare them [26].

# 7.2. Lattices: algorithms and cryptology

### 7.2.1. *Linearly Homomorphic Encryption from DDH*

We design a linearly homomorphic encryption scheme whose security relies on the hardness of the decisional Diffie-Hellman problem. Our approach requires some special features of the underlying group. In particular, its order is unknown and it contains a subgroup in which the discrete logarithm problem is tractable. Therefore, our instantiation holds in the class group of a non maximal order of an imaginary quadratic field. Its algebraic structure makes it possible to obtain such a linearly homomorphic scheme whose message space is the whole set of integers modulo a prime $p$ and which supports an unbounded number of additions modulo $p$ from the ciphertexts. A notable difference with previous works is that, for the first time, the security does not depend on the hardness of the factorization of integers. As a consequence, under some conditions, the prime $p$ can be scaled to fit the application needs. [13]

### 7.2.2. *Secure Efficient History-Hiding Append-Only Signatures in the Standard Model*

As formalized by Kiltz et al. (ICALP'05), append-only signatures (AOS) are digital signature schemes where anyone can publicly append extra message blocks to an already signed sequence of messages. This property is useful, e.g., in secure routing, in collecting response lists, reputation lists, or petitions. Bethencourt, Boneh and Waters (NDSS'07) suggested an interesting variant, called history-hiding append-only signatures (HH-AOS), which handles messages as sets rather than ordered tuples. This HH-AOS primitive is useful when the exact order of signing needs to be hidden. When free of subliminal channels (i.e., channels that can tag elements in an undetectable fashion), it also finds applications in the storage of ballots on an electronic voting terminals or in other archival applications (such as the record of petitions, where we want to hide the influence among messages). However, the only subliminal-free HH-AOS to date only provides heuristic arguments in terms of security: Only a proof in the idealized (non-realizable) random oracle model is given. This paper provides the first HH-AOS construction secure in the standard model. Like the system of Bethencourt et al., our HH-AOS features constant-size public keys, no matter how long messages to be signed are, which is atypical (we note that secure constructions often suffer from a space penalty when compared to their random-oracle-based counterpart). As a second result, we show that, even if we use it to sign ordered vectors as in an ordinary AOS (which is always possible with HH-AOS), our system provides considerable advantages over existing realizations. As a third result, we show that HH-AOS schemes provide improved identity-based ring signatures (i.e., in prime order groups and with a better efficiency than the state-of-the-art schemes). [17]

### 7.2.3. *Compactly Hiding Linear Spans: Tightly Secure Constant-Size Simulation-Sound QA-NIZK Proofs and Applications*

Quasi-adaptive non-interactive zero-knowledge (QA-NIZK) proofs is a powerful paradigm, suggested recently by Jutla and Roy (Asiacrypt'13), which is motivated by the Groth-Sahai seminal techniques for efficient non-interactive zero-knowledge (NIZK) proofs. In this paradigm, the common reference string may depend on specific language parameters, a fact that allows much shorter proofs in important cases. It even makes certain standard model applications competitive with the Fiat-Shamir heuristic in the Random Oracle idealization (such QA-NIZK proofs were recently optimized to constant size by Jutla and Roy (Crypto'14) and Libert et al. (Eurocrypt'14) for the important case of proving that a vector of group elements belongs to a linear subspace). While, e.g., the QA-NIZK arguments of Libert et al. provide unbounded simulation-soundness and constant proof length, their simulation-soundness is only loosely related to the underlying assumption (with a gap proportional to the number of adversarial queries) and it is unknown how to alleviate this limitation without sacrificing efficiency. Here, we deal with the basic question of whether and to what extent we can simultaneously optimize the proof size and the tightness of security reductions, allowing for important applications with tight security (which are typically to date quite lengthy) to be of shorter size. In this paper, we resolve this question by describing a novel simulation-sound QA-NIZK argument showing that a vector v ∈ G n belongs to a subspace of rank t < n using a constant number of group elements. Unlike previous constant-size QA-NIZK proofs of such statements, the unbounded simulation-soundness of our system is nearly tightly related (i.e., the reduction only loses a factor proportional to the security parameter) to the standard Decision

Linear assumption. To show simulation-soundness in the constrained context of tight reductions, we employ a number of techniques, and explicitly point at a technique – which may be of independent interest – of hiding the linear span of a structure-preserving homomorphic signature (which is part of an OR proof). As an application, we design a public-key cryptosystem with almost tight CCA2-security in the multi-challenge, multiuser setting with improved length (asymptotically optimal for long messages). We also adapt our scheme to provide CCA security in the key-dependent message scenario (KDM-CCA2) with ciphertext length reduced by 75% when compared to the best known tightly secure KDM-CCA2 system so far. [18]

### 7.2.4. Short Group Signatures via Structure-Preserving Signatures: Standard Model Security from Simple Assumptions

Group signatures are a central cryptographic primitive which allows users to sign messages while hiding their identity within a crowd of group members. In the standard model (without the random oracle idealization), the most efficient constructions rely on the Groth-Sahai proof systems (Eurocrypt'08). The structure-preserving signatures of Abe et al. (Asiacrypt'12) make it possible to design group signatures based on well-established, constant-size number theoretic assumptions (a.k.a. "simple assumptions") like the Symmetric eXternal Diffie-Hellman or Decision Linear assumptions. While much more efficient than group signatures built on general assumptions, these constructions incur a significant overhead w.r.t. constructions secure in the idealized random oracle model. Indeed, the best known solution based on simple assumptions requires 2.8 kB per signature for currently recommended parameters. Reducing this size and presenting techniques for shorter signatures are thus natural questions. In this paper, our first contribution is to significantly reduce this overhead. Namely, we obtain the first fully anonymous group signatures based on simple assumptions with signatures shorter than 2 kB at the 128-bit security level. In dynamic (resp. static) groups, our signature length drops to 1.8 kB (resp. 1 kB). This improvement is enabled by two technical tools. As a result of independent interest, we first construct a new structure-preserving signature based on simple assumptions which shortens the best previous scheme by 25%. Our second tool is a new method for attaining anonymity in the strongest sense using a new CCA2-secure encryption scheme which is simultaneously a Groth-Sahai commitment. [19]

### 7.2.5. Implementing Candidate Graded Encoding Schemes from Ideal Lattices

Multilinear maps have become popular tools for designing cryptographic schemes since a first approximate realisation candidate was proposed by Garg, Gentry and Halevi (GGH). This construction was later improved by Langlois, Stehlé and Steinfeld who proposed GGHLite which offers smaller parameter sizes. In this work, we provide the first implementation of such approximate multilinear maps based on ideal lattices. Implementing GGH-like schemes naively would not allow instantiating it for non-trivial parameter sizes. We hence propose a strategy which reduces parameter sizes further and several technical improvements to allow for an efficient implementation. In particular, since finding a prime ideal when generating instances is an expensive operation, we show how we can drop this requirement. We also propose algorithms and implementations for sampling from discrete Gaussians, for inverting in some Cyclotomic number fields and for computing norms of ideals in some Cyclotomic number rings. Due to our improvements we were able to compute a multilinear jigsaw puzzle for $\kappa = 52$ (resp. $\kappa = 38$) and $\lambda = 52$ (resp. $\lambda = 80$). [10]

### 7.2.6. Improved security proofs in lattice-based cryptography: using the Rényi divergence rather than the statistical distance

The Rényi divergence is a mean to measure the closeness of two distributions. We show that it can often be used as an alternative to the statistical distance in security proofs for lattice-based cryptography. Using the Rényi divergence is particularly suited for security proofs of primitives in which the attacker is required to solve a search problem (e.g., forging a signature). We show that it may also be used in the case of distinguishing problems (e.g., semantic security of encryption schemes), when they enjoy a public sampleability property. The techniques lead to security proofs for schemes with smaller parameters. [11]

### 7.2.7. Fully Secure Functional Encryption for Inner Products, from Standard Assumptions

Functional encryption is a modern public-key paradigm where a master secret key can be used to derive sub-keys SKF associated with certain functions $F$ in such a way that the decryption operation reveals $F(M)$, if

$M$ is the encrypted message, and nothing else. Recently, Abdalla et al. gave simple and effient realizations of the primitive for the computation of linear functions on encrypted data: given an encryption of a vector y over some specific base ring, a secret key $SK_x$ for the vector $x$ allows computing $< x, y >$. Their technique surprisingly allows for instantiations under standard assumptions, like the hardness of the Decision Diffie-Hellman (DDH) and Learning-with-Errors (LWE) problems. Their constructions, however, are only proved secure against selective adversaries, which have to declare the challenge messages $M_0$ and $M_1$ at the outset of the game. In this paper, we provide constructions that provably achieve security against more realistic adaptive attacks (where the messages $M_0$ and $M_1$ may be chosen in the challenge phase, based on the previously collected information) for the same inner product functionality. Our constructions are obtained from hash proof systems endowed with homomorphic properties over the key space. They are (almost) as efficient as those of Abdalla et al. and rely on the same hardness assumptions. In addition, we obtain a solution based on Paillier's composite residuosity assumption, which was an open problem even in the case of selective adversaries. We also propose LWE-based schemes that allow evaluation of inner products modulo a prime $p$, as opposed to the schemes of Abdalla et al. that are restricted to evaluations of integer inner products of short integer vectors. We finally propose a solution based on Paillier's composite residuosity assumption that enables evaluation of inner products modulo an RSA integer $N = pq$. We demonstrate that the functionality of inner products over a prime field is very powerful and can be used to construct bounded collusion FE for all circuits. [23]

### 7.2.8. *Fully Homomophic Encryption over the Integers Revisited*

Two main computational problems serve as security foundations of current fully homomorphic encryption schemes: Regev's Learning With Errors problem (LWE) and Howgrave-Graham's Approximate Greatest Common Divisor problem (AGCD). Our first contribution is a reduction from LWE to AGCD. As a second contribution, we describe a new AGCD-based fully homomorphic encryption scheme, which outperforms all prior AGCD-based proposals: its security does not rely on the presumed hardness of the so-called Sparse Subset Sum problem, and the bit-length of a ciphertext is only $\widetilde{O}\lambda$, where $\lambda$ refers to the security parameter. [15]

### 7.2.9. *Cryptanalysis of the Multilinear Map over the Integers*

We describe a polynomial-time cryptanalysis of the (approximate) multilinear map of Coron, Lepoint and Tibouchi (CLT). The attack relies on an adaptation of the so-called zeroizing attack against the Garg, Gentry and Halevi (GGH) candidate multilinear map. Zeroizing is much more devastating for CLT than for GGH. In the case of GGH, it allows to break generalizations of the Decision Linear and Subgroup Membership problems from pairing-based cryptography. For CLT, this leads to a total break: all quantities meant to be kept secret can be efficiently and publicly recovered. [14]

### 7.2.10. *Cryptanalysis of Gu's ideal multilinear map*

In March, 2015 Gu Chunsheng proposed a candidate ideal multilinear map [eprint 2015/269]. An ideal multilinear map allows to perform as many multiplications as desired, while in $k$-multilinear maps like GGH [EC 2013] or CLT [CR2013, CR2015] one we canperform at most a predetermined number $k$ of multiplications. In this note, we show that the extraction Multilinear Computational Diffie-Hellman problem (ext-MCDH) associated to Gu's map can be solved in polynomial-time: this candidate ideal multilinear map is insecure. We also give intuition on why we think that the two other ideal multilinear maps proposed by Gu in [eprint 2015/269] are not secure either. [39]

### 7.2.11. *Worst-case to average-case reductions for module lattices*

Most lattice-based cryptographic schemes are built upon the assumed hardness of the Short Integer Solution (SIS) and Learning With Errors (LWE) problems. Their efficiencies can be drastically improved by switching the hardness assumptions to the more compact Ring-SIS and Ring-LWE problems. However, this change of hardness assumptions comes along with a possible security weakening: SIS and LWE are known to be at least as hard as standard (worst-case) problems on euclidean lattices, whereas Ring-SIS and Ring-LWE are only known to be as hard as their restrictions to special classes of ideal lattices, corresponding to ideals of some

polynomial rings. In this work, we define the Module-SIS and Module-LWE problems, which bridge SIS with Ring-SIS, and LWE with Ring-LWE, respectively. We prove that these average-case problems are at least as hard as standard lattice problems restricted to module lattices (which themselves generalize arbitrary and ideal lattices). As these new problems enlarge the toolbox of the lattice-based cryptographer, they could prove useful for designing new schemes. Importantly, the worst-case to average-case reductions for the module problems are (qualitatively) sharp, in the sense that there exist converse reductions. This property is not known to hold in the context of Ring-SIS/Ring-LWE: Ideal lattice problems could reveal easy without impacting the hardness of Ring-SIS/Ring-LWE. [6]

### 7.2.12. Reducing Communication Overhead of the Subset Difference Scheme

In Broadcast Encryption (BE) systems like Pay-TV, AACS, online content sharing and broadcasting, reducing the header length (communication overhead per session) is of practical interest. The Subset Difference (SD) scheme due to Naor-Naor-Lotspiech (NNL) is the most popularly used BE scheme. This work introduced the $(a, b, \gamma)$ augmented binary tree subset difference ($(a, b, \gamma)$-ABTSD) scheme which is a generalization of the NNL-SD scheme. By varying the parameters $(a, b, \gamma)$, it is possible to obtain $O(n \log n)$ different schemes. In addition to the underlying binary tree structure of the NNL-SD scheme, the new scheme uses an additional binary tree structure of height $a$ augmented with each internal node. The SD subsets in this scheme arise due to nodes that are at a distance at most $b$ from each other. In the augmented tree of height $a$, at most $c$ leaves are considered together in creating the SD subsets for the scheme. The average header length achieved by the new schemes is smaller than all known schemes having the same decryption time as that of the NNL-SD scheme and achieving non-trivial trade-offs between the user storage and the header size. The amount of key material that a user is required to store increases. For the earlier mentioned applications, reducing header size and achieving fast decryption is perhaps more of a concern than the user storage

## 7.3. Algebraic computing and high performance kernels

### 7.3.1. Complexity of the F5 Gröbner basis algorithm

We study the complexity of Gröbner bases computation, in particular in the generic situation where the variables are in simultaneous Noether position with respect to the system. We give a bound on the number of polynomials of degree $d$ in a Gröbner basis computed by Faugère's F5 algorithm (2002) in this generic case for the grevlex ordering (which is also a bound on the number of polynomials for a reduced Gröbner basis, independently of the algorithm used). Next, we analyse more precisely the structure of the polynomials in the Gröbner bases with signatures that F5 computes and use it to bound the complexity of the algorithm. Our estimates show that the version of F5 we analyse, which uses only standard Gaussian elimination techniques, outperforms row reduction of the Macaulay matrix with the best known algorithms for moderate degrees, and even for degrees up to the thousands if Strassen's multiplication is used. The degree being fixed, the factor of improvement grows exponentially with the number of variables. [1]

### 7.3.2. Faster Algorithms for Multivariate Interpolation with Multiplicities and Simultaneous Polynomial Approximations

The interpolation step in the Guruswami-Sudan algorithm is a bivariate interpolation problem with multiplicities commonly solved in the literature using either structured linear algebra or basis reduction of polynomial lattices. This problem has been extended to three or more variables; for this generalization, all fast algorithms proposed so far rely on the lattice approach. In this work, we reduce this multivariate interpolation problem to a problem of simultaneous polynomial approximations, which we solve using fast structured linear algebra. This improves the best known complexity bounds for the interpolation step of the list-decoding of Reed-Solomon codes, Parvaresh-Vardy codes, and folded Reed-Solomon codes. In particular, for Reed-Solomon list-decoding with re-encoding, our approach has complexity $\widetilde{O}(\ell^{\omega-1} m^2 (n-k))$, where $\ell, m, n, k$ are the list size, the multiplicity, the number of sample points and the dimension of the code, and $\omega$ is the exponent of linear algebra; this accelerates the previously fastest known algorithm by a factor of $\ell/m$. [3]

### 7.3.3. Recursion based parallelization of exact dense linear algebra routines for Gaussian elimination

We present block algorithms and their implementation for the parallelization of sub-cubic Gaussian elimination on shared memory architectures. Contrarily to the classical cubic algorithms in parallel numerical linear algebra, we focus here on recursive algorithms and coarse grain parallelization. Indeed, sub-cubic matrix arithmetic can only be achieved through recursive algorithms making coarse grain block algorithms perform more efficiently than fine grain ones. This work is motivated by the design and implementation of dense linear algebra over a finite field, where fast matrix multiplication is used extensively and where costly modular reductions also advocate for coarse grain block decomposition. We incrementally build efficient kernels, for matrix multiplication first, then triangular system solving, on top of which a recursive PLUQ decomposition algorithm is built. We study the parallelization of these kernels using several algorithmic variants: either iterative or recursive and using different splitting strategies. Experiments show that recursive adaptive methods for matrix multiplication, hybrid recursive-iterative methods for triangular system solve and tile recursive versions of the PLUQ decomposition, together with various data mapping policies, provide the best performance on a 32 cores NUMA architecture. Overall, we show that the overhead of modular reductions is more than compensated by the fast linear algebra algorithms and that exact dense linear algebra matches the performance of full rank reference numerical software even in the presence of rank deficiencies. [4]

### 7.3.4. Computing the Rank Profile Matrix

The row (resp. column) rank profile of a matrix describes the staircase shape of its row (resp. column) echelon form. In an ISSAC'13 paper, we proposed a recursive Gaussian elimination that can compute simultaneously the row and column rank profiles of a matrix as well as those of all of its leading sub-matrices, in the same time as state of the art Gaussian elimination algorithms. Here we first study the conditions making a Gaus-sian elimination algorithm reveal this information. Therefore, we propose the definition of a new matrix invariant, the rank profile matrix, summarizing all information on the row and column rank profiles of all the leading sub-matrices. We also explore the conditions for a Gaussian elimination algorithm to compute all or part of this invariant, through the corresponding PLUQ decomposition. As a consequence, we show that the classical iterative CUP decomposition algorithm can actually be adapted to compute the rank profile matrix. Used, in a Crout variant, as a base-case to our ISSAC'13 implementation, it delivers a significant improvement in efficiency. Second, the row (resp. column) echelon form of a matrix are usually computed via different dedicated triangular decompositions. We show here that, from some PLUQ decompositions, it is possible to recover the row and column echelon forms of a matrix and of any of its leading sub-matrices thanks to an elementary post-processing algorithm. [16]

### 7.3.5. Formulas for Continued Fractions. An Automated Guess and Prove Approach

We describe a simple method that produces automatically closed forms for the coefficients of continued fractions expansions of a large number of special functions. The function is specified by a non-linear differential equation and initial conditions. This is used to generate the first few coefficients and from there a conjectured formula. This formula is then proved automatically thanks to a linear recurrence satisfied by some remainder terms. Extensive experiments show that this simple approach and its straightforward generalization to difference and $q$-difference equations capture a large part of the formulas in the literature on continued fractions. [20]

### 7.3.6. Algebraic Diagonals and Walks

The diagonal of a multivariate power series $F$ is the univariate power series $\mathrm{Diag}\, F$ generated by the diagonal terms of $F$. Diagonals form an important class of power series; they occur frequently in number theory, theoretical physics and enumerative combinatorics. We study algorithmic questions related to diagonals in the case where $F$ is the Taylor expansion of a bivariate rational function. It is classical that in this case $\mathrm{Diag}\, F$ is an algebraic function. We propose an algorithm that computes an annihilating polynomial for $\mathrm{Diag}\, F$. Generically, it is its minimal polynomial and is obtained in time quasi-linear in its size. We show that this minimal polynomial has an exponential size with respect to the degree of the input rational function. We then

address the related problem of enumerating directed lattice walks. The insight given by our study leads to a new method for expanding the generating power series of bridges, excursions and meanders. We show that their first $N$ terms can be computed in quasi-linear complexity in N, without first computing a very large polynomial equation. [12]

# COMPSYS Project-Team

# 7. New Results

## 7.1. Studying Optimal Spilling in the Light of SSA

**Participants:**  Florian Brandner [ENSTA ParisTech, previously Compsys], Quentin Colombet [Apple, previously Compsys], Alain Darte.

Recent developments in register allocation, mostly linked to static single assignment (SSA) form, have shown the benefits of decoupling the problem in two phases: a first spilling phase places load and store instructions so that the register pressure at all program points is small enough, and a second assignment and coalescing phase maps the variables to physical registers and reduces the number of move instructions among registers. We focused on the first phase, for which many open questions remain: in particular, we studied the notion of optimal spilling (what can be expressed?) and the impact of SSA form (does it help?).

To identify the important features for optimal spilling on load-store architectures, we developed a new integer linear programming formulation, more accurate and expressive than previous approaches. Among other features, we can express SSA $\phi$-functions, memory-to-memory copies, and the fact that a value can be stored simultaneously in a register and in memory. Based on this formulation, we presented a thorough analysis of the results obtained for the SPECINT 2000 and EEMBC 1.1 benchmarks, from which we have drawn, among others, the following conclusions: (1) rematerialization is extremely important; (2) SSA complicates the formulation of optimal spilling, especially because of memory coalescing when the code is not in conventional SSA (CSSA); (3) micro-architectural features are significant and thus have to be accounted for; and (4) significant savings can be obtained in terms of static spill costs, cache miss rates, and dynamic instruction counts.

Parts of this work were published at CASES 2011 [18]. The journal publication [1] contains more detailed discussions, more examples illustrating new concepts and existing approaches, and additional experiments covering the observed worst-case behavior, a new post-latency heuristic, and empiric evidence showing why static spill costs are a poor metric. Three configurations were added: Appel and George under SSA, Koes and Goldstein, and the heuristic of Braun and Hack.

## 7.2. Symbolic Range of Pointers in C programs

**Participants:**  Vitor Paisante [Univ. Mineas Gerais, Brazil], Maroua Maalej, Leonardo Barbosa [Univ. Mineas Gerais, Brazil], Laure Gonnord, Fernando Pereira [Univ. Mineas Gerais, Brazil].

Alias analysis is one of the most fundamental techniques that compilers use to optimize languages with pointers. However, in spite of all the attention that this topic has received, the current state-of-the-art approaches inside compilers still face challenges regarding precision and speed. In particular, pointer arithmetic, a key feature in C and C++, is yet to be handled satisfactorily. We designed a new alias analysis algorithm to solve this problem. The key insight of our approach is to combine alias analysis with symbolic range analysis. This combination lets us disambiguate fields within arrays and structs, effectively achieving more precision than traditional algorithms. To validate our technique, we have implemented it on top of the LLVM compiler. Tests on a vast suite of benchmarks show that we can disambiguate several kinds of C idioms that current state-of-the-art analyses cannot deal with. In particular, we can disambiguate 1.35x more queries than the alias analysis currently available in LLVM. Furthermore, our analysis is very fast: we can go over one million assembly instructions in 10 seconds.

This work has been accepted at CGO'16 [30]. An extended version of the related work is available as an Inria research report [27] and will be the basis of a journal submission.

## 7.3. Analyzing C Programs with Arrays

**Participants:** Laure Gonnord, David Monniaux [CNRS/VERIMAG].

Automatically verifying safety properties of programs is hard, and it is even harder if the program acts upon arrays or other forms of maps. Many approaches exist for verifying programs operating upon Boolean and integer values (e.g., abstract interpretation, counter-examples guided abstraction refinement using interpolants), but transposing them to array properties has been fraught with difficulties.

In contrast to most preceding approaches, we do not introduce a new abstract domain or a new interpolation procedure for arrays. Instead, we generate an abstraction as a scalar problem and feed it to a preexisting solver. The intuition is that if there is a proof of safety of the program, it is likely that it can be expressed by elementary steps between properties involving only a small (tunable) number $N$ of cells from the array.

Our transformed problem is expressed using Horn clauses over scalar variables, a common format with clear and unambiguous logical semantics, for which there exist several solvers. In contrast, solvers directly operating over Horn clauses with arrays are still very immature.

An important characteristic of our encoding is that it creates a non-linear Horn problem, with tree unfoldings, contrary to the linear problems obtained by flatly encoding the control-graph structure. Our encoding thus cannot be expressed by encoding into another control-flow graph problem, and truly leverages the Horn clause format.

Experiments with our prototype VAPHOR (see Section 6.9 ) show that this approach can prove automatically the functional correctness of several classical examples of the literature, including *selection sort*, *bubble sort*, *insertion sort*, as well as examples from previous articles on array analysis.

This work is presented in a research report [28] and is currently under submission.

## 7.4. Termination of C Programs

**Participants:** Laure Gonnord, David Monniaux [CNRS/VERIMAG], Gabriel Radanne [Univ Paris 7/ PPS].

The work of Compsys on the generation of multi-dimensional ranking functions [15], through a mix of polyhedral and abstract interpretation techniques, and its implementation in the tool RanK [16], was continued by Laure Gonnord in collaboration with D. Monniaux. A complete method for synthesizing lexicographic linear ranking functions (and thus proving termination), supported by inductive invariants, was designed in the case where the transition relation of the program includes disjunctions and existentials (large block encoding of control flow).

Previous work would either synthesize a ranking function at every basic block head, not just loop headers, which reduces the scope of programs that may be proved to be terminating, or expand large block transitions including tests into (exponentially many) elementary transitions, prior to computing the ranking function, resulting in a very large global constraint system. In contrast, the new algorithm incrementally refines a global linear constraint system according to extremal counterexamples: only constraints that exclude spurious solutions are included.

Experiments with our tool Termite 6.8  show marked performance and scalability improvements compared to other systems.

This work has been published at the PLDI'15 conference [7].

## 7.5. Data-aware Process Networks

**Participants:** Christophe Alias, Alexandru Plesco [XtremLogic SAS].

High-level circuit synthesis (HLS, high-level synthesis) consists in compiling a program described in a high-level programming language (as C) to a circuit. The circuit must be as efficient as possible while using properly the resources (power consumption, silicon area, FPGA elementary units, memory accesses, etc). Although a lot of progress was achieved on the back-end (low-level) aspects (pipeline generation, place/route), the front-end aspects (parallelism, I/O) are still rudimentary compared to the techniques developed by the HPC community, notably the analysis stemming from the *polyhedral model*.

We introduced data-aware process networks (DPN), a parallel execution model adapted to the hardware constraints of high-level synthesis, where the data transfers are made explicit. We have shown that the DPN model is consistent in the sense that any translation of a sequential program produces an equivalent DPN without deadlocks. Finally, we show how to compile a sequential program to a DPN and how to optimize the input/output and the parallelism.

This work has been published as an Inria research report [9] and will be submitted to a journal.

## 7.6. Mono-parametric Tiling

**Participants:** Guillaume Iooss, Sanjay Rajopadhye [Colorado State University], Christophe Alias, Yun Zou [Colorado State University].

Tiling is a crucial program transformation with many benefits. It improves locality, exposes parallelism, allows for adjusting the ops-to-bytes balance of codes, and can be applied at multiple levels. Allowing tile sizes to be symbolic parameters at compile time has many benefits, including efficient auto-tuning, and run-time adaptability to system variations. For polyhedral programs, parametric tiling in its full generality is known to be non-linear, breaking the mathematical closure properties of the polyhedral model. Most compilation tools therefore either avoid it by only performing fixed size tiling, or apply it only in the final, code generation step. Both strategies have limitations.

We first introduced mono-parametric partitioning, a restricted parametric, tiling-like transformation that can be used to express a tiling. We showed that, despite being parametric, it is a polyhedral transformation. We first proved that applying mono-parametric partitioning (i) to a polyhedron yields a union of polyhedra, and (ii) to an affine function produces a piecewise-affine function. We then used these properties to show how to partition an entire polyhedral program, including one with reductions. Next, we generalized this transformation to tiles with arbitrary tile shapes that can tessellate the iteration space (e.g., hexagonal, trapezoidal, etc). We showed how mono-parametric tiling can be applied at multiple levels, and how it enables a wide range of polyhedral analyses and transformations to be applied.

This work has been published as an Inria research report [14] and will be submitted to a journal. It is the extended version of the work published at IMPACT'14 [26].

## 7.7. Exact and Approximated Data-Reuse Optimizations for Tiling with Parametric Sizes

**Participants:** Alain Darte, Alexandre Isoard.

As mentioned in Section 7.6 , loop tiling is a loop transformation widely used to improve spatial and temporal data locality, to increase computation granularity, and to enable blocking algorithms, which are particularly useful when offloading kernels on computing units with smaller memories. When caches are not available or used, data transfers and local storage must be software-managed, and some useless remote communications can be avoided by exploiting data reuse between tiles. An important parameter of tiling is the sizes of the tiles, which impact the size of the required local memory. However, for most analyses involving several tiles, which is the case for inter-tile data reuse, the tile sizes induce non-linear constraints, unless they are numerical constants. This complicates or prevents a parametric analysis with polyhedral optimization techniques.

We showed that, when tiles are executed in sequence along tile axes, the parametric (with respect to tile sizes) analysis for inter-tile data reuse is nevertheless possible, i.e., one can determine, at compile-time and in a parametric fashion, the copy-in and copy-out data sets for all tiles, with inter-tile reuse, as well as sizes for the induced local memories (this is also connected to the liveness analysis described in Section 7.12 ). When approximations of transfers are performed, the situation is much more complex, and involves a careful analysis to guarantee correctness when data are both read and written. We provide the mathematical foundations to make such approximations possible, thanks to the introduction of the concept of *pointwise functions*. Combined with hierarchical tiling, this result opens perspectives for the automatic generation of blocking algorithms, guided by parametric cost models, where blocks can be pipelined and/or can contain parallelism. Previous work on FPGAs and GPUs already showed the interest and feasibility of such automation with tiling, but in a non-parametric fashion.

Our method is currently implemented with the iscc calculator of ISL, a library for the manipulation of integer sets defined with Presburger arithmetic, a complete implementation within the PPCG compiler is in progress (see also Section 6.7 ).

We believe that our approximation technique can be used for other applications linked to the extension of the polyhedral model as it turns out to be fairly powerful. Our future work will be to derive efficient approximation techniques, either because the program cannot be fully analyzable, or because approximations can speed-up or simplify the results of the analysis without losing much in terms of memory transfers and/or memory sizes.

A preliminary version of this work has been presented at the IMPACT'14 workshop [19]. A revised version was published at the International Conference on Compiler Construction (CC'15) [3].

## 7.8. Analysis of X10 Programs

**Participants:** Paul Feautrier, Alain Ketterlin [Inria/CAMUS], Sanjay Rajopadhye [Colorado State University], Vijay Saraswat [IBM Research], Eric Violard [Inria/CAMUS], Tomofumi Yuki.

While, historically, Compsys has applied polyhedral analysis to sequential programs, it was recently realized that it also applies to parallel programs or specifications, with the aim of checking their correctness or improving their performance. The prospect of having to program exascale architectures, with their millions of cores, has led to the development of new programming languages, whose objective is to increase the programmer productivity. Compsys has first applied polyhedral techniques to synchronous languages [24], [25] and pipelined specifications (see Section 7.7 ), before concentrating on IBM's high-productivity language X10 (see this section as well as Section 7.9 ) and on the OpenStream language (see Section 7.10 ).

X10 is based on the creation of independent *activities* (light-weight threads), which can synchronize either by a generalization of the fork/join scheme, or with *clocks*, an improved version of the familiar barriers. X10 is deadlock-free by construction but it is the programmer responsibility to insure determinism by a proper use of synchronizations. Non-determinism bugs may have a very low occurrence probability thus be very difficult to detect by testing, hence the interest for detecting races at compile time. In collaboration with CSU (S. Rajopadhye, T. Yuki) and IBM (V. Saraswat), we first extended array dataflow analysis to polyhedral clock-free X10 programs [34]. We have been working on clocked programs too. Race detection becomes undecidable [35], but realistic problems may still be solved by heuristics.

In cooperation with Eric Violard and Alain Ketterlin (Inria Team Camus, Strasbourg), and in order to obtain a more secure and precise analysis, we are currently attempting to formalize the "happens before" analysis used in these two previous papers [34], [35], using the proof assistant Coq.

## 7.9. Revisiting Loop Transformations with X10 Clocks

**Participant:** Tomofumi Yuki.

Loop transformations are known to be important for performance of compute-intensive programs, and are often used to expose parallelism. However, many transformations involving loops often obfuscate the code, and are cumbersome to apply by hand. In this work, we explored alternative methods for expressing parallelism that are more friendly to the programmer. In particular, we seek to expose parallelism without significantly changing the original loop structure. We illustrated how clocks in X10 can be used to express some of the traditional loop transformations, in the presence of parallelism, in a manner that we believe to be less invasive. Specifically, expressing parallelism corresponding to one-dimensional affine schedules can be achieved without modifying the original loop structure and/or statements.

This work was published at the international workshop on X10 [8].

## 7.10. Static Analysis of OpenStream Programs

**Participants:** Albert Cohen [Inria Parkas team], Alain Darte, Paul Feautrier.

In the context of the ManycoreLabs project (see Section 8.1 ), we also studied the applicability of polyhedral techniques to the parallel language OpenStream [31]. When applicable, polyhedral techniques are indeed invaluable for compile-time debugging and for generating efficient code well suited to a target architecture. OpenStream is a two-level language in which a control program directs the initialization of parallel task instances that communicate through *streams*, with possibly multiple writers and readers. It has a fairly complex semantics in its most general setting, but we restricted ourselves to the case where the control program is sequential, which is representative of the majority of the OpenStream applications.

In contrast to X10, this restriction offers deterministic concurrency by construction, but deadlocks are still possible. We showed that, if the control program is polyhedral, one may statically compute, for each task instance, the read and write indices to each of its streams, and thus reason statically about the dependences among task instances (the only scheduling constraints in this polyhedral subset). If the control program has nested loops, communications use one-dimensional channels in a form of linearization, and these indices may be polynomials of arbitrary degree, thus requiring to extend to polynomials the standard polyhedral techniques for dependence analysis, scheduling, and deadlock detection. Modern SMT allow to solve polynomial problems, albeit with no guarantee of success; the approach previously developed by P. Feautrier [6] may offer an alternative solution.

The usual way of disproving deadlocks is by exhibiting a schedule for the program operations, a well-known problem for polyhedral programs where dependences can be described by affine constraints. In the case of OpenStream, we established two important results related to deadlocks: 1) a characterization of deadlocks in terms of dependence paths, which implies that streams can be safely bounded as soon as a schedule exists with such sizes, 2) the proof that deadlock detection is undecidable, even for polyhedral OpenStream.

Details of this work are available in a research report [10]. It will be presented at the international workshop IMPACT'16 [2]. Some further developments are in progress for scheduling OpenStream programs using polynomial techniques, see Section 6.4 .

## 7.11. Handling Polynomials for Program Analysis and Transformation

**Participant:** Paul Feautrier.

As shown in Section 7.10 , many problems in parallel programs analysis and verification can be reduced to proving or disproving properties of polynomials in the variables of the program. For instance, the so-called "linearizations" (replacing a multi-dimensional object by a one-dimensional one) generate polynomial access functions. These polynomials then reappear in dependence testing, scheduling, and invariant construction. It may also happen that polynomials are absent from the source program, but are created either by an enabling analysis, as for OpenStream, or are imposed by complexity consideration. The usual solution is to construct a multi-dimensional function (e.g., a schedule for parallelization or a ranking function for termination [15]), which can then be converted into polynomials by counting. However, a direct approach is preferable, especially when the resulting schedule is to be used for further analysis, e.g., in real-time situations or WCET evaluation.

What is needed here is a replacement for the familiar emptiness tests and for Farkas lemma (deciding whether an affine form is positive inside a polyhedron). Recent mathematical results by Handelman and Schweighofer on the *Positivstellensatz* allow one to devise algorithms that are able to solve these problems. The difference is that one gets only sufficient conditions, and that complexity is much higher than in the affine cases. A paper presenting applications of these ideas to three use cases – dependence testing, scheduling, and transitive closure approximation – was presented at the 5th International Workshop on Polyhedral Compilation Techniques (IMPACT'15) [6] in Amsterdam in January 2015. A tool implementing polyhedral schedules complements this work, see Section 6.6 .

## 7.12. Liveness Analysis in Explicitly-Parallel Programs

**Participants:** Alain Darte, Alexandre Isoard, Tomofumi Yuki.

In the light of the parallel specifications encountered in our other works (from Section 7.7 to Section 7.11 ), we revisited scalar and array element-wise liveness analysis for programs with parallel specifications. In earlier work on memory allocation/contraction (register allocation or intra- and inter-array reuse in the polyhedral model), a notion of "time" or a total order among the iteration points was used to compute the liveness of values. In general, the execution of parallel programs is not a total order, and hence the notion of time is not applicable.

We first revised how conflicts are computed by using ideas from liveness analysis for register allocation, studying the structure of the corresponding conflict/interference graphs. Instead of considering the conflict between two pairs of live ranges, we only consider the conflict between a live range and a write. This simplifies the formulation from having four instances involved in the test down to three, and also improves the precision of the analysis in the general case.

Then we extended the liveness analysis to work with partial orders so that it can be applied to many different parallel languages/specifications with different forms of parallelism. An important result is that the complement of the conflict graph with partial orders is directly connected to memory reuse, even in presence of races. However, programs with conditionals do not even have a partial order, and our next step will be to handle such cases with more accuracy.

Details of this work are available in a research report [13]. It will be presented at the international workshop IMPACT'16 [4].

## 7.13. Extended Lattice-Based Memory Allocation

**Participants:** Alain Darte, Alexandre Isoard, Tomofumi Yuki.

We extended lattice-based memory allocation [20], an earlier work on memory (array) reuse analysis. The main motivation is to handle in a better way the more general forms of specifications we see today, e.g., with loop tiling, pipelining, and other forms of parallelism available in explicitly parallel languages. Our extension has two complementary aspects. We showed how to handle more general specifications where conflicting constraints (those that describe the array indices that cannot share the same location) are specified as a (non-convex) union of polyhedra. Unlike convex specifications, this also requires to be able to choose suitable directions (or basis) of array reuse. For that, we extended two dual approaches, previously proposed for a fixed basis, into optimization schemes to select suitable basis. Our final approach relies on a combination of the two, also revealing their links with, on one hand, the construction of multi-dimensional schedules for parallelism and tiling (but with a fundamental difference that we identify) and, on the other hand, the construction of universal reuse vectors (UOV), which was only used so far in a specific context, for schedule-independent mapping.

This algorithmic work, connected to the parametric tiling of Section 7.7 and the liveness analysis results of Section 7.12 , is complemented by a set of prototype scripting tools, see Section 6.3 .

Details of this work are available in a research report. It has also been submitted to a conference.

## 7.14. Stencil Accelerators

**Participants:** Steven Derrien [University of Rennes 1, Inria/CAIRN], Xinyu Niu [Imperial College London], Sanjay Rajopadhye [Colorado State University], Tomofumi Yuki.

Stencil computations have been known to be an important class of programs for scientific calculations. Recently, various architectures (mostly targeting FPGAs) for stencils are being proposed as hardware accelerators with high throughput and/or high energy efficiency. There are many different challenges for such design: How to maximize compute-I/O ratio? How to partition the problem so that the data fits on the on-chip memory? How to efficiently pipeline? How to control the area usage? We seek to address these challenges by combining techniques from compilers and high-level synthesis tools.

One project in collaboration with the CAIRN team and Colorado State University targets stencils with regular dependence patterns. Although many architectures have been proposed for this type of stencils, most of them use a large number of small processing elements (PE) to achieve high throughput. We are exploring an alternative design that aims for a single, large, deeply-pipelined PE. The hypothesis is that the pipelined parallelism is more area-efficient compared to replicating small PEs. We have published a work-in-progress paper on this topic at IMPACT'16 [5].

Another type of stencil accelerators that we are working on, in collaboration with Xinyu Niu, targets stencil programs with dynamic dependences (i.e., sparse computations). The collaboration is in the context of the EURECA project [0] where the dynamic reconfigurability of modern FPGAs are used to efficiently handle dynamic access patterns.

## 7.15. PolyApps

**Participant:** Tomofumi Yuki.

Loop transformation frameworks using the polyhedral model have gained increased attention since the rise of the multi-core era. We now have several research tools that have demonstrated their power on important kernels found in scientific computations. However, there remains a large gap between the typical kernels used to evaluate these tools and the actual applications used by the scientists.

PolyApps is an effort to collect applications from other domains of science to better establish the link between the compiler tools and "real" applications. The applications are modified to bypass some of the front-end issues of research tools, while keeping the ability to produce the original output. The goal is to assess how the state-of-the-art automatic parallelizers perform on full applications, and to identify new opportunities that only arise in larger pieces of code.

We showed that, with a few enhancements, the current tools will be able to reach and/or exceed the performance of existing parallelizations of the applications. One of the most critical element missing in current tools is the ability to modify the memory mappings.

---

[0] http://www.doc.ic.ac.uk/~nx210/2015/09/01/eureca.html

<p align="center"><span style="color:red">**CONVECS Project-Team**</span></p>

# 6. New Results

## 6.1. New Formal Languages and their Implementations

### 6.1.1. Definition of LNT

**Participants:** Hubert Garavel, Frédéric Lang, Wendelin Serwe.

LNT is a next generation formal description language for asynchronous concurrent systems, which attempts to combine the best features of imperative programming languages and value-passing process algebras. LNT is increasingly used by CONVECS for industrial case studies and applications (see § 6.5 ) and serves also in university courses on concurrency, in particular at ENSIMAG (Grenoble) and at Saarland University.

In 2015, the theoretical foundations of LNT have been explored in a journal article [14] that, after examining the various ways sequential composition is handled in mainstream value-passing process calculi, shows that these various approaches are subsumed by the LNT approach, which is easier to learn and leads to more readable and more concise specifications.

The LNT language has also been enhanced in several aspects:

- The "case" construct now supports multiple (tuple-like) expressions and patterns.
- Two new parameter-passing modes "in var" and "out var" have been introduced to allow finer data-flow analyses.
- Exceptions are better handled and a new "assert" statement was added to LNT.
- The "none" channel is now implicitly predefined.
- Finally, the LNT reference manual has been extended and updated at many places.

### 6.1.2. Translation from LNT to LOTOS

**Participants:** Hubert Garavel, Frédéric Lang, Wendelin Serwe.

In 2015, the translator from LNT to LOTOS was further improved. In addition to 22 bug fixes and improved error messages, the following enhancements have been brought:

- The "-root" option of LNT2LOTOS now accepts value parameters for LNT processes and supports gate parameters in named style. It also accepts the name of a process not present in the current module.
- Negative number constants of the form "$-2^{k-1}$", where integer numbers are represented using $k$ bits, are now supported.
- Better warning messages are emitted for "in" and "in out" (formerly "inout") parameters.

### 6.1.3. Translation from LOTOS to Petri nets and C

**Participants:** Hubert Garavel, Wendelin Serwe.

The LOTOS compilers CAESAR and CAESAR.ADT, which were once the flagship of CADP, now play a more discrete role since LNT (rather than LOTOS) has become the recommended specification language of CADP. Thus, CAESAR and CAESAR.ADT are mostly used as back-end translators for LOTOS programs automatically generated from LNT or other formalisms such as Fiacre, and are only modified when this appears to be strictly necessary.

In 2015, in addition to a few bug fixes, the "`-root`" option of the CAESAR compiler has been generalized to support processes having value parameters; this makes compositional verification easier by removing the need for introducing extra wrapper processes having no value parameters. The EXEC/CAESAR interface has been extended with two new primitives "`CAESAR_KERNEL_DELAY`" and "`CAESAR_KERNEL_EXIT()`". Also, optimizations have been introduced to generate shorter and simpler C code, and to make sure that this C code compiles without spurious warnings.

A systematic comparison between CAESAR.ADT and available interpreters/compilers for other languages that support rewrite rules or pattern matching has been undertaken. This comparison reuses the benchmarks developed for the three Rewrite Engine Competitions (2006, 2009, and 2010). As a preliminary step, we developed a tenth translators from the REC formalism in which these benchmarks are written to languages such as Haskell, LOTOS, Maude, mCRL, OCAML, Opal, Rascal, Scala, and Tom.

### 6.1.4. NUPN

**Participants:** Hubert Garavel, Frédéric Lang.

The CAESAR.BDD tool that analyzes NUPN (*Nested-Unit Petri Nets*) models and serves to prepare the yearly Model Checking Contest [0] has been enhanced in several ways. In addition to 7 bug fixes, 14 new command-line options have been added to CAESAR.BDD ("`-arcs`", "`-bits`", "`-creator`", "`-density`", "`-encodings`", "`-height`", "`-hwb`", "`-multiple-arcs`", "`-multiple-initial-tokens`", "`-places`", "`-redundant-units`", "`-transitions`", "`-units`", and "`-width`"). The output format produced by the "`-exclusive-places`" option has been revised. The "`-mcc`" option now computes the extended free choice property. A new option "`-network nupn`" was also added to EXP.OPEN to produce NUPN models from automata networks.

Particular efforts have been put to increase the scalability of CAESAR.BDD for large models. Reading large NUPN files was made much faster. The "`-exclusive-places`" option of CAESAR.BDD was made faster too. The size of the largest data structure allocated by CAESAR.BDD, has been divided by four. CAESAR.BDD has also been optimized to save memory when handling NUPN models having a simple hierarchical structure. Finally, user-specified timeouts are better supported.

A conference article was published [24], which formally defines the NUPN model and investigates its mathematical properties. Additionally, the assembly of a collection of large NUPN models was undertaken, and various prototype tools to handle NUPN models ("nupn_pack", "nupn_reduce", and "nupn_merge") have been developed.

### 6.1.5. *Translation from GRL to LNT*

**Participants:** Fatma Jebali, Jingyan Jourdan-Lu, Frédéric Lang, Eric Léo, Radu Mateescu.

In the context of the Bluesky project (see § 8.1.2.1 ), we study the formal modeling of GALS (*Globally Asynchronous, Locally Synchronous*) systems, which are composed of several synchronous subsystems evolving cyclically, each at its own pace, and communicating with each other asynchronously. Designing GALS systems is challenging due to both the high level of (synchronous and asynchronous) concurrency and the heterogeneity of computations (deterministic and nondeterministic). To bring our formal verification techniques and tools closer to the GALS paradigm, we designed a new formal language named GRL (*GALS Representation Language*), as an intermediate format between GALS models and purely asynchronous concurrent models. GRL combines the main features of synchronous dataflow programming and asynchronous process calculi into one unified language, while keeping the syntax homogeneous for better acceptance by industrial GALS designers. GRL allows a modular composition of synchronous systems (blocks), environmental constraints (environments), and asynchronous communication mechanisms (mediums), to be described at a level of abstraction that is appropriate to verification. GRL also supports external C and LNT code. A translator named GRL2LNT has been developed, allowing an LNT program to be generated from a GRL specification automatically. Additionally, an OPEN/CAESAR-compliant compiler named GRL.OPEN (based on GRL2LNT and LNT.OPEN) makes possible the on-the-fly exploration of the LTS underlying a GRL specification using CADP.

---

[0] http://mcc.lip6.fr/

In 2015, we have revised the GRL syntax to make GRL easier to learn and to understand. Our data base of examples has been updated to take those changes into account. We have also added some language features, such as named constants, and a dedicated construct called *activation signal* to define constraints on the asynchronous activation of blocks. This new construct is easier to use than the previous solution based on ad-hoc data signals, and semantically more elegant as it avoids unexpected deadlocks. Activation signals permit realistic situations such as halting, priorities, scenarios, and pace relations between synchronous components to be modeled at a suitable level of abstraction. They can be smoothly translated into LNT without affecting the rest of the translation.

As regards the specification of properties, to reduce the complexity of using full-fledged temporal logics, we have also proposed a property specification language dedicated to GALS systems, based upon a set of temporal logic patterns, which capture frequently encountered behaviours, encompassing both time-critical and untimed aspects of GALS systems. Those patterns include deadlock, livelock, safety, liveness, and fairness properties. The semantics of the proposed patterns have been defined by translation into the MCL language.

As regards the GRL2LNT tool, nine successive versions have been released, to take into account the syntactic changes in the GRL language, to correct about 20 bugs, to eliminate compilation warnings, and to implement the following new features:

- The generated LNT code has been corrected so as to eliminate compilation warnings and to take into account recent changes in the syntax of LNT (see § 6.1.1 ).

- GRL system specifications can now be parameterized with data values and instantiated using the new "-root" option of GRL2LNT.

- The interface between GRL and external C code has been revised in two ways: (1) external blocks with several outputs are now mapped to a single external function instead of one function per output previously, and (2) conversion functions between GRL and C numeric types have been defined, handling runtime verification of overflows. Those conversion functions have been packaged in a new code library, which is automatically included by GRL2LNT.

- Several verifications on the usage of signals and communication channels have been implemented, leading either to error messages, or to warnings corresponding to potential errors. About 20 new error messages and 10 new warnings have been added.

In addition, three manual pages have been written to document respectively the GRL language, the GRL2LNT translator tool, and the GRL.OPEN shell script. The GRL non-regression test base has been extended and now contains 230 correct examples and 400 incorrect examples.

An article describing the GRL language and its associated tools has been submitted to an international journal.

### 6.1.6. Translation from BPMN to LNT

**Participant:** Gwen Salaün.

Business processes support the modeling and the implementation of software as workflows of local and inter-process activities. Taking over structuring and composition, evolution has become a central concern in software development. We believe this should be taken into account as soon as the modeling of business processes, which can thereafter be made executable using process engines or model-to-code transformations. We advocate that business process evolution can be formally analyzed in order to compare different versions of processes, identify precisely the differences between them, and ensure the desired consistency.

To reach this objective, we developed, in collaboration with Pascal Poizat (LIP6, Paris), a model transformation from the BPMN standard notation to the LNT process algebra. We then proposed a set of relations for comparing business processes at the formal model level. With reference to related work, we proposed a richer set of comparison primitives supporting renaming, refinement, property- and context-awareness. Thanks to the implementation of a tool that integrates with the Eclipse IDE and behind-the-scene interaction with the CADP verification toolbox, we put the checking of evolution within the reach of business process designers. Our approach is fully automated and has been applied for evaluation to a large set of BPMN processes.

### 6.1.7. *Other Language Developments*

**Participants:** Hugues Evrard, Hubert Garavel, Frédéric Lang, Eric Léo, Wendelin Serwe.

The ability to compile and verify formal specifications with complex, user-defined operations and data structures is a key feature of the CADP toolbox since its very origins. A long-run effort has been recently undertaken to ensure a uniform treatment of types, values, and functions across all the various CADP tools.

In 2015, the connection to external software development tools has progressed. The support of the LOTOS and LNT languages in the Emacs/XEmacs, jEdit, and Vim editors has improved. More text editors are now supported, including Nano, Notepad++, and all the text editors compliant with GtkSourceView 3.0 (including the Gedit editor of Gnome). Pretty-printers such as a2ps and the LaTeX "listings" package are also supported. Configuration files for three CADP languages (MCL, SVL, and XTL) and three CADP formats (BES, NUPN, and RBC) have been added.

Also, with the help of its principal author Pierre Boullier (Inria, Alpage), we corrected a memory allocation bug in the SYNTAX parser generator, which is used in most of the compilers developed by the CONVECS team.

## 6.2. Parallel and Distributed Verification

### 6.2.1. *Distributed Code Generation for LNT*

**Participants:** Hugues Evrard, Frédéric Lang.

Rigorous development and prototyping of a distributed algorithm using LNT involves the automatic generation of a distributed implementation. For the latter, a protocol realizing process synchronization is required. As far as possible, this protocol must itself be distributed, so as to avoid the bottleneck that would inevitably arise if a unique process would have to manage all synchronizations in the system. A particularity of such a protocol is its ability to support branching synchronizations, corresponding to situations where a process may offer a choice of synchronizing actions (which themselves may nondeterministically involve several sets of synchronizing processes) instead of a single one. Therefore, a classical barrier protocol is not sufficient and a more elaborate synchronization protocol is needed.

Using a synchronization protocol that we verified formally in 2013, we developed a prototype distributed code generator, named DLC (*Distributed LNT Compiler*), which takes as input the model of a distributed system described as a parallel composition of LNT processes.

In 2015, we finalized the development of DLC: the code was cleaned and the different compiler components were better integrated. A new option was added for the generated executables to dump at runtime an execution trace in the SEQUENCE format of CADP, for further analysis. A complete description of DLC, its synchronization protocol, performance data and usage examples were presented in Hugues Evrard's PhD thesis [9], defended in July 2015. An overview of DLC was presented in an international conference paper [23], and an extended version has been prepared for a journal special issue currently under construction. A tool paper was accepted in an international conference that will take place in 2016 [22].

### 6.2.2. *Verification of Asynchronously Communicating Systems*

**Participants:** Lakhdar Akroun, Gwen Salaün.

Analyzing systems communicating asynchronously via reliable FIFO buffers is an undecidable problem. A typical approach is to check whether the system is bounded, and if not, whether the corresponding state space can be made finite by limiting the presence of communication cycles in behavioral models or by fixing the buffer size. In this work, our focus is on systems that are likely to be unbounded and therefore result in infinite systems. We do not want to restrict the system by imposing any arbitrary bound. We introduced a notion of stability and proved that once the system is stable for a specific buffer bound, it remains stable whatever larger bounds are chosen for buffers. This enables one to check certain properties on the system for that bound and to ensure that the system will preserve them whatever larger bounds are used for buffers. We also proved that computing this bound is undecidable but we showed how we can succeed in computing these bounds for many typical examples using heuristics and equivalence checking.

### *6.2.3. Analysis of Verification Counterexamples*

**Participants:** Gianluca Barbon, Gwen Salaün.

Model checking is an established technique for automatically verifying that a model, e.g., a Labelled Transition System (LTS), obtained from higher-level specification languages (such as process algebras) satisfies a given temporal property, e.g., the absence of deadlocks. 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 (even thousands) of actions, (ii) the debugging task is mostly achieved manually, and (iii) the counterexample does not give any clue on the state of the system (e.g., parallelism or data expressions) when the error occurs.

In collaboration with the SLIDE team of the LIG laboratory, we work on new solutions for simplifying the comprehension of counterexamples and thus favouring usability of model checking techniques. To do so, we apply pattern mining techniques to a set of correct traces (extracted from the LTS) and incorrect traces (corresponding to counterexamples), to identify specific patterns indicating more precisely the source of the problem.

## 6.3. Timed, Probabilistic, and Stochastic Extensions

### *6.3.1. Model Checking for Extended PCTL*

**Participants:** Radu Mateescu, José Ignacio Requeno.

In the context of the SENSATION project (see § 8.2.1.1 ), we study the specification and verification of quantitative properties of concurrent systems.

In 2015, we developed a probabilistic version of ACTL (Action-based CTL) [41], named PACTL. This logic represents an action-based counterpart for PCTL (*Probabilistic Computation Tree Logic*) [50] and is interpreted naturally over DTMCs with labeled transitions, such as those produced from IPCs (*Interactive Probabilistic Chains*) [40]. The PACTL operators generalize those of ACTL: they characterize sequences of transitions in the DTMC by specifying both the states and the actions labeling the transitions. We implemented PACTL as an XTL library, which allows the designer to combine properties on actions, data, probabilities, and discrete time. We have experimented the PACTL library on several DTMCs imported from the probabilistic model checker PRISM [55] to ensure that both implementations provide the same numerical results.

## 6.4. Component-Based Architectures for On-the-Fly Verification

### *6.4.1. Compositional Verification*

**Participants:** Hubert Garavel, Frédéric Lang.

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 2015, we corrected one bug in BCG_CMP and eight bugs in SVL. We extended the SVL language and compiler as follows:

- A new statement was added to translate a LOTOS file or a process in a LOTOS file to an LNT file automatically.

- LNT processes with data parameters can now be instantiated directly in the SVL script, without requiring a parameterless intermediate process to be defined.

- LNT processes with gate parameters can now be instantiated in the SVL script using the named parameter-passing style of LNT.

- Specification of a diagnostic file is now optional in the "`comparison`", "`deadlock`", and "`live-lock`" statements of SVL.

- The "`property`" statement has been extended so that it can now contain any kind of statement, provided it contains at least one verification statement.

- Within SVL properties, it is now possible to define shell lines followed by an "`expected`" clause to specify the expected result of the shell line.

- It is now possible to add a "`result`" clause after a verification statement, so as to store the result of the verification in a shell variable that can be subsequently used in the SVL script.

We improved several demo examples of CADP by using these new SVL constructs, and we added a new demo example on the verification of an airplane-ground communication protocol.

We also improved the PMC tool, by correcting five bugs and adding a new "`-order`" option, which permits the user to define a particular order for quotienting. We improved the presentation of the demo examples released in the PMC distribution. Those examples are now given in LNT and translated automatically into networks of automata in the EXP language, instead of being given directly as networks of automata.

### 6.4.2. *On-the-Fly Test Generation*

**Participants:** Hubert Garavel, Radu Mateescu, Wendelin Serwe.

In the context of the collaboration with STMicroelectronics, we study techniques for testing if a (hardware) implementation is conform to a formal model described in LNT. Our approach is inspired by the theory of conformance testing [63], as implemented for instance in TGV [53] and JTorX [33]. We have developed three prototype tools to support this approach. The first tool implements a dedicated OPEN/CAESAR-compliant compiler for the particular asymmetric synchronous product between the model and the test purpose. The second tool, based on slightly extended generic components for graph manipulation ($\tau$-compression, $\tau$-confluence reduction, determinization) and resolution of Boolean equation systems, generates the complete test graph (CTG), which can be used to extract concrete test cases or to drive the test of the implementation. A third prototype tool takes as input a CTG and extracts either a single test case (randomly chosen or the first encountered one), or the set of *all* test cases. The principal advantage of our approach compared to existing tools is the use of LNT for describing test purposes, which facilitates the manipulation of data values.

In 2015, we corrected the prototype tools to properly handle timers and failure transitions, improved the documentation, and simplified internal data structures.

These prototype tools were used in the case study with STMicroelectronics (see § 6.5.1 ) and the EnergyBus (see § 6.5.4 ).

### 6.4.3. *Other Component Developments*

**Participants:** Soraya Arias, Hubert Garavel, Frédéric Lang, Radu Mateescu.

We separated the MCL library defining the operators of ACTL (Action-based CTL) [41] in two parts: the first one defines the operators of ACTL∖X (the fragment of ACTL without the next-time operators), including optimized definitions of derived temporal operators, and the second one defines the next-time operators, including the definitions of silent next-time operators, which complement the visible next-time operators already present in the library.

We also added an MCL library defining the operators of the L$\mu$-dsbr fragment of modal $\mu$-calculus [6], which includes the ACTL∖X library. The L$\mu$-dsbr library also defines the absence of deadlock property as an MCL formula adequate w.r.t. divergence-sensitive branching bisimulation (divbranching for short) and allowing one to hide all visible actions in the LTS and to reduce it modulo divbranching prior to verification, which may bring significant performance gains.

A new major version 1.2 of the BCG format for storing Labelled Transition Systems was released as part of CADP 2015-a. Following this change, various minor residual bugs have been identified and fixed in 2015, and the type system of XTL has been modified to require fewer explicit type coercions.

In addition to bug fixes in various tools (e.g., CUNCTATOR, EUCALYPTUS, TST, XTL, etc.), the installation procedures of CADP have been revisited and updated; in particular, work is going on and many preliminary changes have been silently brought to ease installation of CADP on Windows.

# 6.5. Real-Life Applications and Case Studies

## 6.5.1. *ACE Cache Coherency Protocol*

**Participants:**  Abderahman Kriouile, Radu Mateescu, Wendelin Serwe.

In the context of a CIFRE convention with STMicroelectronics, we studied system-level cache coherency, a major challenge faced in the current System-on-Chip architectures. Because of their increasing complexity (mainly due to the significant number of computing units), the validation effort using current simulation-based techniques grows exponentially. As an alternative, we study formal verification.

We focused on the ACE (AXI Coherency Extensions) cache coherency protocol, a system-level coherency protocol proposed by ARM  [31]. In previous years, we developed a parametric formal model (about $3,400$ lines of LNT) of a system consisting of an ACE-based cache coherent interconnect, processors, and a main memory. We also specified temporal properties expressing cache coherence, data integrity, and successful completion of each transaction. Note that the former property required to transform state-based properties into action-based ones, by adding information about the cache state to the actions executed by the cache.

In 2015, we continued to exploit the formal model to improve the validation of the architecture under design at STMicroelectronics, in particular by integrating tests derived from the formal model into the official test plans. This work led to a publication in an international conference [25], and the defense of the PhD corresponding to the CIFRE convention [10].

## 6.5.2. *Deployment and Reconfiguration Protocols for Cloud Applications*

**Participants:**  Rim Sakka Abid, Gwen Salaün.

Cloud applications are complex applications composed of a set of interconnected software components running on different virtual machines, hosted on remote physical servers. Deploying and reconfiguring this kind of applications are very complicated tasks especially when one or multiple virtual machines fail when achieving these tasks. Hence, there is a need for protocols that can dynamically reconfigure and manage running distributed applications.

In 2015, we proposed a novel protocol, which aims at reconfiguring cloud applications. This protocol is able to ensure communication between virtual machines and resolve dependencies by exchanging messages, (dis)connecting, and starting/stopping components in a specific order. The interaction between machines is assured via a publish-subscribe messaging system. Each machine reconfigures itself in a decentralized way. The protocol supports virtual machine failures, and the reconfiguration always terminates successfully even in the presence of a finite number of failures. Due to the high degree of parallelism inherent to these applications, the protocol was specified in LNT and verified using CADP. The use of formal specification languages and tools helped to detect several bugs and to improve the protocol. These results were published in [12].

Another line of work concerns autonomic computing in cloud environments. Managing distributed cloud applications is a challenging problem because manual administration is no longer realistic for these complex distributed systems. Thus, autonomic computing is a promising solution for monitoring and updating these applications automatically. This is achieved through the automation of administration functions and the use of control loops called autonomic managers. An autonomic manager observes the environment, detects changes, and reconfigures dynamically the application. Multiple autonomic managers can be deployed in the same system and must make consistent decisions. Using them without coordination may lead to inconsistencies and error-prone situations.

In 2015, we propose an approach for coordinating stateful autonomic managers, which relies on a simple coordination language, new techniques for asynchronous controller synthesis and Java code generation. We used our approach for coordinating real-world cloud applications. These results were published in [19].

### 6.5.3. *Networks of Programmable Logic Controllers*

**Participants:**  Fatma Jebali, Jingyan Jourdan-Lu, Frédéric Lang, Eric Léo, Radu Mateescu.

In the context of the Bluesky project (see § 8.1.2.1 ), we study the software applications embedded on the PLCs (Programmable Logic Controllers) manufactured by Crouzet Automatismes. One of the objectives of Bluesky is to enable the rigorous design of complex control applications running on several PLCs connected by a network. Such applications are instances of GALS (*Globally Asynchronous, Locally Synchronous*) systems composed of several synchronous automata embedded on individual PLCs, which interact asynchronously by exchanging messages. A formal analysis of these systems can be naturally achieved by using the formal languages and verification techniques developed in the field of asynchronous concurrency.

For describing the applications embedded on individual PLCs, Crouzet provides a dataflow language with graphical syntax and synchronous semantics, equipped with an ergonomic user-interface that facilitates the learning and use of the language by non-experts. To equip the PLC language of Crouzet with functionalities for automated verification, the solution adopted in Bluesky was to translate it into GRL (see § 6.1.5 ), which enables the connection to testing and verification tools covering the synchronous and asynchronous aspects.

In 2015, we have provided support to Crouzet, who started to integrate GRL in the PLC design process by developing both a library of GRL blocks corresponding to function blocks present in their PLC programming tool, and an automated translation from a PLC program into a GRL block. The GRL2LNT and GRL.OPEN tools (see § 6.1.5 ) provide a direct connection to all verification functionalities of CADP, in particular model checking and equivalence checking.

We also investigated the equivalence checking for networks of PLCs, with the objective of proposing a general methodology usable in industrial context. We identified several rules (formalized as templates) for describing the asynchronous and synchronous parts of PLC networks, as well as their external behaviour (service), in order to facilitate the equivalence checking modulo branching bisimulation.

### 6.5.4. *EnergyBus Standard for Connecting Electric Components*

**Participants:**  Hubert Garavel, Wendelin Serwe.

The EnergyBus [0] is an upcoming industrial standard for electric power transmission and management, based on the CANopen field bus. It is developed by a consortium assembling all major industrial players (such as Bosch, Panasonic, and Emtas) in the area of light electric vehicles (LEV); their intention is to ensure interoperability between all electric LEV components. At the core of this initiative is a universal plug integrating a CAN-Bus [0] with switchable power lines. The central and innovative role of the EnergyBus is to manage the safe electricity access and distribution inside an EnergyBus network.

In the framework of the European FP7 project SENSATION (see § 8.2.1.1 ) a formal specification in LNT of the main EnergyBus protocols is being developed by Alexander Graf-Brill and Holger Hermanns at Saarland University  [48], with the active collaboration of CONVECS.

In 2015, we pursued the analysis of the LNT model, involving both verification (by means of state-space exploration and model checking techniques) and validation (using test cases automatically derived from the formal LNT model).

### 6.5.5. *AutoFlight Control System*

**Participant:**  Fatma Jebali.

---

[0]http://www.energybus.org
[0]http://www.can-cia.org

In collaboration with Eric Jenn (IRT Saint Exupery, Toulouse), we studied an AutoFlight Control System (AFCS), provided by Thales Avionics. The goal of an AFCS is to improve the quality of flight and enhance the operational capability of the aircraft. The architecture of the AFCS comprises two parts. The first part (FCP, Flight Control Panel) consists of a control panel, which enables the pilot to interact with the system. The second part (AFS, Automatic Flight System) is in charge of performing functions such as guidance and automatic pilot. For safety purposes, each part is organized into a command and monitoring channels. The command channel ensures the function allocated to the component. The monitoring channel ensures that the command channel operates correctly. To ensure a sufficient availability level, a high level of redundancy is built inside the system. Components communicate using various communication means with different latencies (AFDX, A429, discrete).

Since AFCSs have stringent safety and time-critical requirements, formal verification is required to ensure their correctness. To this aim, we have applied the GRL approach for the formal modelling and verification of GALS systems (see § 6.1.5 ). In a first step, we have addressed the AFCS without redundancy. We have written a GRL description (750 lines), which can be parameterized by the activation paces of different synchronous components. We have written a set of correctness properties in MCL, which we have verified on the GRL model.

### 6.5.6. *Graphical User-Interfaces and Plasticity*

**Participants:**  Hubert Garavel, Frédéric Lang, Raquel Oliveira.

In the context of the Connexion project (see § 8.1.1.2 ) and in close collaboration with Gaëlle Calvary and Sophie Dupuy-Chessa (IIHM team of the LIG laboratory), we study the formal description and validation of graphical user-interfaces using the most recent features of the CADP toolbox. The case study assigned to LIG in this project is a prototype graphical user-interface  [38] designed to provide human operators with an overview of a running nuclear plant. The main goal of the system is to inform the operators about alarms resulting from faults, disturbances, or unexpected events in the plant. Contrary to conventional control rooms, which employ large desks and dedicated hardware panels for supervision, this new-generation interface uses standard computer hardware (i.e., smaller screen(s), keyboard, and mouse), thus raising challenging questions on how to best provide synthetic views of the plant status. Another challenge is to introduce plasticity in such interface, so as to enable several supervision operators, including mobile ones outside of the control room, to get accurate information in real time.

We formally specified this new-generation interface in LNT, encompassing not only the usual components traditionally found in graphical user-interfaces, but also a model of the physical world (namely, a nuclear reactor with various fault scenarios) and a cognitive model of a human operator in charge of supervising the plant. Also, several desirable properties of the interface have been expressed in MCL and verified on the LNT model using CADP. At last, we used our formal model to check conformance of execution traces generated by an industrial control room prototype provided by a partner in the project.

In 2015, we finalized our approach to formally verifying safety critical interactive systems provided with plastic user interfaces, either using equivalence checking (to check whether different versions of user interfaces present the same interaction capabilities and appearance) or model checking (to check a set of properties over a model of the system). The results have been published in international conferences [26], [27] and journals [17], and in Raquel Oliveira's PhD thesis [11].

### 6.5.7. *Fault-Tolerant Routing for Network-on-Chip Architectures*

**Participant:**  Wendelin Serwe.

Fault-tolerant architectures provide adaptivity for on-chip communications, but also increase the complexity of the design, so that formal verification techniques are needed to check their correctness. In collaboration with Chris Myers and Zhen Zhang (University of Utah, USA), we studied an extension of the link-fault tolerant Network-on-Chip (NoC) architecture introduced by Wu *et al*  [64] that supports multiflit wormhole routing. A major difference with similar architectures existing in the literature is that the considered routing algorithm is not statically proven free of deadlocks, but rather implements deadlock avoidance (by dynamically detecting possible deadlock situations and avoiding them by dropping packets).

In 2015, we detected a potential livelock in the previously developed formal LNT model  [65]. The correction of this problem led to an improved routing algorithm, for which the state space for 2x2 NoCs could be generated compositionally. We also experimented with the analysis of larger configurations on Grid'5000, but even a 2x3 NoC is still too large, so that compositional state space generation fails with an intermediate state space of several billions of states. This work led to a publication accepted in an international journal [18] and a PhD thesis  [66].

### 6.5.7.1. Other Case Studies

The demo examples of CADP, which have been progressively accumulated since the origins of the toolbox, are a showcase for the multiple capabilities of CADP, as well as a test bed to assess the new features of the toolbox. In 2015, the effort to maintain and enhance these demos has been pursued. The progressive migration to LNT has continued, by translating five demos (16, 21, 22, 36, and 38) from LOTOS to LNT. A new demo 05 (airplane-ground communication protocol) has been added. The code of many demos was updated to use the latest features of LNT, such as "in var" parameters and "assert" statements. Demos 14 and 16 have been greatly simplified by inlining MCL and XTL temporal logic formulas in SVL scripts, using the "property", "check", and "|=" statements recently added to SVL. Nine demos (02, 08, 17, 20, 27, 28, 31, 33, and 36) have been simplified by using the new possibility to pass value parameters to LOTOS and LNT processes directly in SVL scripts. XTL formulas have been shortened in demos 23 and 27. The illustration of the EXEC/CAESAR framework in demo 38 has been integrated as a property into the SVL script. Finally, demo 38 led to a publication in an international workshop [29].

<p style="text-align:center"><span style="color:red">**CORSE Team**</span></p>

# 6. New Results

## 6.1. An interval constrained memory allocator for the Givy GAS runtime

**Participants:**  François Gindraud, Fabrice Rastello, Albert Cohen [ENS Ulm], Francois Broquedis.

This work presents a memory allocator for a global address space (GAS) execution environment targeting manycore architectures with distributed memory. Among the family of Multi Processor System on Chip (MPSoC), these devices are composed of multiple nodes linked by an on-chip network; most nodes have multiple processors sharing a small local memory. An MPSoC has an excellent performance-per-Watt ratio, but it is hard to program due to multilevel parallelism, explicit resource and memory management, and hardware constraints (limited memory, network topology).

Practical programming frameworks let the programmer in charge of the hard, target-specific work (e.g., threads or node-local OpenMP plus explicit communications). Automatic, more abstract frameworks exist for specific (scientific) applications, but they target big systems and do not model the hardware constraints of MPSoC. Givy is a runtime system to execute dynamic task graphs on MPSoC. It has a focus on supporting irregular applications, and uses data-flow semantics to coordinate dynamic task scheduling and data transfer. To simplify the programmer's view of memory, both runtime and program data objects live in a GAS. To avoid address collisions when objects are dynamically allocated, and to maintain the consistency of these addresses across explicit data transfers and virtual memory remapping, a GAS-aware memory allocator is required. The allocator proposed in this work has the following properties: (1) it is free of inter-node synchronizations; (2) it is well suited for small memory systems; (3) its performances match that of existing state-of-the-art allocators.

## 6.2. On Characterizing the Data Access Complexity (IO) of Programs and Using it for Architectural Design Exploration

**Participants:**  Venmugil Elango [OSU], Naser Sedaghati [OSU], Fabrice Rastello, Louis-Noël Pouchet [UCLA], J. Ramanujam [LSU], Radu Teodorescu [OSU], P. Sadayappan [OSU].

Technology trends will cause data movement to account for the majority of energy expenditure and execution time on emerging computers. Therefore, computational complexity will no longer be a sufficient metric for comparing algorithms, and a fundamental characterization of data access complexity will be increasingly important. The problem of developing lower bounds for data access complexity has been modeled using the formalism of Hong & Kung's red/blue pebble game for computational directed acyclic graphs (CDAGs). However, previously developed approaches to lower bounds analysis for the red/blue pebble game are very limited in effectiveness when applied to CDAGs of real programs, with computations comprised of multiple sub-computations with differing DAG structure. We address this problem by developing an approach for effectively composing lower bounds based on graph decomposition. We also develop a static analysis algorithm to derive the asymptotic data-access lower bounds of programs, as a function of the problem size and cache size.

The roofline model is a popular approach to "bounds and bottleneck" performance analysis. It focuses on the limits to performance of processors because of limited bandwidth to off-chip memory. It models upper bounds on performance as a function of operational intensity, the ratio of computational operations per byte of data moved from/to memory. While operational intensity can be directly measured for a specific implementation of an algorithm on a particular target platform, it is of interest to obtain broader insights on bottlenecks, where various semantically equivalent implementations of an algorithm are considered, along with analysis for variations in architectural parameters. This is currently very cumbersome and requires performance modeling and analysis of many variants.

We alleviate this problem by using the roofline model in conjunction with upper bounds on the operational intensity of computations as a function of cache capacity, derived using lower bounds on data movement. This enables bottleneck analysis that holds across all dependence-preserving semantically equivalent implementations of an algorithm. We demonstrate the utility of the approach in in assessing fundamental limits to performance and energy efficiency for several benchmark algorithms across a design space of architectural variations.

This work is the fruit of the collaboration 8.4  with OSU. The first contribution (static analysis for lower bound) will be presented at ACM POPL'15 [10]. The second contribution (architectural exploration) is to be published at ACM TACO'15 [3].

## 6.3. A Tiling Perspective for Register Optimization

**Participants:**  Duco Van Amstel, Lukasz Domagala, P. Sadayappan [OSU], Fabrice Rastello.

Register allocation is a much studied problem. A particularly important context for optimizing register allocation is within loops, since a significant fraction of the execution time of programs is often inside loop code. A variety of algorithms have been proposed in the past for register allocation, but the complexity of the problem has resulted in a decoupling of several important aspects, including loop unrolling, register promotion, and instruction reordering.

In this work, we develop an approach to register allocation and promotion in a unified optimization framework that simultaneously considers the impact of loop unrolling and instruction scheduling. This is done via a novel instruction tiling approach where instructions within a loop are represented along one dimension and innermost loop iterations along the other dimension. By exploiting the regularity along the loop dimension, and imposing essential dependence based constraints on intra-tile execution order, the problem of optimizing register pressure is cast in a constraint programming formalism. Experimental results are provided from thousands of innermost loops extracted from the SPEC benchmarks, demonstrating improvements over the current state-of-the-art.

This work is the fruit of both the collaboration 8.4  with OSU and with Kalray 7.1  7.2 .

## 6.4. Hybrid Data Dependence Analysis for Loop Transformations

**Participants:**  Diogo Nunes Sampaio, Alain Ketterlin, Fabrice Rastello, Fernando Pereira, Alexandros Labrineas, Péricles Alves, Fabian Gruber.

Loop optimizations such as tiling, vectorization, or parallel task extraction are extremely important to achieve high performance. All such transformations rely on accurate memory dependence information to assess their validity. There are many practical situations, though, where dependence analysis fails to provide precise enough information. In this common scenario, the compiler will conservatively choose not to do any transformation. This happens in particular with low-level IRs (which are more and more common to address performance portability), but also in legacy code with pointers (e.g. C), linearized arrays, etc.

This work addresses the important problem of may-dependence disambiguation through the angle of a combination of static and dynamic analyses (sometimes called a hybrid analysis), similarly to what is already implemented in mainstream compilers, such as GCC, for auto-vectorization. This technique consists of adding a run-time test to disambiguate may-dependencies which static dependence analysis was not able to rule out. We propose two contributions to address this important problem.

The first approach proposes hybrid may-alias disambiguation. It combines two approaches: one that statically computes a symbolic expression of the interval of memory values a pointer may point to and uses dynamic overlap tests on these intervals to prove non-aliasing for each pair of pointers; another that hooks the memory allocator to find the base-pointer of a pointer and thus determine dynamically if a pointer pair belongs to two different allocations (and is thus disjoint) or not. We have applied these ideas on Polly-LLVM, a loop optimizer built on top of the LLVM compilation infrastructure. Our experiments indicate that our method is precise, effective and useful: we can disambiguate every pair of pointer in the loop intensive Polybench benchmark suite. The result of this precision is code quality: the binaries that we generate are 10% faster than those that Polly-LLVM produces without our optimization, at the -O3 optimization level of LLVM.

The second technique extends the non-overlapping intervals approach to may-dependence disambiguation. For this purpose, a powerful quantifier elimination scheme on multivariate-polynomials over integers has been developed. The quality of the presented scheme is important to make this approach realistic. In particular it must be precise (the integer aspect makes this problem very challenging), so that the test succeeds in practical cases, and must lead to negligible overhead. We evaluate preciseness and overhead on a set of 30+ benchmarks using complex loop transformations including loop fusion, skewing, and tiling.

This work is the fruit of the collaboration with UFMG 8.4 , Kalray 7.1 7.2 , STMicroelectronics 7.2 , and with EPI CAMUS in the context of IPL Multicore 8.2 . The first contribution has been presented at ACM OOPSLA'15 [19]. The second has been submitted to PLDI'16.

## 6.5. Power Efficiency and Computing Performance

**Participants:** Emilio Francesquini [UNICAMP, Campinas, Brazil], Edson Luiz Padoin [PhD: UFRGS and UNIJUI, Brazil], Marcio Castro [UFSC, Florianapolis, Brazil], Pedro Penna [PUC Minas, Belo Horizonte, Brazil], Henrique Cota de Freitas [PUC Minas, Belo Horizonte, Brazil], Fabrice Dupros [BRGM, Orléans, France], Philippe Navaux [UFRGS, Porto Alegre, Brazil], Jean François Méhaut.

Until the last decade, performance of HPC architectures has been almost exclusively quantified by their processing power. However, energy efficiency is being recently considered as important as raw performance and has become a critical aspect to the development of scalable systems. These strict energy constraints guided the development of a new class of so-called light-weight manycore processors. This study evaluates the computing and energy performance of two well-known irregular NP-hard problems – the Traveling-Salesman Problem (TSP) and K-Means clustering – and a numerical seismic wave propagation simulation kernel – Ondes3D – on multicore, NUMA, and manycore platforms. First, we concentrate on the nontrivial task of adapting these applications to a manycore, specifically the Kalray/MPPA-256 manycore processor. Then, we analyze their performance and energy consumption on those different machines. Our results show that applications able to fully use the resources of a manycore can have better performance and may consume from $3.8 \times$ to $13 \times$ less energy when compared to low-power and general-purpose multicore processors, respectively.

This work is the fruit of collaborations with Brazil and several universities (UFRGS, UFSC, UNICAMP, PUC Minas, USP). This work has been published in the journal of parallel and distributed computing [6] and in the journal of IET Computers and Digtal Techniques [7]. This work was also part of several international projects (LICIA, CNPq/Inria HOSCAR project, Exase).

Emilio Francesquini and Marcio Castro are also former PhD students of University Grenoble Alpes (UGA) and the LIG Laboratory.

## 6.6. Modeling and Simulating of Dynamic Task-Based Runtime Systems

**Participants:** Luka Stanisic [PhD, Inria, Mescal], Samuel Thibault [Univ. Bordeaux, Inria, Storm], Brice Videau, Arnaud Legrand [CNRS, Inria, Mescal], Jean François Méhaut.

Multi-core architectures comprising several GPUs have become mainstream in the field of High-Performance Computing. However, obtaining the maximum performance of such heterogeneous machines is challenging as it requires to carefully offload computations and manage data movements between the different processing units. The most promising and successful approaches so far build on task-based runtimes that abstract the machine and rely on opportunistic scheduling algorithms. As a consequence, the problem gets shifted to choosing the task granularity, task graph structure, and optimizing the scheduling strategies. Trying different combinations of these different alternatives is also itself a challenge. Indeed, getting accurate measurements requires reserving the target system for the whole duration of experiments. Furthermore, observations are limited to the few available systems at hand and may be difficult to generalize. In this work, we show how we crafted a coarse-grain hybrid simulation/emulation of StarPU, a dynamic runtime for hybrid architectures, over SimGrid, a versatile simulator for distributed systems. This approach allows to obtain performance predictions of classical dense linear algebra kernels accurate within a few percents and in a matter of seconds, which allows both runtime and application designers to quickly decide which optimization to enable or whether it is worth

investing in higher-end GPUs or not. Additionally, it allows to conduct robust and extensive scheduling studies in a controlled environment whose characteristics are very close to real platforms while having reproducible behavior.

This work is part of the Luka Stanisic's thesis. Luka stanisic was coadvised by Arnaud Legrand, Brice Videau and Jean-François Méhaut. This thesis was defended in November 2015. Luka Stanisic currently holds a postdoc position at Inria Bordeaux in the Storm and HiePacs teams. This work was published in the CCPE journal [9].

## 6.7. Fast and Accurate Simulation of Multithreaded Sparse Linear Algebra Solvers

**Participants:** Luka Stanisic [PhD, Inria, Mescal], Arnaud Legrand [CNRS, Inria, Mescal], Emmanuel Agullo [Inria, HiePacs], Alfredo Buttari [CNRS, IRIT, Toulouse], Florent Lopez [CNRS, IRIT, Toulouse], Brice Videau.

The ever growing complexity and scale of parallel architectures imposes to rewrite classical monolithic HPC scientific applications and libraries as their portability and performance optimization only comes at a prohibitive cost. There is thus a recent and general trend in using instead a modular approach where numerical algorithms are written at a high level independently of the hardware architecture as Directed Acyclic Graphs (DAG) of tasks. A task-based runtime system then dynamically schedules the resulting DAG on the different computing resources, automatically taking care of data movement and taking into account the possible speed heterogeneity and variability. Evaluating the performance of such complex and dynamic systems is extremely challenging especially for irregular codes. In this article, we explain how we crafted a faithful simulation, both in terms of performance and memory usage, of the behavior of qr_mumps , a fully-featured sparse linear algebra library, on multi-core architectures. In our approach, the target high-end machines are calibrated only once to derive sound performance models. These models can then be used at will to quickly predict and study in a reproducible way the performance of such irregular and resource-demanding applications using solely a commodity laptop.

This work is part of the Luka Stanisic's thesis. Luka stanisic was coadvised by Arnaud Legrand, Brice Videau and Jean-François Méhaut. This thesis was defended in November 2015. Luka Stanisic currently holds a postdoc position at Inria Bordeaux in the Storm and HiePacs teams. This work was published in the ICPADS'2015 conference [18].

## 6.8. OpenMP Loop Scheduling

**Participants:** Pedro Penna [Master, PUC Minas, UFSC], Marcio Castro [Professor, UFSC], Henrique Cota de Freitas [Professor, PUC Minas], Francois Broquedis, Jean François Méhaut.

In High Performance Computing, the application's workload must be well balanced among the threads to achieve better performance. In this work, we propose a methodology that enables the design and exploration of new loop scheduling strategies. In this methodology, a simulator is used to evaluate the most relevant existing scheduling strategies, and a genetic algorithm is employed to explore the solution space of the problem itself. The proposed methodology allowed us to design a new loop scheduling strategy, which showed to be up to 32.3x better than the existing policies in terms of load balance.

## 6.9. BOAST: a Metaprogramming framework for computing kernels

**Participants:** Brice Videau [Postdoc CNRS, Mont-Blanc], Kevin Pouget [UJF, Nano2017], Luigi Genovese [Researcher, CEA INAC], Thierry Deutsch [Researcher, CEA INAC], Anthony Leonard [CNRS, Polytech Grenoble, Internship, from May 2015 until Aug 2015], Frederic Desprez, Jean François Méhaut.

Porting and tuning HPC applications to new platforms is tedious and costly in terms of human resources. Nonetheless, it is a very important aspect of the Mont-Blanc project. Indeed, for the Mont-Blanc project, more than ten applications were selected to be ported and optimized for the prototype platform.

Unfortunately, portability efforts are often lost when migrating to a new architecture. Worse, code may lose maintainability because several versions of some functionalities coexist, usually with a lot of duplication. Thus productivity of porting and tuning efforts is low as a huge fraction of those developments are never used after the platform they were intended for is decommissioned. Genericity of HPC codes is often limited. One of the reason is that producing generic code in Fortran 90/95 is difficult as the language does not really fit for it. Sometimes, adding genericity degrades performance as optimization opportunities that come from over-specification are lost. Functionality of HPC codes is tied to the previous point. Without genericity, adding new functionalities can be quite costly.

BOAST is a metaprogramming framework to produce portable and efficient computing kernels for HPC application. BOAST offers an embedded domain specific language to describe the kernels and their possible optimization. BOAST also supplies a complete runtime to compile, run, benchmark, and check the validity of the generated kernels. BOAST is being used in two flagship HPC applications BigDFT and SPECFEM3D, to improve performance portability of those codes.

BOAST is developped in the context of Mont-Blanc projects. It will be also used in the C2S@Exa IPL and the H2020/HPC4E project.

## 6.10. Performance comparison between Java and JNI for optimal implementation of computational micro-kernels

**Participants:**  Nassim Halli [PhD student, CIFRE Aselta Nanographics], Henri-Pierre Charles [CEA LIST, CRI PILSI], Jean François Méhaut.

General purpose CPUs used in high performance computing (HPC) support a vector instruction set and an out-of-order engine dedicated to increase the instruction level parallelism. Hence, related optimizations are currently critical to improve the performance of applications requiring numerical computation. Moreover, the use of a Java run-time environment such as the HotSpot Java Virtual Machine (JVM) in high performance computing is a promising alternative. It benefits from its programming flexibility, productivity and the performance is ensured by the Just-In-Time (JIT) compiler. Though, the JIT compiler suffers from two main drawbacks. First, the JIT is a black box for developers. We have no control over the generated code nor any feedback from its optimization phases like vectorization. Secondly, the time constraint narrows down the degree of optimization compared to static compilers like GCC or LLVM. So, it is compelling to use statically compiled code since it benefits from additional optimization reducing performance bottlenecks. Java enables to call native code from dynamic libraries through the Java Native Interface (JNI). Nevertheless, JNI methods are not inlined and require an additional cost to be invoked compared to Java ones. Therefore, to benefit from better static optimization, this call overhead must be leveraged by the amount of computation performed at each JNI invocation. In this work we tackle this problem and we propose to do this analysis for a set of micro-kernels. Our goal is to select the most efficient implementation considering the amount of computation defined by the calling context. We also investigate the impact on performance of several different optimization schemes which are vectorization, out-of-order optimization, data alignment, method inlining and the use of native memory for JNI methods.

This work was presented in the ADAPT'2015 workshop. It's alsop part of the Nassim Halli's thesis.

## 6.11. Reducing trace size in multimedia applications endurance tests

**Participants:**  Serge Emteu [PhD ST Microelectronics, LIG/Slide, CORSE], Miguel Santana [ST Microelectrnics], Alexadre Termier [Prof. Univ. Rennes I, IRISA/Inria/Dream], René Quiniou [CR Inria, IRISA/Inria/Dream], Brice Videau [Postdoc CNRS, Inria/Corse], Jean François Méhaut.

The consumer electronics market is dominated by embedded systems due to their ever-increasing processing power and the large number of functionnalities they offer. To provide such features, architectures of embedded systems have increased in complexity : they rely on several heterogeneous processing units, and allow concurrent tasks execution. This complexity degrades the programmability of embedded system architectures and makes application execution difficult to understand on such systems. The most used approach for analyzing application execution on embedded systems consists in capturing execution traces (event sequences, such as system call invocations or context switch, generated during application execution). This approach is used in application testing, debugging or profiling. However in some use cases, execution traces generated can be very large, up to several hundreds of gigabytes. For example endurance tests, which are tests consisting in tracing execution of an application on an embedded system during long periods, from several hours to several days. Current tools and methods for analyzing execution traces are not designed to handle such amounts of data.

We propose an approach for monitoring an application execution by analyzing traces on the fly in order to reduce the volume of recorded traces. Our approach is based on features of multimedia applications which contribute the most to the success of popular devices such as set-top boxes or smartphones. This approach consists in identifying automatically the suspicious periods of an application execution in order to record only the parts of traces which correspond to these periods. The proposed approach consists of two steps : a learning step which discovers regular behaviors of an application from its execution trace, and an anomaly detection step which identifies behaviors deviating from the regular ones.

The many experiments, performed on synthetic and real-life datasets, show that our approach reduces the trace size by an order of magnitude while maintaining a good performance in detecting suspicious behaviors.

This work was presented at the DATE conference in Grenoble. It was also part of the Serge Emteu's thesis wth ST Microelectronics.

## 6.12. Data Mining Approach to Temporal Debugging of Embedded Streaming Applications

**Participants:** Oleg Iegorov [PhD ST Microelectronics, LIG/Slide, CORSE], Miguel Santana [ST Microelectrnics], Alexadre Termier [Prof. Univ. Rennes I, IRISA/Inria/Dream], Vincent Leroy [Associate Professor UJF, LIG/Slide], Jean François Méhaut.

One of the greatest challenges in the embedded systems area is to empower software developers with tools that speed up the debugging of QoS properties in applications. Typical streaming applications, such as multimedia (audio/video) decoding, fulfill the QoS properties by respecting the realtime deadlines. A perfectly functional application, when missing these deadlines, may lead to cracks in the sound or perceptible artifacts in the image.

We start from the premise that most of the streaming applications that run on embedded systems can be expressed under a dataflow model of computation, where the application is represented as a directed graph of the data flowing through computational units called actors. It has been shown that in order to meet real-time constraints the actors should be scheduled in a periodic manner. We exploit this property to propose SATM – a novel approach based on data mining techniques that automatically analyzes execution traces of streaming applications, and discovers significant breaks in the periodicity of actors, as well as potential causes of these breaks. We show on a real use case that our debugging approach can uncover important defects and pinpoint their location to the application developer.

This work was presented at the EMSOFT conference in Amsterdam. It was also part of the Oleg Iegorov's thesis wth ST Microelectronics.

## 6.13. Tiling Bitwise Computations Using Look-up Instructions

**Participants:** Florent Bouchez - Tichadou, Cyril Six [Inria, Internship, from Feb 2015 until Jun 2015].

The BWLU is an instruction of a Very Long Instruction Word processor (VLIW) that performs a series of bit-independent computations in only one step through the use of a "look-up table" (LUT). The Bit-Wise Look-Up table instruction (BWLU) takes as input four registers as well as a 32-bit integer (the "table"), and is able to output two bit-independent computations based on the input registers into two output registers.

The goal is to make the best use possible of this instruction by replacing during compilation as much as possible groups of bitwise computation using BWLUs so as to reduce the number of instructions required to perform a computation. The problem is represented by a data-flow graph representing a computation, and the goal is use BLWUs as tiles to "match" groups of bitwise instruction.

We proved the problem NP-complete for a general data-flow graph, so it is not practical to try to find the optimal solution.

It is easy to devise a greedy algorithm that will produce a solution, but we wanted a way to check whether the solutions found where far from the optimal. An optimal algorithm is of course exponential in the size of the input graph, however, we devised a complete space exploration algorithm based on dynamic programming that manages to find the optimal solution for data-flow graphs with small width or height.

<p style="text-align:center"><span style="color:red">**DICE Team**</span></p>

# 6. New Results

## 6.1. The economy of intermediation and the anthropocene

Better understanding the economy, in a broad sense, of intermediation as it is performed by online platforms, is one of the major goals of the team. The paper [12] published in 1024, introduces the topics of algorihtmic intermediation and its social impact to a large audience.

Two contemporary revolutions are shaking the world. On one side, the digital revolution, which seems to introduce to a new economic era, allowing more sharing, and according to some authors the end of capitalism. On the other hand, the challenges of the preservation of our planet, and the limitation of resources that we are now facing. Clearly, there is an expectation that digital means will help face the challenges of the planet. In [14], we go one step further and analyse the possible relationship between the two phenomena, by drawing comparisons with biology where stress on ressources can lead to a horizontalisation of the species, much like what happens with digital technologies and intermediation platforms.

This later work is made in the framework of the study of the anthropocene, for which we are involved in the organisation of a workshop in the framework program of the HKW in Berlin on the technosphere

- URL: http://www.hkw.de/en/programm/projekte/2015/curriculum_campus_technosphere/campus_the_technosphere_issue.php

## 6.2. Geopolitics of intermediation platforms

Our study of the geopolitics of intermediation aims at grasping the balance of power between platforms, as well as between states - in their relation to platforms - and between platforms and states. We have designed coarse metrics [1] which capture the importance of a platform and the importance of a country in the digital landscape.

Our study focuses on the top 25 websites in a hundred countries. We emphasize the weight of intermedations platforms on the web. We also underline the imbalance between two digital powers - the United States and China - and the rest of the world. Indeed, most platforms belong to these two countries. We have extented our study to a deep analysis of the Asian case [8]. We develop our analysis in an interdisciplinary context as we collaborate with cartographers and economists. Two outcomes of our work are especially notable:

- We produce a set of maps and data vizualisations to illustrate the intermediation economy [11].
- We highlight the determinants of the imbalance in the intermediation landscape. National policies and incentives are of primer importance. The digita landscapes of Korea and Taiwan for instance, show that countries can still play a main role in their domestic web [8].

## 6.3. Public administration and intermediation platforms

Building on the sucess of platforms such as Uber and the analyses of their externalities, we study the potential role of platforms in public administration. Indeed, cities such as Boston exhibit the interest of a collaboration between administrations and platforms in city planning and maintenance. We also address the role of platforms at a wider level as we study cases such as the settlement of the right to be forgotten in Europe. Our work benefits from the collaboration with administrations, such as Lyon metropole and social scientists. In particular, we have designed three possible scenarii of collobation between platforms and institutions:

- Coexistence: platforms and institutions ignore themselves;
- Conflict: the services developped by platforms conflict with existing policies and institutional practices;
- Partenership: platforms and institutions partner around the development and promotion of services.

A working group has been established on digital sovereignty with CLTC, Center for Long Term Cybersecurity at UC Berkeley, Chaire Castex at Ecole Militaire, and Dice. This working group aims at getting a better understanding of the concept as well as the discrepancy of perception on both sides of the Atlantic. A first seminar was organised in Les Houches in december 2015.

This is work in progress with both academic and public administration actors.

## 6.4. Architecture design for intermediation platforms

Dice team designs software architectures for intermediation platforms. C3PO and BitBallot targets spontaneous and ephemeral social networks whereas Jumplyn focuses on pure central based system. All these architectures share a common JavaScript layout both at the client and the server sides. In the research context we validate state-of-the art technologies promoted by web leaders such as Google AngularJS, Facebook ReactJS and many others such as Netflix, Wallmart, or the Linux foundation for node.js. The Web environment raises many big issues since all equipments are basically connected to the Internet and the balance between end-user equipment cost and processing power is still a work in progress. Our main research track in such context is to find proper software toolkits hiding Web complexity. We mainly focus on time jitter, cornerstone of Web development, since it implies both end-user and network TCP indecisions. Due to this jitter combination the Web programming model has mutated toward the promises paradigm. It is a complex event based development model provided without external API help. It handles future execution whether successful or not, in a time jittered context. AngularJS, ReactJS, CoffeeScript, NodeJS, MongoDB, ElasticSearch are all time jitter compliant technologies designed for the Web constrains and revolutionising the way we build intermediation platforms.

C3PO explores network transport laziness with the use of a DTN that imposes a larger jitter than classical TCP/UDP. We build a JavaScript mockup [5] that uses a Java based DTN that stores, carries and forwards message from source to destination. C3PO is a software framework extending AngularJS through plugins, without central server, even during deployment phases. We use the dynamic nature of JavaScript to build application on the fly from network messages containing the application description. Our C3PO architecture enables us to build ephemeral and spontaneous social network, on demand and in a matter of days.

Our joint work with Worldline explores the promises paradigm model to enable automation extraction of independent micro-service. These micro-services called *fluxion* [9], from the contraction of flow and functions, may be dynamically and transparently moved over a cluster of servers. Our novelty resides in the fact that the original code is not redesigned for the cluster architecture. *Fluxion* are extracted from the initial code, and an equivalence is maintained between the initially *promissified* code and the *fluxionized* one. Code has two facets, a promise one, used to express software services and a *fluxion* one, used to express software bottlenecks.

Eventually our work with Jumplyn explores complex centralised social network. We want to design a software system to later support our technical research hot topics. The target theme is a software platform that helps students handle their projects. University depends more and more on external resources to teach students. Theses resources are both known by students and their teachers, and the pace and range of explored technologies leads to difficulties in teaching state-of-the-art subjects. The more dedicated a professor needs to be in his research activity, the more broad knowledge he has to teach. For instance 20 years ago one could cope software development teaching with one or two programming languages. Nowadays, a single code involves more then four programming languages to be fully understood. This technology spreading issue stands still in many teaching domains, since past technologies are still actives and future one are promising. We build Jumplyn to cope with this unbalanced game. To help student improving their project and avoid working with obsolete technologies, and to help teacher face the universal and inexpensive availability of knowledge. Jumplyn is a complex JavaScript development stack that collects resources for improving student work and providing services to help them from day to day activities. The current stack integrates the following technologies : MaterialDesign, AngularJS, CoffeeScript, NodeJs, MongoDb, ElasticSearch. Managing and developing software service above this stack is a complex research issue for a small sized development team. We do not have any publication on Jumplyn since our first goal is to build a support intermediation platform to

study classical issues such as recommendation or web crawling, scraping and indexation with our own sources of raw data.

<div align="center">

**PRIVATICS Project-Team**

</div>

# 6. New Results

## 6.1. Surveillance

**Participants:** Claude Castelluccia, Javier Parra Arnau.

In recent times, we are witnessing an increasing concern by governments and intelligence agencies to deploy mass-surveillance systems that help them fight terrorism. In [40], we conduct a formal analysis of the overall cost of such surveillance systems. Our analysis starts with a fairly-known result in statistics, namely, the false-positive paradox. We propose a quantitative measure of the total cost of a monitoring program, and study a detection system that is designed to minimize it, subject to a constraint in the number of terrorists the agency wishes to capture. In the absence of real, accurate behavioral models, we perform our analysis on the basis of several simple but insightful examples. With these examples, we illustrate the different parameters involved in the design of the detection system, and provide some indicative and representative figures of the cost of the monitoring program.

## 6.2. Security or privacy ?

**Participants:** Amrit Kumar, Cédric Lauradoux.

Security softwares such as anti-viruses, IDS or filters help Internet users to protect their privacy from hackers. The cost of this protection is that the users privacy is stripped away by the companies providing these security solutions. Currently, Internet users can choose between no security with the risk of being hacked or use security softwares and lose all personal data to security companies. As a example of this dilemma, we analyze the solution proposed by Google for Safe Browsing in [29] and shows that their privacy policies do not match the reality: Google can perform tracking.

## 6.3. Users characterization

**Participants:** Jagdish Achara, Gergely Acs, Claude Castelluccia.

Prior works have shown that the list of apps installed by a user reveal a lot about user interests and behavior. These works rely on the semantics of the installed apps and show that various user traits could be learnt automatically using off-the-shelf machine-learning techniques. In this work, we focus on the re-identifiability issue and thoroughly study the unicity of smartphone apps on a dataset containing 54,893 Android users collected over a period of 7 months. Our study finds that any 4 apps installed by a user are enough (more than 95% times) for the re-identification of the user in our dataset. As the complete list of installed apps is unique for 99% of the users in our dataset, it can be easily used to track/profile the users by a service such as Twitter that has access to the whole list of installed apps of users. As our analyzed dataset is small as compared to the total population of Android users, we also study how unicity would vary with larger datasets. This work emphasizes the need of better privacy guards against collection, use and release of the list of installed apps.

## 6.4. Data anonymization

**Participants:** Claude Castelluccia, Gergely Acs.

Set-valued dataset contains different types of items/values per individual, for example, visited locations, purchased goods, watched movies, or search queries. As it is relatively easy to re-identify individuals in such datasets, their release poses significant privacy threats. Hence, organizations aiming to share such datasets must adhere to personal data regulations. In order to get rid of these regulations and also to benefit from sharing, these datasets should be anonymized before their release. In this paper, we revisit the problem of anonymizing set-valued data. We argue that anonymization techniques targeting traditional $k^m$-anonymity model, which limits the adversarial background knowledge to at most $m$ items per individual, are impractical for large real-world datasets. Hence, we propose in [25] a probabilistic relaxation of $k^m$-anonymity and present an anonymization technique to achieve it. This relaxation also improves the utility of the anonymized data. We also demonstrate the effectiveness of our scalable anonymization technique on a real-world location dataset consisting of more than 4 million subscribers of a large European telecom operator. We believe that our technique can be very appealing for practitioners willing to share such large datasets.

## 6.5. Wi-Fi and privacy

**Participants:** Jagdish Achara, Mathieu Cunche, Vincent Roca, Celestin Matte.

- **Geolocation spoofing attack** Our work at WiSec 2015 [17] shows how it is possible to manipulate the geolocation information of a single device and how to exploit this information as a side channel to identify the owner of the device on geottaged platforms such as social networks.

- **Extraction of sementical information from Wi-Fi network identifiers** Methods based on text similarity metrics can be used to infer user's interests based on the list of their preferred networks. We present in [23] a method identifying the physical entity (shop, restaurant, company ...) associated to Wi-Fi networks identifiers (SSID).

## 6.6. Formal and legal issues of privacy

**Participants:** Thibaud Antignac, Daniel Le Metayer.

- **Privacy by design** Privacy by design will become a legal obligation in the European Community when the Data Protection Regulation eventually gets adopted. However, taking into account privacy requirements in the design of a system is a challenging task. We have proposed an approach based on the specification of privacy architectures and illustrated our formal framework through several case studies. In collaboration with Morpho, we have applied it in the context of biometrics systems. The choice of particular techniques and the role of the components (central server, secure module, terminal, smart card, etc.) in the architecture have a strong impact on the privacy guarantees provided by a biometric system. However, existing proposals were made on a case by case basis, which makes it difficult to compare them and to provide a rationale for the choice of specific options. We have shown that a general framework for the definition of privacy architectures can be used to specify these options and to reason about them in a formal way. In 2015 the results on biometrics were presented at the conferences FM2015 [16] and ISC 2015 [15] (best paper award) and the general approach itself has led to Thibaud Antignac's PhD defense.

- **Verification of privacy properties**

  Electric vehicles are an up-and-coming technology that provides significant environmental benefits. A major challenge of these vehicles is their somewhat limited range, requiring the deployment of many charging stations. To effectively deliver electricity to vehicles and guarantee payment, a protocol was developed as part of the ISO 15118 standardization effort. A privacy-preserving variant of this protocol, POPCORN, has been proposed in recent work, claiming to provide significant privacy for the user, while maintaining functionality. We have proposed an approach for the verification of privacy properties of the protocol. We have provided a formal model of the expected privacy properties in the applied Pi-Calculus and used ProVerif to check them. We have identified weaknesses in the protocol in [11] and suggest improvements to address them.

- **Control over personal data**

More than ever the notion of control plays a pivotal and pervasive role in the discourses of privacy and data protection. Privacy scholarship and regulators propose to increase individual control over personal information as the ultimate prescriptive solution to tackle the issues raised by emergent data processing technologies. Conceived as the claim of individuals to determine for themselves when, how, and to what extent information about them is communicated to others, the notion of control is not new. It is often considered as the unique means of empowerment of the data subject. The mechanisms of this empowerment remain however surprisingly vague and understudied. What does it really mean to be in control of one's data in the context of contemporary socio-technical environments and practices? What are the characteristics, purposes and potential limits of such control and how can we guarantee data subjects effective control over their own data? We have carried out an interdisciplinary review of the concept of control to explore such questions in the fields of law and computer science and suggested conditions for the effective application of this concept (see [5]).

- **Accountability** The use of body-worn cameras by police forces around the world is spreading quickly. The resulting mobile and ubiquitous surveillance is often marketed as an instrument for accountability and an effective way of reducing violence. It also involves remarkable potential for intrusion into the privacy of both individuals and police agents. We have studied in [4] the deployment of police body-worn cameras in five countries, investigated their suitability as an accountability tool given the associated privacy threats, and analyzed the societal impact of their deployment as well as the risk of function creep.

## 6.7. Buidling blocks

**Participant:** Marine Minier.

- **Symmetric cryptography** During this year, a fruitful work in collaboration with Céline Blondeau from University of AAlto has appeared in FSE 2015 [8] concerning the equivalence between the key recovery parts of the three attacks (Zero-Correlation, impossible and integral) using the matrix method.

   With Thierry Berger, Julien Francq and also Gaël Thomas, we have proposed 2 new lightweight block ciphers : Lilliput and CubeCipher.

   Concerning symmetric cryptography, we obtain some results in both sides: on the one hand, we provide 2 new families of lightweight block ciphers: CubeCipher familiy and Lilliput; on the other hand, we work on the matrix method to simplify the representation of some attacks such as zero-correlation attack, impossible and integral attacks.

   We also published the extended version of our Secrypt 2013 paper in the journal Security and Communication Networks [2] concerning the performances on a dedicated platform.

- **Passwords Cracking** Passwords are widely used for user authentication, and will likely remain in use in the foreseeable future, despite several weaknesses. One important weakness is that human-generated passwords are far from being random, which makes them susceptible to guessing attacks. Understanding the adversaries' capabilities for guessing attacks is a fundamental necessity for estimating their impact and advising countermeasures. We develop OMEN [9], a new Markov model-based password cracker that extends ideas proposed by Narayanan and Shmatikov (CCS 2005). The main novelty of our tool is that it generates password candidates according to their occurrence probabilities, i.e., it outputs most likely passwords first. As shown by our extensive experiments, OMEN significantly improves guessing speed over existing proposals. In particular, we compare the performance of OMEN with the Markov mode of John the Ripper, which implements the password indexing function by Narayanan and Shmatikov. OMEN guesses more than 40% of passwords correctly with the first 90 million guesses, while JtR-Markov (for T = 1 billion) needs at least eight times as many guesses to reach the same goal, and OMEN guesses more than 80% of passwords correctly at 10 billion guesses, more than all probabilistic password crackers we compared against.

- **Time-memory trade-off** Cryptanalytic time-memory trade-offs (TMTO) are well-known tools available in any security expert toolbox. They have been used to break ciphers such as A5/1, but their efficiency to crack passwords made them even more popular in the security community. While symmetric keys are generated randomly according to a uniform distribution, passwords chosen by users are in practice far from being random, as confirmed by recent leakage of databases. Unfortunately, the technique used to build TMTOs is not appropriate to deal with non-uniform distributions. In [6], we introduce an efficient construction that consists in partitioning the search set into subsets of close densities, and a strategy to explore the TMTOs associated to the subsets based on an interleaved traversal. This approach results in a significant improvement compared to currently used TMTOs. We experimented our approach on a classical problem, namely cracking 7-character NTLM Hash passwords using an alphabet with 34 special characters, which resulted in a 16 × speedup over rainbow tables, which are considered as the most efficient variant of time-memory trade-offs.

<div align="center">

## SPADES Project-Team

</div>

# 6. New Results

## 6.1. Components and contracts

**Participants:**  Sophie Quinton, Jean-Bernard Stefani.

### 6.1.1. Multi-viewpoint contracts for the negotiation of embedded software updates

In the context of the CCC project (http://ccc-project.org/) we address the issue of change after deployment in safety-critical embedded system applications. Our goal is to substitute lab-based verification with in-field formal analysis to determine whether an update may be safely applied. This is challenging because it requires an automated process able to handle multiple viewpoints such as functional correctness, timing, etc. For this purpose, we propose an original methodology for contract-based negotiation of software updates. The use of contracts allows us to cleanly split the verification effort between the lab and the field. In addition, we show how to rely on existing viewpoint-specific methods for update negotiation. We have started validating our approach on a concrete example inspired by the automotive domain in collaboration with our German partners from TU Braunschweig.

### 6.1.2. Location Graphs

The design of configurable systems can be streamlined and made more systematic by adopting a component-based structure, as demonstrated with the FRACTAL component model [2]. However, the formal foundations for configurable component-based systems, featuring higher-order capabilities where components can be dynamically instantiated and passivated, and non-hierarchical structures where components can be contained in different composites at the same time, are still an open topic. We have recently introduced the location graph model  [88], where components are understood as graphs of locations hosting higher-order processes, and where component structures can be arbitrary graphs.

We have continued the development of the location graph model and extended it in several directions. First we have introduced basic capabilities and predicate parameters in the model to allow for different forms of architectural invariants, such as different forms of encapsulation, to be maintained even in presence of dynamic graph modifications. Second, we have started developing the premises of a refinement theory for location graphs, showing in particular how one could refine a location process into a whole graph. Finally, we have shown how to handle heterogeneous forms of composition in the same location graph, turning each location into a composition operator. This work has not yet been published.

## 6.2. Real-Time multicore programming

**Participants:**  Vagelis Bebelis, Adnan Bouakaz, Pascal Fradet, Alain Girault, Gregor Goessler, Xavier Nicollin, Jean-Bernard Stefani.

### 6.2.1. A time predictable programming language for multicores

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 [59]. 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 [92]. 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 [38].

Following our past results on the PRET-C programming language [36], we have proposed a time predictable synchronous programming language for multicores, called FOREC. It extends C with a small set of ESTEREL-like synchronous primitives to express concurrency, interaction with the environment, looping, and a synchronization barrier [93] (like the `pause` statement in ESTEREL). FOREC threads communicate with each other via shared variables, the values of which are *combined* at the end of each tick to maintain deterministic execution. FOREC is compiled into threads that are then statically scheduled for a target multicore chip. Our WCET analysis takes into account the access to the shared TDMA bus and the necessary administration for the shared variables. We achieve a very precise WCET (the over-approximation being less than 2%) thanks to a reachable space exploration of the threads' states.

Recent results have addressed the semantics, the compiler, and the experiments. In particular, we have seeked to provide several combine policies for shared variables, in a way similar as concurrent revisions [49].

This work has been conducted within the RIPPES associated team.

### 6.2.2. *Modular distribution of synchronous programs*

Synchronous programming languages describe functionally centralized systems, where every value, input, output, or function is always directly available for every operation. However, most embedded systems are nowadays composed of several computing resources. The aim of this work is to provide a language-oriented solution to describe *functionally distributed reactive systems*. This research started within the Inria large scale action SYNCHRONICS and is a joint work with Marc Pouzet (ENS, PARKAS team from Rocquencourt) and Gwenaël Delaval (UGA, CTRL-A team from Grenoble).

We are working on defining a *fully-conservative* extension of a synchronous data-flow programming language (the HEPTAGON language, inspired from LUCID SYNCHRONE [51]). The extension, by means of *annotations* adds *abstract location parameters* to functions, and *communications* of values between locations. At deployment, every abstract location is assigned an actual one; this yields an executable for each actual computing resource. Compared to the PhD of Gwenaël Delaval [56], [57], the goal here is to achieve *modular* distribution even in the presence of non-static clocks, *i.e.*, clocks defined according to the value of inputs.

By *fully-conservative*, we have three aims in mind:

1. A non-annotated (*i.e.*, centralized) program will be compiled exactly as before;
2. An annotated program eventually deployed onto only one computing location will behave exactly as its centralized couterpart;
3. The input-output semantics of a distributed program is the same as its centralized counterpart.

By *modular*, we mean that we want to compile each function of the program into a single function capable of running on any computing location. At deployment, the program of each location may be optimized (by simple Boolean-constant-propagation, dead-code and unused-variable elimination), yielding different optimized code for each computing location.

We have formalized the type-system for inferring the location of each variable and computation. In the presence of local clocks, added information is computed from the existing clock-calculus and the location-calculus, to infer necessary communication of clocks between location. The overall structure of the new compiler is defined, including new algorithms for deployment (and code optimization), achieving the three aims detailed above.

### 6.2.3. *Analysis and scheduling of parametric dataflow models*

Recent data-flow 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.

Recently, we have introduced the *Schedulable Parametric Data-Flow* (SPDF) MoC for dynamic streaming applications [62], which extends the standard dataflow model by allowing rates to be parametric, and the *Boolean Parametric Data Flow* (BPDF) MoC [42], [41] which combines integer parameters (to express dynamic rates) and boolean parameters (to express the activation and deactivation of communication channels). High dynamicity is provided by integer parameters which can change at each basic iteration and boolean parameters which can change even within the iteration. We have presented static analyses that ensure the liveness and the boundedness of BDPF graphs.

This year, we have focused on the *symbolic* analysis of parametric data-flow graphs. This work has been conducted within the RIPPES associated team.

### 6.2.4. 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.*, [90] 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 [83]. 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 [63].

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. We have been exploring two approaches to overcome this state-space explosion.

We are currently investigating an approach using mode sequences of given length as symbolic states for our abstractions. 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.

### 6.2.5. Typical Worst-Case Analysis of real-time systems

We focus on the problem of computing tight deadline miss models for real-time systems, which bound the number of potential deadline misses in a given sequence of activations of a task. In practical applications, such guarantees are often sufficient because many systems are in fact not hard real-time. Our major contribution this year is a general formulation of that problem in the context of systems where some tasks occasionally experience sporadic overload [26]. Based on this new formulation, we present an algorithm that can take into account fine-grained effects of overload at the input of different tasks when computing deadline miss bounds. We show in experiments with synthetic as well as industrial data that our algorithm produces bounds that are much tighter than in previous work, in sufficiently short time. In addition, we improve, in the preemptive case, the criterion proposed in [71] for establishing how much overload can be tolerated in a time window while still guaranteeing absence of deadline misses: our new criterion is a necessary and sufficient condition (as opposed to the sufficient condition of [71]) and therefore yields better results.

In parallel, we have developed an extension of sensitivity analysis for budgeting in the design of weakly-hard real-time systems. During design, it often happens that some parts of a task set are fully specified while other parameters, e.g. regarding recovery or monitoring tasks, will be available only much later. In such cases, sensitivity analysis can help anticipate how these missing parameters can influence the behavior of the whole system so that a resource budget can be allocated to them. It is, however, sufficient in many application contexts to budget these tasks in order to preserve weakly-hard rather than hard guarantees. We have thus developed an extension of sensitivity analysis for deriving task budgets for systems with hard and weakly-hard requirements.

We currently validate our approach on synthetic test cases and a realistic case study given by our partner Thales.

## 6.3. Language Based Fault-Tolerance

**Participants:** Dmitry Burlyaev, Pascal Fradet, Alain Girault, Yoann Geoffroy, Gregor Goessler, Jean-Bernard Stefani, Atena Abdi, Ismail Assayad.

### 6.3.1. *Fault Ascription in Concurrent Systems*

The failure of one component may entail a cascade of failures in other components; several components may also fail independently. In such cases, elucidating the exact scenario that led to the failure is a complex and tedious task that requires significant expertise.

The notion of causality *(did an event e cause an event e'?)* has been studied in many disciplines, including philosophy, logic, statistics, and law. The definitions of causality studied in these disciplines usually amount to variants of the counterfactual test "$e$ is a cause of $e'$ if both $e$ and $e'$ have occurred, and in a world that is as close as possible to the actual world but where $e$ does not occur, $e'$ does not occur either". In computer science, almost all definitions of logical causality — including the landmark definition of [70] and its derivatives — rely on a causal model that may not be known, for instance in presence of black-box components. For such systems, we have been developing a framework for blaming that helps us establish the causal relationship between component failures and system failures, given an observed system execution trace. The analysis is based on a formalization of counterfactual reasoning [7].

We have instantiated our approach to a synchronous data flow framework defined by a subset of the LUSTRE [69] language, and implemented the analysis in LoCA (see Section 5.2 ).

In [25] we have shown that we can improve precision of the analysis if (1) we can emulate execution of components instead of relying on their specifications, and (2) take into consideration input/output dependencies between components to avoid blaming components for faults induced by other components. We have demonstrated the utility of the extended analysis with a case study for a closed-loop patient-controlled analgesia system.

We have further proposed in [23] a general semantic framework for fault ascription. Our framework relies on configuration structures to handle concurrent systems, partial and distributed observations in a uniform way. It defines basic conditions for a counterfactual analysis of necessary and sufficient causes, and it presents a refined analysis that conforms to our basic conditions while avoiding various infelicities.

### 6.3.2. *Tradeoff exploration between energy consumption and execution time*

We have continued our work on multi-criteria scheduling, in two directions. First, in the context of dynamic applications that are launched and terminated on an embedded homogeneous multi-core chip, under execution time and energy consumption constraints, we have proposed a two layer adaptive scheduling method. In the first layer, each application (represented as a DAG of tasks) is scheduled statically on subsets of cores: 2 cores, 3 cores, 4 cores, and so on. For each size of these sets (2, 3, 4, ...), there may be only one topology or several topologies. For instance, for 2 or 3 cores there is only one topology (a "line"), while for 4 cores there are three distinct topologies ("line", "square", and "T shape"). Moreover, for each topology, we generate statically several schedules, each one subject to a different total energy consumption constraint, and consequently with a different Worst-Case Reaction Time (WCRT). Coping with the energy consumption constraints is achieved thanks to Dynamic Frequency and Voltage Scaling (DVFS). In the second layer, we use these pre-generated static schedules to reconfigure dynamically the applications running on the multi-core each time a new application is launched or an existing one is stopped. The goal of the second layer is to perform a dynamic global optimization of the configuration, such that each running application meets a pre-defined quality-of-service constraint (translated into an upper bound on its WCRT) and such that the total energy consumption be minimized. For this, we (1) allocate a sufficient number of cores to each active application, (2) allocate the unassigned cores to the applications yielding the largest gain in energy, and (3) choose for each application the best topology for its subset of cores (*i.e.*, better than the by default "line" topology). This is a joint work with Ismail Assayad (U. Casablanca, Morocco) who visited the team in September 2015.

Second, in the context of a static application (again represented a DAG of tasks) running on an homogeneous multi-core chip, we have worked on the static scheduling minimizing the WCRT of the application under the multiple constraints that the reliability, the power consumption, and the temperature remain below some given threshold. There are multiple difficulties: (1) the reliability is not an invariant measure w.r.t. time, which makes it impossible to use backtrack-free scheduling algorithms such as list scheduling [37]; to overcome this, we adopt instead the Global System Failure Rate (GSFR) as a measure of the system's reliability that is invariant with time [64]; (2) keeping the power consumption under a given threshold requires to lower the voltage and frequency, but this has a negative impact both on the WCRT and on the GSFR; keeping the GSFR below a given threshold requires to replicate the tasks on multiple cores, but this has a negative impact both on the WCRT, on the power consumption, and on the temperature; (3) keeping the temperature below a given threshold is even more difficult because the temperature continues to increase even after the activity stops, so each scheduling decision must be assessed not based on the current state of the chip (*i.e.*, the temperature of each core) but on the state of the chip at the end of the candidate task, and cooling slacks must be inserted. This is a joint work with Atena Abdi (Amirkabir U., Iran) who is a PhD visitor in the team.

### 6.3.3. *Automatic transformations for fault tolerant circuits*

In the past years, we have studied the implementation of specific fault tolerance techniques in real-time embedded systems using program transformation [1]. We are now investigating the use of automatic transformations to ensure fault-tolerance properties in digital circuits. To this aim, we consider program transformations for hardware description languages (HDL). We consider both single-event upsets (SEU) and single-event transients (SET) and fault models of the form *"at most 1 SEU or SET within $n$ clock cycles"*.

We have proposed novel fault-tolerance transformations based on time-redundancy. In particular, we have presented a transformation using double-time redundancy (DTR) coupled with micro-checkpointing, rollback and a speedup mode [19]. The approach is capable to mask any SET every 10 cycles and keeps the same input/output behavior regardless error occurrences. Usually transparent masking requires triple redundancy and voting. Experimental results on the ITC'99 benchmark suite indicate that the hardware overhead of DTR is 2.7 to 6.1 times smaller than full TMR with a double loss in throughput. The method does not require any specific hardware support and is an interesting alternative to Triple Modular Redundancy (TMR) for logic intensive designs.

We have also designed a transformation that allows the circuit to change its level of time-redundancy. This feature allows the circuit to dynamically and temporarily lower (resp. increase) fault-tolerance and speed up (resp. slow down) its computation without interruption [20]. The motivations for such changes can be based on the current radiation environment or the processing of critical data. When hardware size is limited and fault-tolerance is only occasionally needed, that scheme is a better choice than static TMR, which involves a constant high area overhead

These time redundancy transformations (DTR and adaptive fault-tolerance) have been patented [50]

We have described how to formally certify fault-tolerant transformations using the CoQ proof assistant [53] (see Section 5.3 ). The transformations are described on a simple gate-level hardware description language LDDL (Low-level Dependent Description Language). This combinator language is equiped with dependent types and ensures that circuits are well-formed by construction (gates correctly plugged, no dangling wires, no combinational loops, ...). Fault-models are specified in the operational semantics of the language. The main theorem states that, for any circuit, for any input stream and for any SET allowed by the fault-model, its transformed version produces a correct output [18]. The primary motivation of this work was to certify DTR whose intricacy requested a formal proof to make sure that no single-point of failure existed. We have first applied this approach to the correctness proofs of TMR, TTR (Triple Time Redundancy) and finally DTR.

This research is part of Dmitry Burlyaev's pHD thesis [11] defended in November 2015.

### 6.3.4. *A formal approach for the synthesis and implementation of fault-tolerant embedded systems*

We have been working for several years on the usage of discrete controller synthesis (DCS) [83] to provide the automated addition of fault-tolerance in embedded systems with formal guarantees [65]. The first key idea is that the initial system model (usually an LTS) includes both the expected behaviors, the unexpected ones (that is, the failures), and the reconfigurations (typically repair actions). The second key idea is that the failures are modeled as *uncontrollable* events. Then, thanks to an exhaustive state space traversal, DCS is able to generate a *controller* that will prevent the system from entering a "bad" state (*e.g.*, a configuration of the system where a task is active on a faulty processor). From the point of view of fault-tolerance, this approach combines the advantages of static guarantees with that of dynamic reconfiguration (hence without the penalty of static redundancy).

Through this new work, we have demonstrated the feasibility of a complete workflow to synthesize and implement correct-by-construction fault tolerant distributed embedded systems consisting of real-time periodic tasks [24]. Correct-by-construction is provided by the use of DCS, which allows us to guarantee that the synthesized controlled system guarantees the functionality of its tasks even in the presence of processor failures. For this step, our workflow uses the HEPTAGON domain specific language [58] and the SIGALI DCS tool [79]. The correct implementation of the resulting distributed system is a challenge, all the more since the controller itself must be tolerant to the processor failures. We achieve this step thanks to the libDGALS real-time library [89] (1) to generate the glue code that will migrate the tasks upon processor failures, maintaining their internal state through migration, and (2) to make the synthesized controller itself fault-tolerant.

<span style="color:red">BIPOP Project-Team</span>

# 6. New Results

## 6.1. The contact complementarity problem, and Painlevé paradoxes

**Participants:** Bernard Brogliato, Florence Bertails-Descoubes, Alejandro Blumentals.

The contact linear complementarity problem is an set of equalities and complementarity conditions whose unknowns are the acceleration and the contact forces. It has been studied in a frictionless context with possibly singular mass matrix and redundant constraints, using results on well-posedness of variational inequalities obtained earlier by the authors [26]. This is also the topic of the first part of the Ph.D. thesis of Alejandro Blumentals where the frictional case is treated as a perturbation of the frictionless case [37]. The contact LCP is directly related to the so-called Painlevé's paradox of contact mechanics. In collaboration with C. Liu (Beijing university PKU) some results have been obtained from the analysis of a compliant model in the limit [34]. It shows on the classical sliding rod system that the inconsistent mode yield to instantaneous transition to a sticking mode. This is quite coherent with previous results obtained by Le xuan Anh in 1991 on the Painlevé-Klein system (bilateral constraints with Coulomb friction). With R. Kikuuwe from Kyushu University, we have also proposed a new formulation of the Baumgarte's stabilisation method, for unilateral constraints and Coulomb's friction , which sheds new light on Painlevé paradoxes as well [29].

## 6.2. Analysis of compliant nonlinear contact models

**Participants:** Bernard Brogliato, Guillaume James, Alexandre Vieira.

The master thesis of A. Vieira consisted of the study of suitable numerical method for compliant contact/impact models like the Simon-Hunt-Crossley and the Kuwabara-Kono models. These two models extend Hertz' contact by adding a dissipative force that takes the form of nonlinear viscous friction (nonlinear spring/dashpot). The fact that the Kuwabara-Kono dissipation is non-Lipschtiz requires particular care.

## 6.3. Discrete-time sliding mode control

**Participants:** Vincent Acary, Bernard Brogliato, Olivier Huber.

This topic concerns the study of time-discretized sliding-mode controllers. Inspired by the discretization of nonsmooth mechanical systems, we propose implicit discretizations of discontinuous, set-valued controllers [3]. This is shown to result in preservation of essential properties like simplicity of the parameters tuning, suppression of numerical chattering, reachability of the sliding surface after a finite number of steps, and disturbance attenuation by a factor $h$ or $h^2$ [22]. This work was part of the ANR project CHASLIM. Within the framework of CHASLIM we have performed many experimental validations on the electropneumatic setup of IRCCyN (Nantes), which nicely confirm our theoretical and numerical predictions: the implicit implementation of sliding mode control, drastically improves the input and output chattering behaviours, both for the classical order-one ECB-SMC and the twisting algorithms [33], [42], [27], [28]. In particular the high frequency bang-bang controllers which are observed with explicit discretizations, are completely suppressed. The implicit discretization has been applied to the classical equivalent-based-control SMC, and also to the twisting sliding-mode controller. The case of a nonlinear controller is studied in [35].

## 6.4. Lur'e set-valued dynamical systems: State observers

**Participants:** Bernard Brogliato, Christophe Prieur.

Lur'e systems are quite popular in Automatic Control since the fifties. Set-valued Lur'e systems possess a static feedback nonlinearity that is a multivalued function. We study in [53], [32] state observers for particular Lur'e systems which are Moreau's sweeping processes modelling Lagrange dynamics with frictionless unilateral constraints.

## 6.5. Measure Driven ODEs

**Participants:**  Bernard Brogliato, Christophe Prieur.

Measure driven Ordinary differential equations are analyzed in [31] from the point of view of input-to-state stability (ISS). This relies on the solution concept introduced by Bressan and Rampazzo. Lyapunov-like functions are used to caracterize the ISS. The link with impulsive ODEs and switching systems is made.

## 6.6. 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 [2]. They form a particular class of dynamical systems whose simulation requires the development of specific simulators.

### 6.6.1. *Numerical time–integration methods for event-detecting schemes.*

**Participants:**  Vincent Acary, Bernard Brogliato, Mounia Haddouni.

The CIFRE thesis of M. Haddouni concerns the numerical simulation of mechanical systems subject to holonomic bilateral constraints, unilateral constraints and impacts. This work is performed in collaboration with ANSYS and the main goal is to improve the numerical time–integration in the framework of event-detecting schemes. Between nonsmooth events, time integration amounts to numerically solving a differential algebraic equations (DAE) of index 3. We have compared dedicated solvers (Explicit RK schemes, Half-explicit schemes, generalizes $\alpha$-schemes) that solve reduced index formulations of these systems. Since the drift of the constraints is crucial for the robustness of the simulation through the evaluation of the index sets of active contacts, we have proposed some recommendations on the use of the solvers of dedicated to index-2 DAE. A manuscript has been submitted to Multibody System Dynamics.

### 6.6.2. *Multibody systems with clearances (dynamic backlash)*

**Participants:**  Vincent Acary, Bernard Brogliato, Narendra Akadkhar.

The PhD thesis of N. Akadkhar under contract with Schneider Electric concerns the numerical simulation of mechanical systems with unilateral constraints and friction, where the presence of clearances in imperfect joints plays a crucial role. A first work deals with four-bar planar mechanisms with clearances at the joints, which induce unilateral constraints and impacts, rendering the dynamics nonsmooth. The objective is to determine sets of parameters (clearance value, restitution coefficients, friction coefficients) such that the system's trajectories stay in a neighborhood of the ideal mechanism (*i.e.* without clearance) trajectories. The analysis is based on numerical simulations obtained with the projected Moreau-Jean time-stepping scheme. These results have been reported in [47]. It is planned to extend these simulations to frictional cases and to mechanisms of circuit breakers.

## 6.7. Nonlinear waves in granular chains

**Participants:**  Guillaume James, Bernard Brogliato, Alexandre Vieira.

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. While much effort has been devoted to the theoretical and experimental analysis of solitary waves in granular chains, there is now an increasing interest in the study of breathers (spatially localized oscillations) in granular systems. Due to their oscillatory nature and associated resonance phenomena, static or traveling breathers exhibit much more complex dynamical properties compared to solitary waves. Such properties have strong potential applications for the design of acoustic metamaterials allowing to efficiently damp or deviate shocks and vibrations. Our contribution to this field is twofold. In the work [52], the existence of static breathers is analyzed in granular metamaterials consisting of hollow beads with internal masses. Using multiple scale analysis and exploiting the unilateral character of Hertzian interactions, we show that long-lived breather solutions exist but time-periodic breathers do not (breather solutions actually disperse on long time scales). Moreover, in a collaboration with Y. Starosvetsky and D. Meimukhin (Technion), we numerically study the persistence of traveling breathers in

granular chains with local potentials under the effect of contact damping. Using a viscoelastic damping model (Hertz-Kuwabara-Kono model), we show that breathers can be generated by simple impacts in granular chains made from various materials (breathers propagate over a significant number of sites before being damped). The design of an experimental setup to test these theoretical predictions is underway. Another series of works concerns more specifically the modeling and numerical analysis of dissipative impacts : introduction of appropriate variables and simplifications for different models of contact damping (James, Brogliato), and comparative tests for various numerical discretizations of the Hunt-Crossley and Kuwabara-Kono models (Vieira, Brogliato, James).

## 6.8. Traveling pulses in the Burridge-Knopoff model

**Participants:** Guillaume James, Jose Eduardo Morales Morales, Arnaud Tonnelier.

The Burridge-Knopoff model describes the earthquake faulting process through the interaction of two plates modeled as a chain of blocks elastically coupled subject to a friction force. We study the existence of soliton-like solutions for the excitable Burridge-Knopoff model with different friction forces. We report for the first time the propagation of a one-pulse solitary wave where the position of the blocks remains unchanged after the passage of the wave. Extensive numerical simulations are done for different friction laws and a systematic investigation of the influence of the pulling velocity and the coupling constant is done. For a piecewise linear frictional law, we prove the existence of a traveling pulse in the weak coupling limit. A lower bound of the propagation speed is derived together with results on the shape of the traveling wave.

## 6.9. Propagation in space-discrete excitable systems

**Participant:** Arnaud Tonnelier.

We introduce a simplified model of excitable systems where the response of an isolated cell to an incoming signal is idealized by a fixed pulse-shape function. When the total activity of the cell reaches a given threshold a signal is sent to its $N$ neighbors. We show that a chain of such excitable cells is able to propagate a set of simple traveling waves where the time interval between the firing of two successive cells remains constant. A comprehensive study is done for a transmission line with $N = 2$ and $N = 3$. It is shown that, depending on initial copnditions, the network may propagate traveling waves with different velocities. Some necessary conditions for multistationarity are derived for an arbitrary $N$.

## 6.10. Inverse modeling with contact and friction

### 6.10.1. Inverse statics of plates and shells with frictional contact
**Participants:** Florence Bertails-Descoubes, Romain Casati, Gilles Daviet.

We pursued our work on the static inversion of thin elastic shells, in the presence of contact and friction with an external object. We have shown how to formulate draping as a local constrained minimization problem, and we have generalized the adjoint method to this constrained case. These new results are included in Romain Casati's PhD thesis, defended in June 2015, and will be part of a paper to be submitted in 2016.

## 6.11. Continnum modeling of granular materials

### 6.11.1. Continuum modeling of granular materials
**Participants:** Florence Bertails-Descoubes, Gilles Daviet.

We have proposed a new numerical framework for the continuous simulation of dilatable materials with pressure-dependent (Coulomb) yield stress, such as sand or cement. Relying upon convex optimization tools, we have shown that the continuous equations of motion coupled to the macroscopic nonsmooth Drucker-Prager rheology can be interpreted as the exact analogous of the solid frictional contact problem at the heart of Discrete Element Methods (DEM), extended to the tensorial space. Combined with a carefully chosen finite-element discretization, this new framework allowed us to avoid regularizing the continuum rheology while benefiting from the efficiency of nonsmooth optimization solvers, mainly leveraged by DEM methods so far. Our numerical results were successfully compared to analytic solutions on model problems, such as the silo discharge, and we retrieved qualitative flow features commonly observed in reported experiments of the literature. This work is currently under review at the Journal of Non Newtonian Fluid Mechanics, and a preliminary version is available as a research report [43]. Furthermore, we have recently extended the approach to account for flows with a varying density, leveraging the Material Point Method to discretize the Drucker Prager yield criterion without linearization. This work will be submitted to ACM SIGGRAPH in 2016.

## 6.12. Nonsmooth optimisation and applications

### 6.12.1. *Semidefinite programming and combinatorial optimization*
**Participant:** Jérôme Malick.

We have worked with Frederic Roupin (Prof. at Paris XIII) and Nathan Krislock (Assistant Prof. at North Illinois University, USA) on the use of semidefinite programming to solve combinatorial optimization problems to optimality. Nathan was the guest of the team during 2 months (June/July).

We have worked on a generic semidefinite-based solver for solve binary quadratic optimization problems. Using the generality of the bounds proposed in [54]. Our article is in revision in ACM Transaction of Mathematical Software. Our solver and our data sets are available online at http://lipn.univ-paris13.fr/BiqCrunch/.

Specializing the method of the k-cluster problem, we have proposed in [51] an algorithm able to solve exactly k-cluster instances of size 160. In practice, our method works particularly fine on the most difficult instances (with a large number of vertices, small density and small k).

### 6.12.2. *Stochastic optimization for electricity production*
**Participant:** Jérôme Malick.

Everyday, electricity generation companies submit a generation schedule to the grid operator for the coming day; computing an optimal schedule is called the unit-commitment problem. In collaboration with W. van Ackooij from EDF, we have proposed in [44] a two-stage formulation of unit-commitment to better include the impact of renewable energies. We present a primal-dual decomposition approach to tackle large-scale instances of these two-stage problems, wherein both the first and second stage problems are full unit-commitment problems. We provide an analysis of the theoretical properties of the algorithm, as well as computational experiments showing the interest of the approach for real-life large-scale unit-commitment instances.

## 6.13. Robotics

### 6.13.1. *Mobile manipulation by humanoid robots*
**Participants:** Pierre-Brice Wieber, Dimitar Dimitrov, Alexander Sherikov, Jory Lafaye, Camille Brasseur, Nestor Alonso Bohorquez Dorante.

This year's contributions to the field of mobile manipulation by humanoid robots have been three-fold: a lexicographic MPC approach to the decision of using optional contacts when necessary to maintain balance (and only when necessary), a robust MPC approach to online generation of dynamic walking motion on uneven ground such as stairs, and an analysis of the role of viability and capturability in collision prevention, using once again a lexicographic MPC approach.

### 6.13.2. Reactive trajectory generation

**Participants:** Pierre-Brice Wieber, Dimitar Dimitrov, Saed Al Homsi.

The goal of the collaboration with Adept Technologies is to generate time optimal trajectories in the presence of moving obstacles in real time. Three approaches with increasing computational complexity have been proposed and validated experimentally. The cheapest approach begins with a standard bang-bang control which is time-optimal in the absence of obstacles, and simply projected on dynamic limits imposed by collision avoidance. This leads to reasonable results where collisions are explicitly avoided, but time-optimality is lost in the process. A more complex approach introduces an MPC scheme minimizing a weighted L1-norm, which is tuned to generate a time-optimal behavior in the absence of obstacles. In the presence of obstacles, time-optimality is once again lost, however, results are much improved with respect to the previous approach. The final, and most complex approach, considers time-optimality as a lexicographic objective: a lexicographic MPC scheme is proposed, which achieves time-optimality in the presence of obstacles, with reasonable online computation time. This work has been submitted to ICRA 2016.

<h1 style="text-align:center; color:red;">MISTIS Project-Team</h1>

# 7. New Results

## 7.1. Mixture models

### 7.1.1. Taking into account the curse of dimensionality

**Participants:** Stéphane Girard, Alessandro Chiancone, Seydou-Nourou Sylla.

**Joint work with:** C. Bouveyron (Univ. Paris 5), M. Fauvel (ENSAT Toulouse) and J. Chanussot (Gipsa-lab and Grenoble-INP)

In the PhD work of Charles Bouveyron (co-advised by Cordelia Schmid from the Inria LEAR team)  [61], we proposed new Gaussian models of high dimensional data for classification purposes. We assume that the data live in several groups located in subspaces of lower dimensions. Two different strategies arise:

- the introduction in the model of a dimension reduction constraint for each group
- the use of parsimonious models obtained by imposing to different groups to share the same values of some parameters

This modelling yields a supervised classification method called High Dimensional Discriminant Analysis (HDDA) [4]. Some versions of this method have been tested on the supervised classification of objects in images. This approach has been adapted to the unsupervised classification framework, and the related method is named High Dimensional Data Clustering (HDDC) [3]. Our recent work consists in adding a kernel in the previous methods to deal with nonlinear data classification and heterogeneous data [13]. We first investigate the use of kernels derived from similary measures on binary data [30]. The targeted application is the analysis of verbal autopsy data (PhD thesis of N. Sylla): Indeed, health monitoring and evaluation make more and more use of data on causes of death from verbal autopsies in countries which do not keep records of civil status or with incomplete records. The application of verbal autopsy method allows to discover probable cause of death. Verbal autopsy has become the main source of information on causes of death in these populations. Second, the kernel classification method is applied to three real hyperspectral data sets,and compared with three others classifiers. The proposed models show good results in terms of classification accuracy and processing time [21].

## 7.2. Semi and non-parametric methods

### 7.2.1. Conditional extremal events

**Participant:** Stéphane Girard.

**Joint work with:** L. Gardes (Univ. Strasbourg), G. Mazo (Univ. Catholique de Louvain), J. Elmethni (Univ. Paris 5) and S. Louhichi (Univ. Grenoble 1)

The goal of the PhD theses of Alexandre Lekina and Jonathan El Methni was 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, the tail heaviness of $Y$ depends on $X$, and thus the tail index as well as the extreme quantiles are also functions of the covariate. We combine nonparametric smoothing techniques  [63] with extreme-value methods in order to obtain efficient estimators of the conditional tail index and conditional extreme quantiles. The strong consistency of such an estimator is established in [53]. When the covariate is functional and random (random design) we focus on kernel methods [23].

Conditional extremes are studied in climatology where one is interested in how climate change over years might affect extreme temperatures or rainfalls. In this case, the covariate is univariate (time). Bivariate examples include the study of extreme rainfalls as a function of the geographical location. The application part of the study is joint work with the LTHE (Laboratoire d'étude des Transferts en Hydrologie et Environnement) located in Grenoble [20] and the "département Génie urbain" of "Université Paris-Est Marne-la-vallée" [11].

### 7.2.2. *Estimation of extreme risk measures*

**Participant:** Stéphane Girard.

**Joint work with:** A. Daouia (Univ. Toulouse), E. Deme (Univ. Gaston-Berger, Sénégal), A. Guillou (Univ. Strasbourg) and G. Stupfler (Univ. Aix-Marseille).

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 $\alpha$-quantile of the loss distribution. The Value-at-Risk however suffers from several weaknesses. First, it provides us only with a pointwise information: VaR$(\alpha)$ 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 $\alpha$-quantile of the loss distribution. This risk measure thus takes into account the whole information contained in the upper tail of the distribution. It is frequently encountered in financial investment or in the insurance industry. In [51], we have established the asymptotic properties of the CTE estimator in case of extreme losses, *i.e.* when $\alpha \to 0$ as the sample size increases. We have exhibited the asymptotic bias of this estimator, and proposed a bias correction based on extreme-value techniques. A second possible coherent alternative risk measure is based on expectiles [59]. 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. Both quantiles and expectiles were embedded in the more general class of M-quantiles as the minimizers of a generic asymmetric convex loss function. It has been proved very recently that the only M-quantiles that are coherent risk measures are the expectiles.

### 7.2.3. *Multivariate extremal events*

**Participants:** Stéphane Girard, Florence Forbes.

**Joint work with:** F. Durante (Univ. Bolzen-Bolzano, Italy) L. Gardes (Univ. Strasbourg) and G. Mazo (Univ. Catholique de Louvain, Belgique).

Copulas are a useful tool to model multivariate distributions  [67].

However, while there exist various families of bivariate copulas, much fewer has been done when the dimension is higher. To this aim an interesting class of copulas based on products of transformed copulas has been proposed in the literature. The use of this class for practical high dimensional problems remains challenging. Constraints on the parameters and the product form render inference, and in particular the likelihood computation, difficult. We proposed a new class of high dimensional copulas based on a product of transformed bivariate copulas [26]. No constraints on the parameters refrain the applicability of the proposed class which is well suited for applications in high dimension. Furthermore the analytic forms of the copulas within this class allow to associate a natural graphical structure which helps to visualize the dependencies and to compute the likelihood efficiently even in high dimension. The extreme properties of the copulas are also derived and an R package has been developed.

As an alternative, we also proposed a new class of copulas constructed by introducing a latent factor. Conditional independence with respect to this factor and the use of a nonparametric class of bivariate copulas lead to interesting properties like explicitness, flexibility and parsimony. In particular, various tail behaviours are exhibited, making possible the modeling of various extreme situations [19], [27], [52]. A pairwise moment-based inference procedure has also been proposed and the asymptotic normality of the corresponding estimator has been established [28].

In collaboration with L. Gardes, we investigate the estimation of the tail copula, which is widely used to describe the amount of extremal dependence of a multivariate distribution. In some situations such as risk management, the dependence structure can be linked with some covariate. The tail copula thus depends on this covariate and is referred to as the conditional tail copula. The aim of our work is to propose a nonparametric estimator of the conditional tail copula and to establish its asymptotic normality [22].

### 7.2.4. *Level sets estimation*
**Participant:**  Stéphane Girard.

**Joint work with:** G. Stupfler (Univ. Aix-Marseille)

The boundary bounding the set of points is viewed as the larger level set of the points distribution. This is then an extreme quantile curve estimation problem. We proposed estimators based on projection as well as on kernel regression methods applied on the extreme values set, for particular set of points [10]. We also investigate the asymptotic properties of existing estimators when used in extreme situations. For instance, we have established in collaboration with G. Stupfler that the so-called geometric quantiles have very counter-intuitive properties in such situations [24], [25] and thus should not be used to detect outliers.

### 7.2.5. *Retrieval of Mars surface physical properties from OMEGA hyperspectral images.*
**Participants:**  Stéphane Girard, Alessandro Chiancone.

**Joint work with:** J. Chanussot (Gipsa-lab and Grenoble-INP).

Visible and near infrared imaging spectroscopy is one of the key techniques to detect, to map and to characterize mineral and volatile (eg. water-ice) species existing at the surface of planets. Indeed the chemical composition, granularity, texture, physical state, etc. of the materials determine the existence and morphology of the absorption bands. The resulting spectra contain therefore very useful information. Current imaging spectrometers provide data organized as three dimensional hyperspectral images: two spatial dimensions and one spectral dimension. Our goal is to estimate the functional relationship $F$ between some observed spectra and some physical parameters. To this end, a database of synthetic spectra is generated by a physical radiative transfer model and used to estimate $F$. The high dimension of spectra is reduced by Gaussian regularized sliced inverse regression (GRSIR) to overcome the curse of dimensionality and consequently the sensitivity of the inversion to noise.

In his PhD thesis work, Alessandro Chiancone studies the extension of the SIR method to different sub-populations. The idea is to assume that the dimension reduction subspace may not be the same for different clusters of the data [14].

### 7.2.6. *Robust Sliced Inverse Regression.*
**Participants:**  Stéphane Girard, Alessandro Chiancone, Florence Forbes.

Sliced Inverse Regression (SIR) has been extensively used to reduce the dimension of the predictor space before performing regression. Recently it has been shown that this techniques is, not surprisingly, sensitive to noise. Different approaches has been proposed to robustify SIR, in this work, we start considering an inverse problem proposed by R.D. Cook and we show that the framework can be extended to take into account a non-Gaussian noise. Generalized Student distribution are considered and all parameters are estimated via EM algorithm. The algorithm is outlined and tested comparing the results with different approaches on simulated data. Results on a real dataset shows the interest of this technique in presence of outliers.

### 7.2.7. *Robust Locally linear mapping with mixtures of Student distributions*

**Participants:** Florence Forbes, Emeline Perthame, Brice Olivier, Leo Nicoletti.

The standard GLLiM model [17] for high dimensional regression assumes Gaussian noise models and is in its unconstrained version equivalent to a joint GMM. The fact that response and independent variables $(X, Y)$ are jointly a mixture of Gaussian distribution is the key for all derivations in the model. In this work, we show that similar developments are possible based on a joint Student Mixture model, joint SMM. It follows a new model referred to as SLLiM for Student Locally linear mapping for which we investigate the robustness to outlying data in a high dimensional regression context.

## 7.3. Markov models

### 7.3.1. *Change-point models for tree-structured data*

**Participant:** Jean-Baptiste Durand.

**Joint work with:** Pierre Fernique (Inria) and Yann Guédon (CIRAD), Inria Virtual Plants.

In the context of plant growth modelling, methods to identify subtrees of a tree or forest with similar attributes have been developed. They rely either on hidden Markov modelling or multiple change-point approaches. The latter are well-developed in the context of sequence analysis, but their extensions to tree-structured data are not straightforward. Their advantage on hidden Markov models is to relax the strong constraints regarding dependencies induced by parametric distributions and local parent-children dependencies. Heuristic approaches for change-point detection in trees were proposed and applied to the analysis of patchiness patterns (consisting of canopies made of clumps of either vegetative or flowering botanical units) in mango trees [45].

### 7.3.2. *Hidden Markov models for the analysis of eye movements*

**Participants:** Jean-Baptiste Durand, Brice Olivier.

*This research theme is supported by a LabEx PERSYVAL-Lab project-team grant.*

**Joint work with:** Marianne Clausel (LJK) 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 make some kinds of press reviews. 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 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 through eye-movements. Our approach was inspired by Simola *et al.* (2008) [68], but we used hidden semi-Markov models for better characterization of phase length distributions. The estimated HMM highlighted contrasted reading strategies (i.e., state transitions), with both individual and document-related variability.

However, the characteristics of eye movements within each phase tended to be poorly discriminated. As a result, high uncertainty in the phase changes arose, and it could be difficult to relate phases to known patterns in EEGs.

This is why, as part of Brice Olivier's PhD thesis, we are developing integrated models coupling EEG and eye movements within one single HMM for better identification of the phases. Here, the coupling should incorporate some delay between the transitions in both (EEG and eye-movement) chains, since EEG patterns associated to cognitive processes occur lately with respect to eye-movement phases. Moreover, EEGs and scanpaths were recorded with different time resolutions, so that some resampling scheme must be added into the model, for the sake of synchronizing both processes.

### 7.3.3. *Lossy compression of tree structures*
**Participant:** Jean-Baptiste Durand.

**Joint work with:** Christophe Godin (Inria, Virtual Plants) and Romain Azais (Inria BIGS)

In a previous work [65], a method to compress tree structures and to quantify their degree of self-nestedness was developed. This method is based on the detection of isomorphic subtrees in a given tree and on the construction of a DAG (Directed Acyclic Graph), equivalent to the original tree, where a given subtree class is represented only once (compression is based on the suppression of structural redundancies in the original tree). In the lossless compressed graph, every node representing a particular subtree in the original tree has exactly the same height as its corresponding node in the original tree. A lossy version of the algorithm consists in coding the nearest self-nested tree embedded in the initial tree. Indeed, finding the nearest self-nested tree of a structure without more assumptions is conjectured to be an NP-complete or NP-hard problem. We improved this lossy compression method by computing a self-nested reduction of a tree that better approximates the initial tree. The algorithm has polynomial time complexity for trees with bounded outdegree. This approximation relies on an indel edit distance that allows (recursive) insertion and deletion of leaf vertices only. We showed in a conference paper accepted at DCC'2016 [55] with a simulated dataset that the error rate of this lossy compression method is always better than the loss based on the nearest embedded self-nestedness tree [65] while the compression rates are equivalent. This procedure is also a keystone in our new topological clustering algorithm for trees. In addition, we obtained new theoretical results on the combinatorics of self-nested structures. The redaction of an article is currently in progress.

## 7.4. Statistical models for Neuroscience

### 7.4.1. *Comparison of stochastic and variational solutions to ASL fMRI data analysis*
**Participants:** Florence Forbes, Aina Frau Pascual.

**Joint work with:** Philippe Ciuciu from Team Parietal and Neurospin, CEA Saclay.

Functional Arterial Spin Labeling (fASL) MRI can provide a quantitative measurement of changes of cerebral blood flow induced by stimulation or task performance. fASL data is commonly analysed using a general linear model (GLM) with regressors based on the canonical hemodynamic response function. In this work [37], we consider instead a joint detection-estimation (JDE) framework which has the advantage of allowing the extraction of both task-related perfusion and hemodynamic responses not restricted to canonical shapes. Previous JDE attempts for ASL have been based on computer intensive sampling (MCMC) methods. Our contribution is to provide a comparison with an alternative variational expectation-maximization (VEM) algorithm on synthetic and real data. Other investigations were related to the use of appropriate physiological information and priors [39], [38].

### 7.4.2. *A differential evolution-based approach for fitting a nonlinear biophysical model to fMRI BOLD data*
**Participants:** Florence Forbes, Pablo Mesejo.

**Joint work with:** Jan Warnking from Grenoble Institute of Neuroscience.

Physiological and biophysical models have been proposed to link neuronal activity to the Blood Oxygen Level-Dependent (BOLD) signal in functional MRI (fMRI). Those models rely on a set of parameter values that cannot always be extracted from the literature. In some applications, interesting insight into the brain physiology or physiopathology can be gained from an estimation of the model parameters from measured BOLD signals. This estimation is challenging because there are more than 10 potentially interesting parameters involved in nonlinear equations and whose interactions may result in identifiability issues. However, the availability of statistical prior knowledge about these parameters can greatly simplify the estimation task. In this work we focus on the extended Balloon model and propose the estimation of 15 parameters using two stochastic approaches: an Evolutionary Computation global search method called Differential Evolution (DE)

and a Markov Chain Monte Carlo version of DE. To combine both the ability to escape local optima and to incorporate prior knowledge, we derive the target function from Bayesian modeling. The general behavior of these algorithms is analyzed and compared with the *de facto* standard Expectation Maximization Gauss-Newton (EM/GN) approach, providing very promising results on challenging real and synthetic fMRI data sets involving rats with epileptic activity. These stochastic optimizers provided a better performance than EM/GN in terms of distance to the ground truth in 4 out of 6 synthetic data sets and a better signal fitting in 12 out of 12 real data sets. Non-parametric statistical tests showed the existence of statistically significant differences between the real data results obtained by DE and EM/GN. Finally, the estimates obtained from DE for these parameters seem both more realistic and more stable or at least as stable across sessions as the estimates from EM/GN. This work will appear in [29]. A preliminary version has also been accepted at the conference MICCAI 2015 [40].

### 7.4.3. *Multi-subject joint parcelation detection estimation in functional MRI*
**Participant:** Florence Forbes.

**Joint work with:** Lotfi Chaari, Mohanad Albughdadi, Jean-Yves Tourneret from IRIT-ENSEEIHT in Toulouse and Philippe Ciuciu from Neurospin, CEA Saclay.

fMRI experiments are usually conducted over a population of interest for investigating brain activity across different regions, stimuli and subjects. Multi-subject analysis usually proceeds in two steps: an intra-subject analysis is performed sequentially on each individual and then a group-level analysis is carried out to report significant results at the population level. This work considers an existing Joint Parcellation Detection Estimation (JPDE) model which performs joint hemodynamic parcellation, brain dynamics estimation and evoked activity detection. The hierarchy of the JPDE model is extended for multi-subject analysis in order to perform group-level parcellation. Then, the corresponding underlying dynamics is estimated in each parcel while the detection and estimation steps are iterated over each individual. Validation on synthetic and real fMRI data shows its robustness in inferring group-level parcellation and the corresponding hemodynamic profiles. This work has been accepted at ISBI 2016.

### 7.4.4. *Tumor classification and prediction using robust multivariate clustering of multiparametric MRI*
**Participants:** Florence Forbes, Alexis Arnaud.

**Joint work with:** Emmanuel Barbier and Benjamin Lemasson from Grenoble Institute of Neuroscience.

Advanced statistical clustering approaches are promising tools to better exploit the wealth of MRI information especially on large cohorts and multi-center studies. In neuro-oncology, the use of multiparametric MRI may better characterize brain tumor heterogeneity. To fully exploit multiparametric MRI (e.g. tumor classification), appropriate analysis methods are yet to be developed. They offer improved data quality control by allowing automatic outlier detection and improved analysis by identifying discriminative tumor signatures with measurable predictive power. In this work, we show on small animals data that advanced statistical learning approaches can help 1) in organizing existing data by detecting and excluding outliers and 2) in building a dictionary of tumor fingerprints from a clustering analysis of their microvascular features. The work also now includes the integration in a joint statistical model of both automatic ROI delineation and clustering for whole brain data analysis. A preliminary version of this work has been accepted to the ISMRM 2015 conference and in the SFMRMB 2015 conference [41].

### 7.4.5. *Functional specifications of a brain segmentation software*
**Participants:** Florence Forbes, Priscillia Previtero.

**Joint work with:** Michel Dojat from Grenoble Institute of Neuroscience and Senan Doyle from Pixyl.

The goal of P. Previtero's internship was to help with a number of software engineering tasks and communications actions around the P-Locus software and the Pixyl start-up. The internship resulted in particular in a new web site for Pixyl.

<div align="center">

## NANO-D Project-Team

</div>

# 7. New Results

## 7.1. Algorithms for Orbital-Free Density Functional Theory

**Participants:**  Francois Rousse, Stephane Redon.

The Schrödinger equation permits, in theory, to model and simulate every molecular systems exactly. Unfortunately it is not computationally doable to solve this equation even on really small systems (2 atoms). Density Functional Theory (DFT) gives a method to solve this equation, find the electronic structure and simulate molecules with the laws of physics on reasonably large system : from 1.000 to 10.000 depending on the basis chosen and the version of DFT used. Unfortunately, the computation of kinetic energy requires the orthogonalization of the basis, which consumes a lot of time and prevents the algorithm from being adaptive : one needs to recompute the whole system if a little change is done in the molecules position. One can deals with this issue by computing the kinetic energy directly with the electronic density and not anymore with the orbitals. That is the idea of Orbital-Free DFT (OF-DFT). It can models great systems (up to 1.000.000 atoms) and be turned adaptive. On the other hand, it looses a lot of accuracy and power to model different kind of systems on the other DFT.

We have already developed our own OF-DFT code. It runs on parallel cores, is implemented in the SAMSON platform as a SAMSON App and gives correct electron's densities. The electronic structures are computed in real space to preserve the possibility of incremental calculations. We are now going to test our implementation, and will then attempt to make the method adaptive. The difficulty will be the determination of the domain that needs to be recomputed when a part of the system has moved, and the criteria that will help to do so.

## 7.2. Parallel adaptively restrained particle simulations

**Participants:**  Krishna Kant Singh, Stephane Redon.

We have continued our work on the development of parallel adaptively restrained particle simulations. We developed new algorithms for neighbor list and incremental force updates. These algorithms have advantages over the state-of-the-art methods for simulating a system using Adaptively Restrained Molecular Dynamics (ARMD). We have simulated systems with different number (500, 4000 and 108000) of LJ particles using adaptively restrained integrator and Lennard-Jones potential in NVE (constant number of particles, Volume and Energy) and NVT ensemble (constant number of particles, Volume and Temperature). All the particles were placed in an orthogonal box. We used periodic periodic boundary conditions with 8.5 angstrom cut-off for the Lennard-Jones potential. The system was simulated using 2 femtoseconds. We compared the LAMMPS algorithm to adaptive algorithm while using adaptively restrained integrator. Our results show that a significant speed-up can be achieved if more than $60\%$ of the particles are restrained (Figure 4 ). Figure 5  shows that ARMD in NVT ensemble preserves the average temperature of the system (irrespective of number of restrained particles).

## 7.3. Incremental algorithms for long-range interactions

**Participants:**  Semeho Edorh, Stephane Redon.

Numerical simulations of molecular dynamics (MD) are very expensive in terms of CPU resources. During Molecular dynamics simulations, the most CPU intensive task is the evaluation of the interaction potential [78]. Due to the large number of particles involved, updating this potential may have, at each time-step, a very high computational cost.
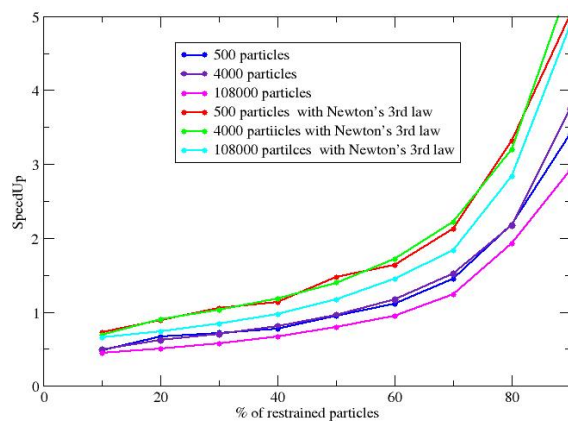
*Figure 4. Speedup using ARMD on different benchmark*
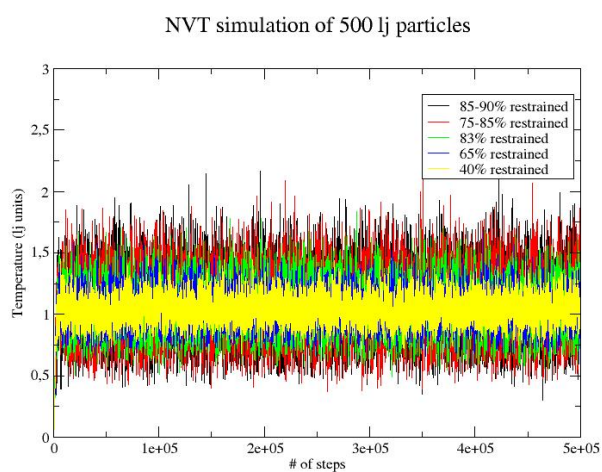
NVT simulation of 500 lj particles



*Figure 5. Temperature profile of 500 LJ particles in NVT ensemble using ARMD*

In large crystalline ionic system, Ewald summation is the most popular method for computing electrostatic interactions. It rewrites the interaction potential $\phi$ as the sum of a short-range term and a long-range term. Ewald summation using optimal parameters requires $\mathcal{O}(N^{3/2})$ operations [47], [30] but it can be modified so that it involves only about $\mathcal{O}(N \log N)$ operations [31], [85] by using the Fast Fourier Transform.

We want to develop a new approach that can reduce the computational cost by using incremental algorithms. The key-idea is to use, for each time-step of the simulation, information that we have computed in previous steps.

The Particle Mesh Ewald (PME) algorithm developped by Darden et al. is the most successful approach for computing long range interactions. In the particle mesh method, just as in standard Ewald summation, the generic interaction potential is separated into two terms.The so-called short-range contribution can be easily calculated in a direct space by using truncation methods. Where as the long-range contribution is calculated using two Fast Fourier transforms ($N \log(N)$ algorithm ). In pratice, the long range contribution algorithm boil than to [30]:

- Map particle charge density $Q$ to a mesh
- Compute the forward Fast Fourier Transform of the approximation $Qm$ of charge density on the mesh
- Multiply $Qm$ by a green function (related to the choice of the mesh).
- Compute a backward Fast Fourier Transform of the result.
- Retrieve the long-range contribution potential by interpolating the previous result at particles positions.

We modified this algorithm to make it incremental. We started from the PME implementation in LAMMPS. Instead of mapping the charge density to the mesh, we mapped the increment of density $dQ$ to the mesh. The FFT solver KissFFT is based on a divide-and-conquer algorithm. We built a sparse input solver as a modified version of FFT solver which computes only needed (non trivial) operations [74]. We built also a sparse output solver inspired by the algorithm proposed by Katabi et al. [36]. Unfortunately, we did not get significant speed-ups with these modifications.

We decided to compute the increment of the long-range contribution related to the increment of density $dQ$ by using multi-resolution methods. These methods are slower than PME but have better adaptive behavior. The multigrid approach was chosen because of its $\mathcal{O}(N)$ behavior and its good scalability [13]. We are currently developing an adaptive multigrid method.

## 7.4. Motion planning architecture for nanosystems

**Participants:** Leonard Jaillet, Stephane Redon.

In the past, we have started the development of original quasi-static simulation methods for nano-scale systems, based on motion planning methods inspired from Robotics. In the continuity of this work, we have proposed an original Motion Planning architecture for nanosystems platform called planning. This platform offers a general framework for motion planning applied to nanosystems. In particular it includes:

- A flexible definition of the degrees of freedom that describe the system, allowing different levels of representation (e.g. Cartesian coordinates, internal coordinates, coarse grain representation, etc.).
- The possibility to define an arbitrary set of initial, final and intermediate states, guiding the search for a solution path.
- The possibility to define an arbitrary set of constraints on the intermediate states of the path (e.g. geometric constraints, energy constraints, etc.)
- Several modular functionalities specific to motion planning (e.g. conformational sampling, exploration strategy, nearest neighbor search, etc.)
- An adapted integration within SAMSON which allows using directly all the existing force fields and state updaters present in the platform.

The planning architecture has been the base of several SAMSON modules. In particular, it led to the Planner-Explore module, which regroups many of the functionalities proposed and that can by combined together through a graphical interface. This module has in particular been used to study two complex problems:

- To capture the transition paths between endiandric acids (see Figure 6 ).
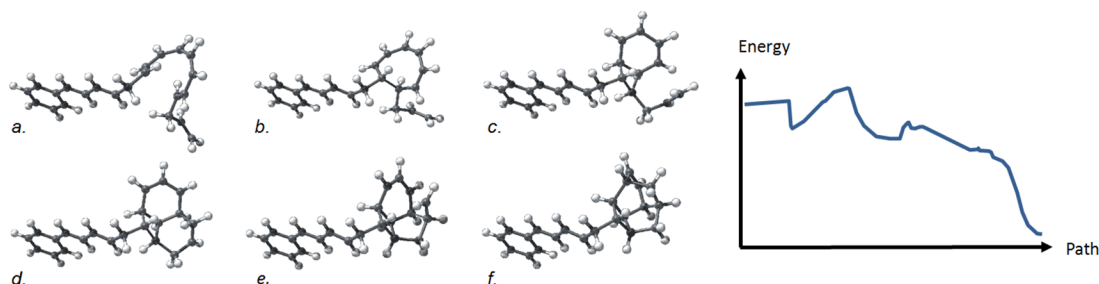- To find the global minima of Lennard-Jones clusters, for dimensions up to one hundred.



*Figure 6. Transition path with its corresponding energy for an homolog of the endiandric acid and produced thanks to the Planner-Explore module.*

## 7.5. Optimization of transition paths

**Participants:** Leonard Jaillet, Stephane Redon.

Motion planning methods allow producing initial paths which represent transitions from one given conformation to another. However, these paths are typically suboptimal because of the probabilistic nature of the search strategy. Hence, it is necessary to develop tools to locally increase the path quality of the solution generated during the first phase. We have proposed several methods to address such a problem. One method developed is a variant of a state-of-the-art approach called nudged elastic band (NEB). It optimizes a set of intermediate images along the path, such that each image finds the lowest energy possible while maintaining equal spacing to neighboring images. Another technique we proposed is to rely on an equivalent of the shortcutting technique developed in Robotics motion planning, but applied to the context of energy landscape. Finally, we also have complemented these methods with additional tools to do simple path edition such as cutting or thinning paths.

## 7.6. As-rigid-as-possible shape interpolation for molecular modeling

**Participants:** Minh Khoa Nguyen, Leonard Jaillet, Stephane Redon.

Computer-aided methods play an important role in the study of molecular structures and interactions. Inspired by the as-rigid-as-possible approaches in the field of computer graphics, we created a tool for studying large deformation of molecular structures. This tool generates interpolated structures between two known conformations of a molecule while satisfying physical constraints. The users may use it for exploring, preprocessing, or combining their model with other biological algorithms. The developed method is flexible and can be extended to include physical properties of molecular structures.

We tested our method on a graphene sheet folding into a nanotube (Figure 8 ) and a few biological molecules, one of which is shown in Figure 9 . The results show realistic transition motions compared to those from the linear interpolation approach.

The ARAP interpolation method has two main advantages: simplicity and preservation of local rigidity. The method is totally geometrical, yet can be extended to include physical or biological properties such as bond strength. It will be proposed as a SAMSON Element for the SAMSON software platform for computational nanoscience.
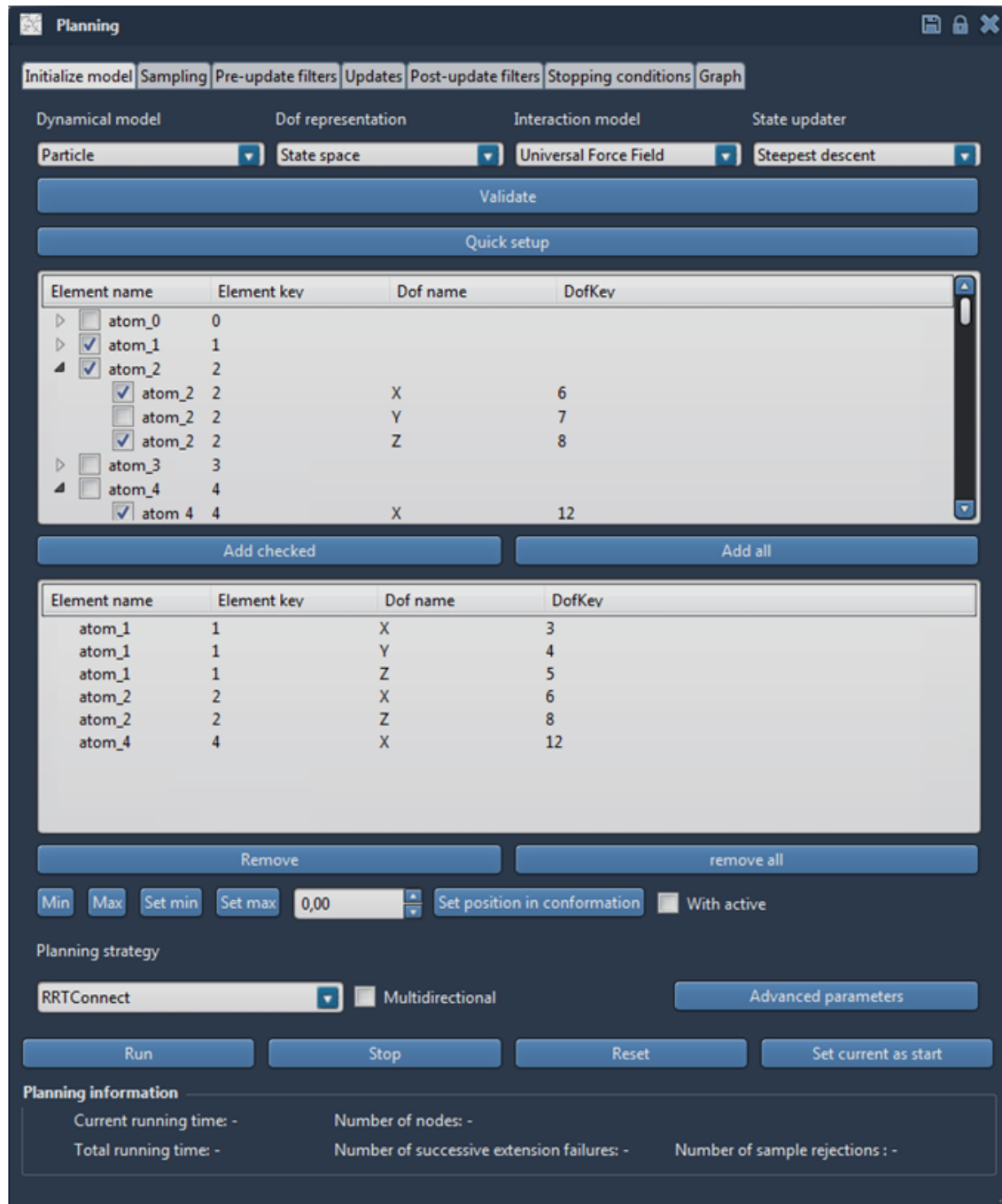
*Figure 7. One tab of the graphical interface of the Planner-Explore modules, which allows initializing the model to be simulated.*
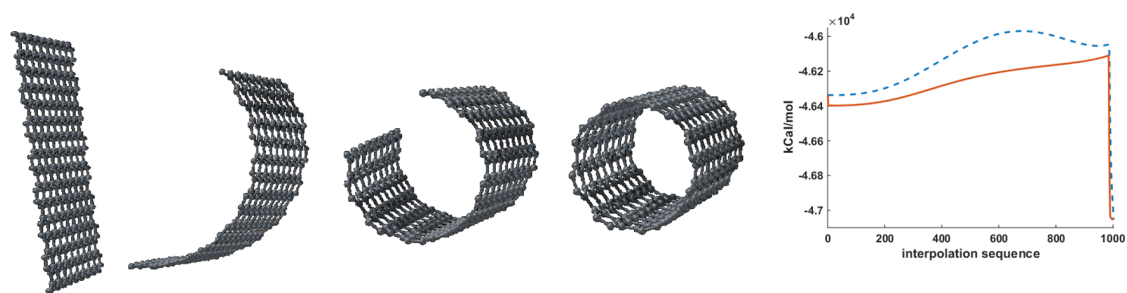
*Figure 8. ARAP interpolation to generate graphene sheet folding into a nanotube. The last figure plots the energy for a sequence of 1000 interpolated images. The energy of ARAP interpolation is shown by the dotted blue curve and the optimal energy after applying NEB is shown by the solid red curve.*
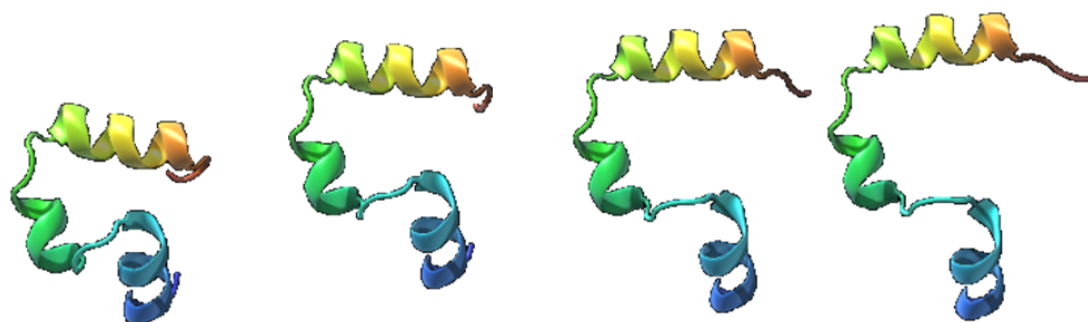


*Figure 9. Transition obtained by the ARAP method of a subdomain of the villin headpiece (protein ID: 1YFR) into its distorted shape generated manually.*

## 7.7. Automatic parameterization for the Universal Force Field

**Participants:**  Svetlana Artemova, Leonard Jaillet, Stephane Redon.

We have continued working on the integration of the Universal Force Field in SAMSON. This force field is a classical non-reactive force field that has parameterizations for all atoms of the periodic table with atomic number lower than 103. Our implementation of this force field includes a new automatic perception scheme for molecular systems that is specifically-tailored for UFF, as well as several corrections and refinements that have been lately proposed in the literature. We have tested this implementation on more benchmarks and improved its computational performance. Additionally, we have compared our implementation to that of the OpenBabel toolbox. As a result, our self-contained implementation was integrated in a new module for SAMSON and is now available on SAMSON-Connect website (see Figure 10 ). The paper describing the obtained results will appear in the Journal of Computational Chemistry.
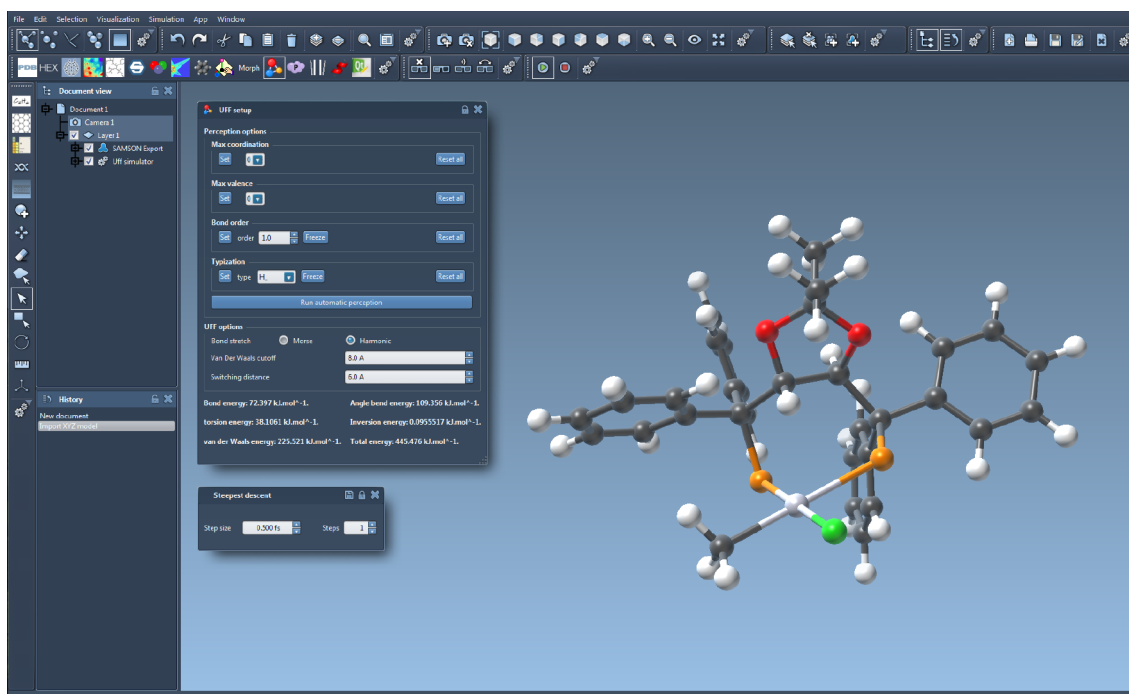


*Figure 10. A molecule being interactively manipulated in SAMSON thanks to the UFF module. The interface of the UFF module allows to setup UFF. The upper part of the interface proposes options to manually adjust the perception of the molecular system. The middle part proposes the UFF options. The lower part prints out each energy contribution with the resulting total energy.*

## 7.8. Interactive modeling with the Universal Force Field

**Participants:**  Leonard Jaillet, Svetlana Artemova, Stephane Redon.

In parallel with the classical Universal Force Field, we have continued working on an extension of this force field that we call Interactive Modeling UFF (IM-UFF). In classical UFF topologies and atoms' typizations are set in the initialization phase and remain fixed for the entire simulation. IM-UFF, on the contrary, allows soft transitions for both topologies and atoms' typizations. This new approach, thus, combines the possibility to

significantly modify molecular structures (as with reactive force fields) with a broad diversity of supported systems thanks to the universality of UFF. Such an extension lets the user easily build and edit molecular systems interactively while being guided by physically-based inter-atomic forces. The validity of this extended version of UFF was tested on the same large set of benchmarks as those used to test classical UFF, and the results of both approaches were compared.

## 7.9. Error Analysis of Modified Langevin Dynamics

**Participants:** Zofia Trstanova, Gabriel Stoltz, Stephane Redon.

Adaptively Restrained Particles Simulations (ARPS) were recently proposed with the purpose of speeding up molecular simulations. The main idea is to modify the Hamiltonian such that the kinetic energy is set to zero for small velocities, which allows to save computational time since particles do not move and forces need not be updated. ARPS can be combined with Langevin dynamics in order to speed up the computation of macroscopic quantities.

The aim of this work is to understand how simulation errors depend on the parameters of the method. We distinguish the statistical error and the systematic error related to the finiteness of the time step $\Delta t$. The statistical error is controlled by variance, that is given by

$$\sigma^2 = -2 \left\langle A - \mu(A), \mathcal{L}^{-1} \left( A - \mu(A) \right) \right\rangle_{L^2(\mu)} \tag{1}$$

where $\mu$ is the invariant measure, $\mathcal{L}$ is the generator of the stochastic process and $A$ an observable. First we demonstrate by use of weighted $L^\infty$ estimations that the ARPS-Langevin dynamics are well defined. In the main part of this work, we quantify the increase of variance of the ARPS-Langevin process as a function of the ARPS parameters. For small parameters, we express the generator of the ARPS-Langevin dynamics as a perturbed generator of the Langevin dynamics, and study the asymptotic expansions of the variance (1 ) in the restrained dynamics parameter $\varepsilon$.

$$\sigma_\varepsilon^2 = \sigma^2 + \mathcal{O}(\varepsilon)$$

For large values of $\varepsilon$, we perform numerical simulations. For a simple 1D system we approximate $\mathcal{L}^{-1}$ by Galerkin approach and for higher-dimensional systems we discretize the stochastic differential equations by a second order method and analyze a model of a dimer surrounded by solvent particles.

## 7.10. Algorithmic speed up of the ARPS method

**Participants:** Zofia Trstanova, Gabriel Stoltz, Stephane Redon.

Adaptively Restrained Particles Simulations (ARPS) allow to save computational time at each time step since particles do not move and forces need not be updated. The associated gain can be quantified by an algorithmic speed-up factor $S_{\text{algo}} \geq 1$. Intuitively, freezing more particles leads to larger algorithmic speed-ups, but also larger correlations in time.

We analyzed the algorithmic speed up with respect to the standard methods. Since the ARPS algorithm is based on adding and subtracting of the forces between active particles, the gain with respect to the standard method, where only one complete computation of all interactions is performed at each time step, is achieved only if the percentage of restrained particles is big enough. Hence we studied the necessary conditions, under which the computational complexity of the forces updating in the ARPS method is lower than the one of the standard method. This allows to achieve an algorithmic speed up that is always bigger than one.

We also propose a simple strategy for choosing optimal simulation parameters.

## 7.11. Numerical analysis for the ARPS method

**Participants:** Zofia Trstanova, Gabriel Stoltz, Stephane Redon.

Previous works have led to understanding of the choice of optimal parameters for the ARPS dynamics. The interest lies in achieving the highest percentage of restrained particles, while minimizing the modification of the variance and the systematic error. We study discretization schemes of the ARPS-Langevin dynamics, such that the systematic error remains of second order in the time step size and we introduce a Metropolis step in order to stabilize the simulations and hence to allow "a sharper" choice of the ARPS parameters, which lead to better algorithmic speed-ups.

## 7.12. New rendering algorithm for secondary structures

**Participants:** Marc Aubert, Stephane Redon.

We developed a new algorithm for rendering secondary structures of proteins (Figure 11 ). The method relies on the determination of the most probable secondary structure elements (e.g. alpha helices and beta sheets) based on geometrical features of a protein. After construction of control points on the CPU, the method generates triangles directly on the Graphics Processing Unit (GPU) through geometry shaders. The number of generated triangles may be adaptively chosen based on e.g. the camera distance and the desired resolution. The secondary structure algorithm and the rendering algorithm are both fast enough to allow for interactive modification of the protein (e.g. thanks to As-Rigid-As-Possible editing algorithms).
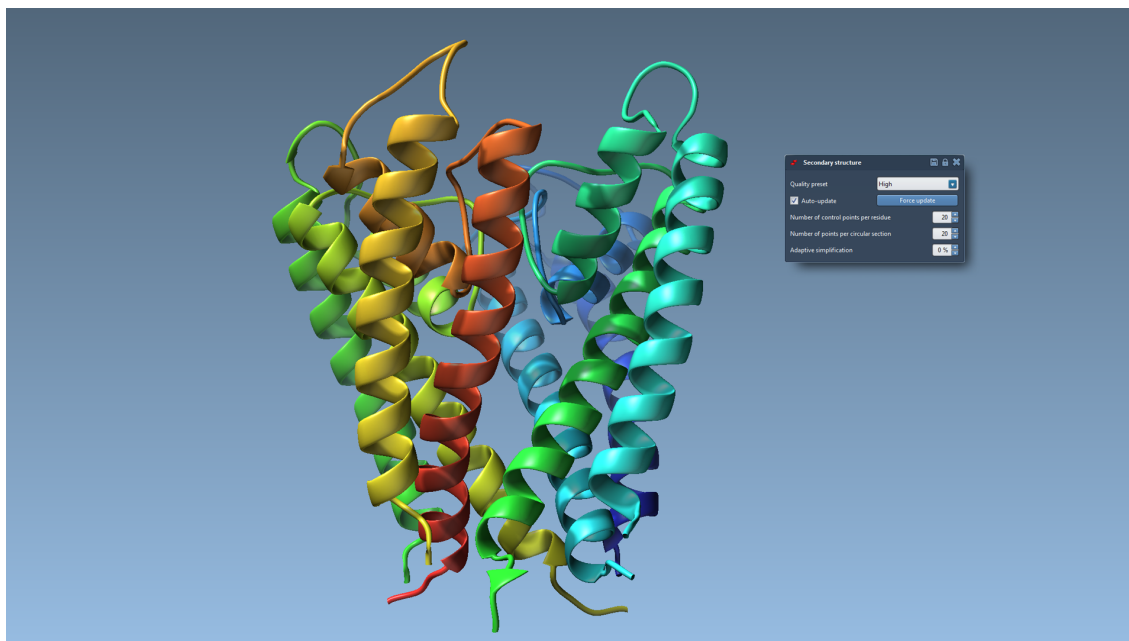


*Figure 11. Protein secondary structure rendering on GPUs.*

## 7.13. Property models

**Participants:** Marc Aubert, Stephane Redon.

We extended the hierarchy of classes in SAMSON for *property models*. Property models are one of the five categories of models in SAMSON, with structural models (for geometry and topology), visual models (for custom graphical representations), dynamical models (to describe degrees of freedom) and interaction models (to represent energies and forces). We have added classes to easily represent in SAMSON various functions, fields (e.g. scalar fields and vector fields), etc. These property models are template classes which may rely on the unit system of SAMSON to perform dimensional analysis at compile time.

## 7.14. Integration of tools in SAMSON

**Participants:** Nadhir Abdellatif, Stephane Redon.

Thanks to funding from the Nanosciences Foundation in Grenoble, we developed SAMSON Elements (modules for SAMSON) that integrate existing tools. In particular, we integrated OpenBabel, a tool to convert between numerous molecular formats (Figure 12 ), ClustalW, a tool for sequence alignment (Figure 13 ), and Pepsi-SAXS, a tool for SAXS developed in the group (Figure 14 ).
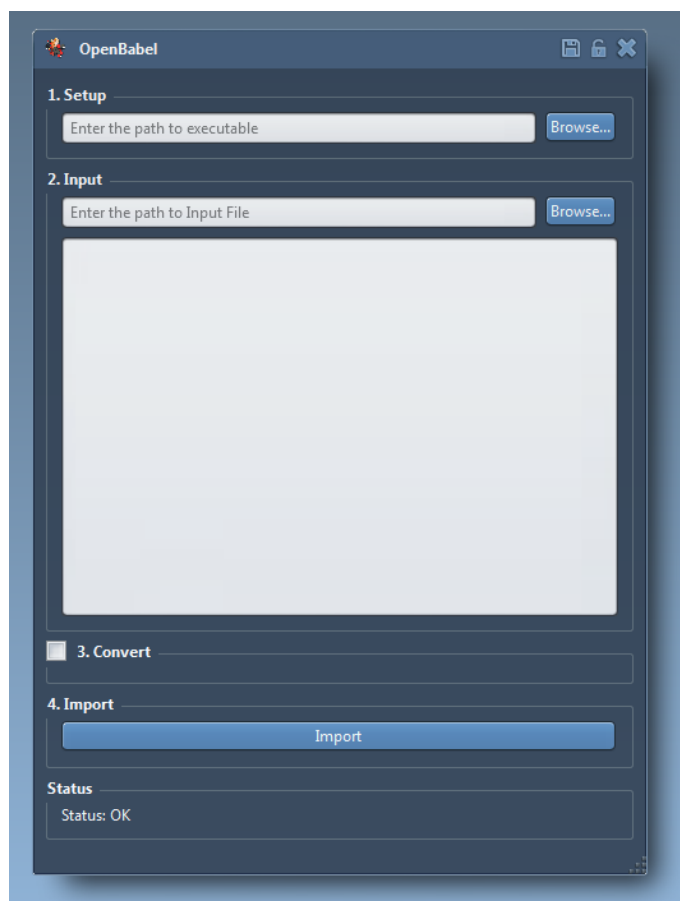


*Figure 12. The OpenBabel connector in SAMSON*

## 7.15. Development of SAMSON Connect

**Participants:** Mohamed Yengui, Jocelyn Gate, Stephane Redon.

We have continued the development of SAMSON Connect (Figure 15 , https://www.samson-connect.net), the online platform for distributing SAMSON and SAMSON Elements (modules for SAMSON). SAMSON Connect is a web application, associated to a database, that functions as the well-known stores for mobile Apps (e.g. Google Play, the Apple App store, etc.). Users may create an account, download SAMSON, and add SAMSON Elements to their configuration based on their needs (Figure 16 ). Adding a SAMSON Element is performed in just one click (to the Add button of the corresponding SAMSON Element), and the SAMSON
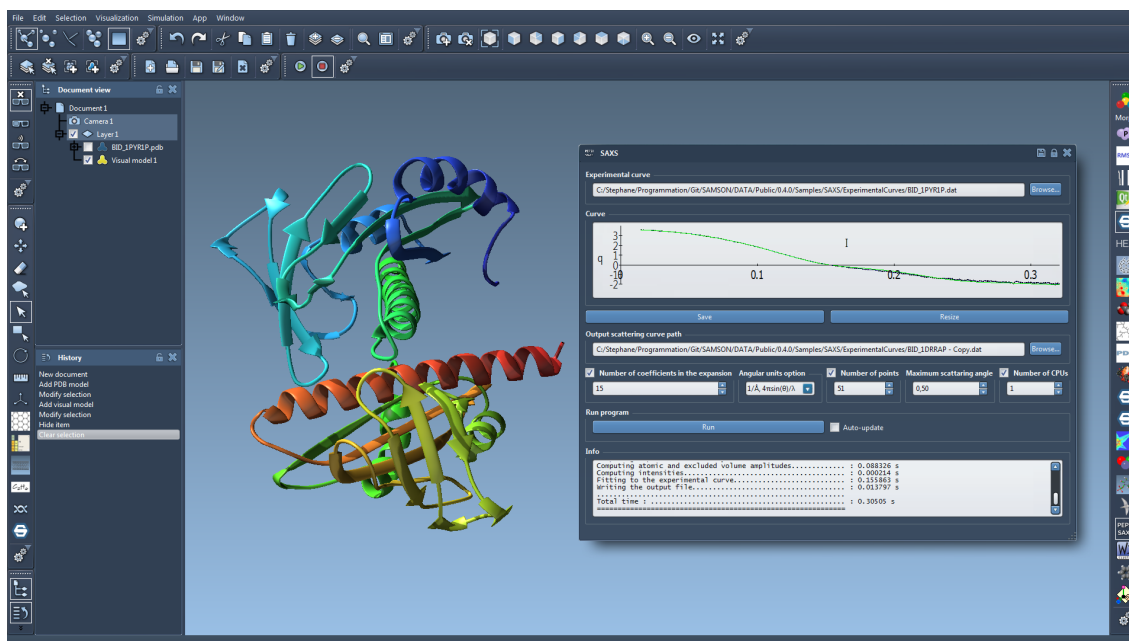
*Figure 13. ClustalW in SAMSON*

*Figure 14. Our Pepsi-SAXS tool integrated in SAMSON*

Element is installed when the user restarts SAMSON. Users may also request an upgrade to a Developer status, after which they can download the SAMSON Software Development Kit used to develop SAMSON Elements. They may then upload their SAMSON Elements to SAMSON Connect in order to share them. The platform opened in March 2015 to release the first beta version of SAMSON. We also produced some video tutorials for SAMSON (Figure 17 ).

On the back-office of SAMSON Connect, we added several functionalities that facilitate publishing new versions of SAMSON and SAMSON Elements (e.g. choosing default SAMSON Elements), email users based on their account type (user, developer, etc.), OS, etc. We also turned to automatic acceptance of new user accounts (once they validate their email address). We also updated the SAMSON web service to enable more message types and retrieve information about the server, the database, etc.

## 7.16. Documenting the SAMSON Software Development Kit

**Participants:** Stephane Redon, Jocelyn Gate, Svetlana Artemova.

The SAMSON Software Development Kit (SDK) is at the core of the SAMSON platform and makes it possible to develop SAMSON Elements (modules). The API of SAMSON contains numerous classes and allows for a variety of modules types (e.g. parsers, force fields, visual models, integrators, apps, editors, etc.), and provides several non-elementary mechanisms (e.g. a unit system, a signals and slots mechanism, memory management, data structures for incremental calculations, etc.). We continued writing the SDK documentation accessible to SAMSON Elements developers. The current PDF version for the beta 0.4.0 version has passed 500 pages.

## 7.17. SAMSON SDK Helpers
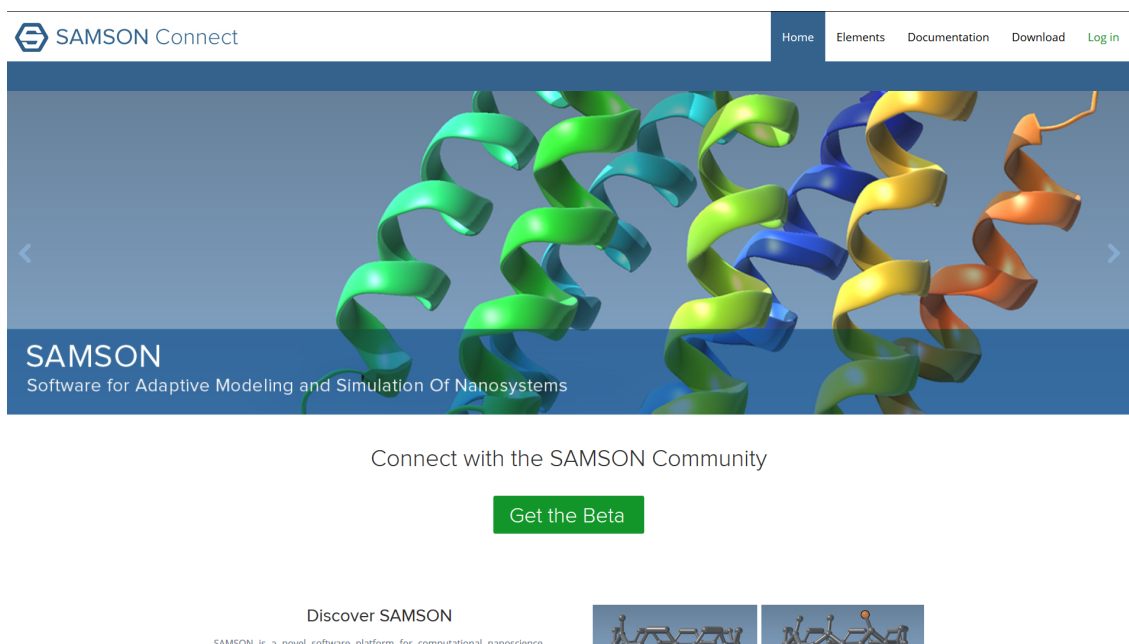
**Participants:** Jocelyn Gate, Stephane Redon.

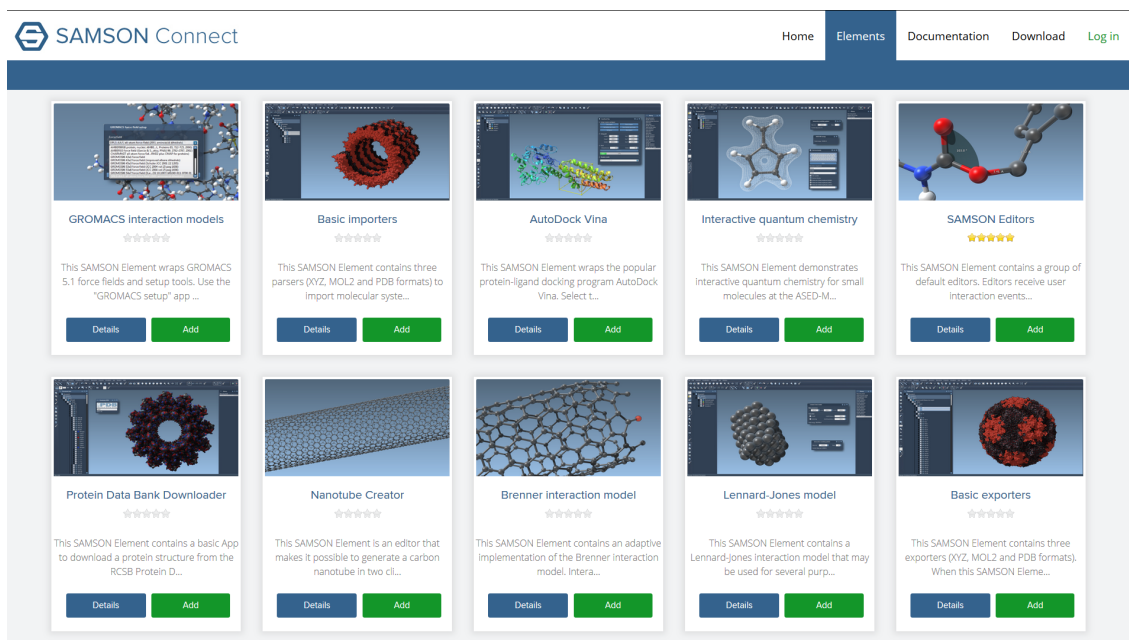*Figure 15. The home page of SAMSON Connect (https://www.samson-connect.net)*



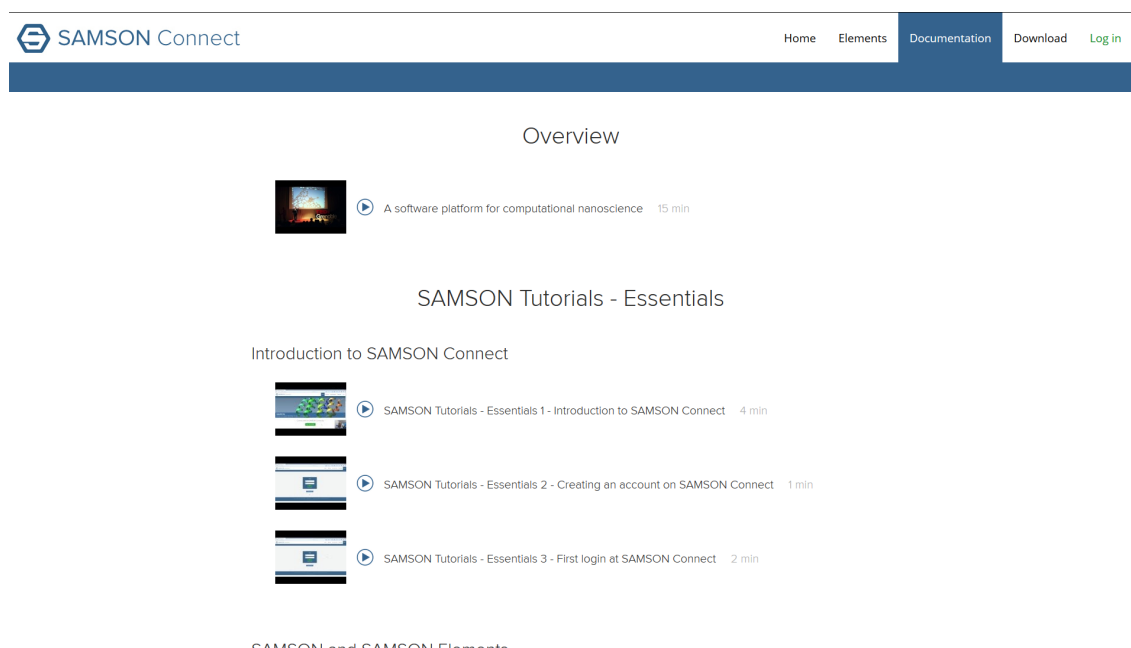*Figure 16. The Elements page, where users may add SAMSON Elements (modules) to their configuration (https://www.samson-connect.net)*

*Figure 17. The documentation page on SAMSON Connect (https://www.samson-connect.net)*

We have developed a number of Helpers in the SAMSON SDK, in order to facilitate the development of SAMSON Elements (modules for SAMSON). For example, the SAMSON Element generator (Figure 18 ) generates code that immediately compiles and runs, and that developers may complete, for a number of SAMSON Classes.
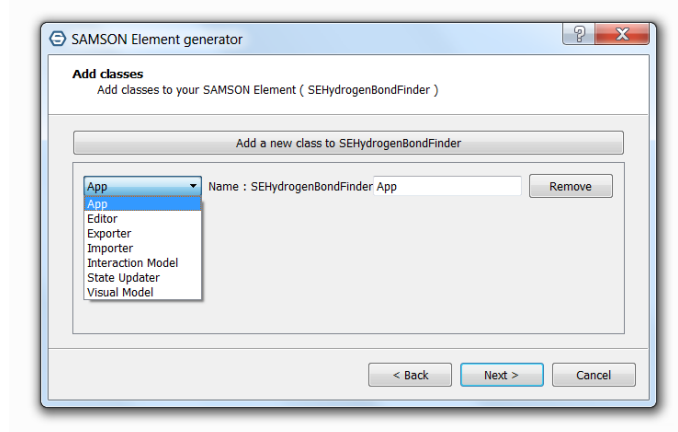


*Figure 18. The SAMSON Element Generator makes it easy to develop modules for SAMSON*

We have also developed for the group a helper able to upload numerous SAMSON Elements to SAMSON Connect at the same time (Figure 19 ), which is especially useful given the rapidly growing number of modules being developed in the team.
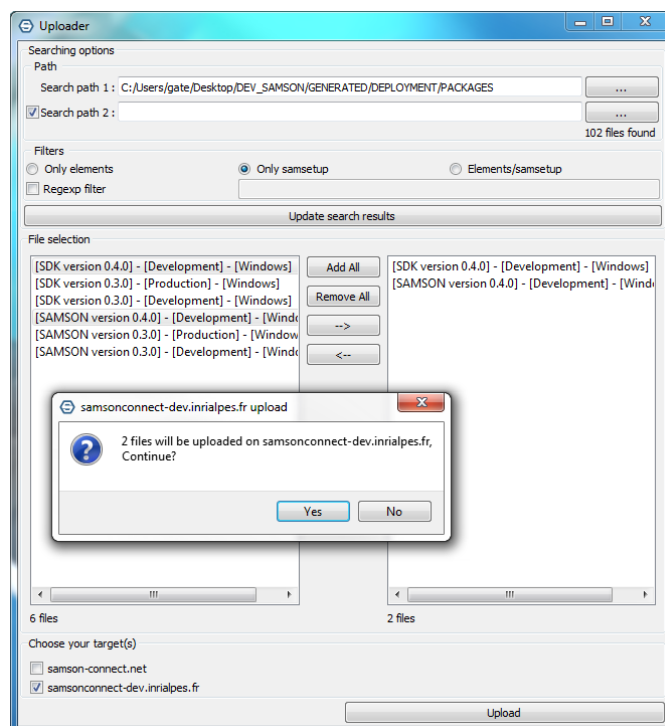


*Figure 19. The SAMSON Element uploader eases the transfer of multiple SAMSON Elements to SAMSON Connect*

## 7.18. Pepsi-SAXS : an adaptive method for rapid and accurate computation of small angle X-ray scattering profiles

**Participant:**  Sergei Grudinin.

We developed a new method called Pepsi–SAXS that calculates small angle X-ray scattering profiles from atomistic models. Our method is based on the multipole expansion scheme and is significantly faster and more precise compared to other tested methods. In particular, using the Nyquist–Shannon–Kotelnikov sampling theorem, we adapt the multipole expansion order to the size of the model and the resolution of the experimental data. We argue that using the adaptive expansion order, our method has the same quadratic dependence on the number of atoms in the model as the Debye-based approach, however, with a much smaller prefactor in the computational complexity.

We have systematically validated our method on an excessive set of over fifty models collected from the BioIsis and SASBDB databases. Using a laptop, we demonstrated that Pepsi-SAXS is about 9, 33 and 43 times faster compared to CRYSOL, FoXS and the 3D-Zernike method in SAStbx, correspondingly, when tested on data from the BioIsis database, and is about 5, 18 and 23 times faster compared to CRYSOL, FoXS and SAStbx, correspondingly, when tested on data from SASBDB. On average, Pepsi-SAXS achieves 17% smaller value of $\chi$ compared to CRYSOL and 15% smaller value of $\chi$ compared to FoXS for BioIsis profiles, and 6% smaller value of $\chi$ compared to CRYSOL and 19% smaller value of $\chi$ compared to FoXS for SASBDB profiles.

## 7.19. Knodle: a Support Vector Machines-based automatic perception of organic molecules from 3D coordinates

**Participants:** Maria Kadukova, Sergei Grudinin.

We addressed the problem of the assignment of atom types and bond orders in low molecular weight compounds. For this purpose, we developed a prediction model based on nonlinear Support Vector Machines (SVM), implemented in a KNOwledge-Driven Ligand Extractor called *Knodle*, a software library for the recognition of atomic types, hybridization states and bond orders in the structures of small molecules. We trained the model using an excessive amount of structural data collected from the PDBbindCN database. Accuracy of the results and the running time of our method is comparable with other popular methods, such as NAOMI, fconf, and I-interpret. More precisely, on the popular Labute's benchmark set consisting of 179 protein-ligand complexes, *Knodle* makes five to six perception errors, NAOMI makes seven errors, I-interpret makes nine errors, and fconv makes thirteen errors. On a larger set of 3,000 protein-ligand structures collected from the PDBBindCN general data set (v2014), *Knodle* along with NAOMI have a comparable accuracy of approximately 6 % of errors, whereas fconv produces approximately 13 % of errors. Overall, our study demonstrates the efficiency of nonlinear SVM in structure perception tasks.

## 7.20. Symmetry Detection Method

**Participants:** Silvia Dias Pinto, Sergei Grudinin.

We developed an algorithm for automatic recognition of the point group symmetry in electron density maps of biological objects. More precisely, the method operates on cryo-Electron Microscopy (cryoEM) data, which typically contain 3D structures of multi-domain proteins and their complexes. We represent the shape using a spherical harmonic decomposition and then operate on the expansion coefficients to quantify the structural symmetry thanks to a mismatch function. Overall, we developed new mathematical and computational frameworks for symmetry detection using the polynomial expansion approach.

## 7.21. Pepsi-Dock: fast predictions of putative docking poses using accurate knowledge-based potentials functions to describe interactions between proteins

**Participants:** Emilie Neveu, Sergei Grudinin, David W. Ritchie, Petr Popov.

Many biological tasks involve finding proteins that can act as an inhibitor for a virus or a bacteria, for example. Such task requires knowledge on the structure of the complex to be formed. Protein Data Bank can help but only a small fraction of its proteins are complexes [16]. Therefore, computational docking predictions, being low-cost and easy to perform, are very attractive if they describe accurately the interactions between proteins while being fast to find which conformation will be the most probable. We have been developing a fast and accurate algorithm that combines the FFT-accelerated docking methods [67] with a precise knowledge-based potential functions [58] describing interactions between the atoms in the proteins .

Interactions between proteins follow complex and non-linear laws which computation is time-consuming. It is of common usage to start the predictions with a simple, approximated, expression of these interactions to then reduce the space search in order to use more complex laws. However we think it is important to use the most accurate free energy not to miss some important docking solutions. Thus, our aim is to integrate the very-detailed knowledge-based potentials into the *Hex* code and to take advantage of its exhaustive search, which is by now still the most efficient and reliable search algorithm [67] .

Last year, we adapted the machine learning process so that the knowledge-based potentials describing atom interactions can be translated into the polynomial basis used in *Hex*. The current evaluations of the knowledge-based scores takes more time than a shape+electrostatic representation but is still fast: exploring $10^9$ conformations of a complex takes on average 5-10 minutes on a regular laptop computer.

This year, we run cross-validation experiments and tested different data sets in order to improve the predictions. Using bound conformations of each proteins to make the predictions, we retrieve up to 70% correct complexes of about 200 complexes. Results show that the knowledge-based potentials, while being general, correctly predict the interactions. Even better results could be achieved without the limitations in the search range by the spherical sampling grid which lacks of precision far away of its origin. Because many complexes have separation distances greater than 30 Å, we are now working on a multi-centre definition of the potentials in order to correctly predict the structures of protein complexes starting from their unbound structures.

## 7.22. Pepsi-Piper: rigid docking predictions using Pepsi potentials into Piper code

**Participants:** Sergei Grudinin, Emilie Neveu, Dima Kozakov, Dzmitry Podgorny.

This work is the continuation of the Pepsi-Dock project that aims to develop fast predictions of putative docking poses using accurate knowledge-based potentials functions to describe interactions between proteins. The goal is to integrate the precise, and yet easy to compute, distance-based pairwise knowledge-based potentials [58] into the Piper search code [48] in order to compare its exhaustive search with the *Hex* one. The former samples the conformations using a cartesian grid while the latter, a spherical one. We proved our potential used in *Hex* can predict the structures of complexes with a really good success rate, the main limitation being the lack of precision of the spherical sampling when the separation distance of the two proteins is too large. We think predicting docking combining our potential and a sampling search based on a cartesian grid as in Piper will achieve greater results, but will require more computational time.
We first adapt our potential to the Piper code and showed that the ranking results on the data set used for training are better than the ranking provided by Piper [25]: when the potential is used to sort the conformations, the correct solution is found in the first ten for 85% cases, while Piper found it in only 25% cases. The next step is to use the cartesian sampling to make docking predictions. When the Piper code will be ready to integrate our potential, we will be able to confront with other knowledge-based potentials such as the one initially used in Piper, DARS.

## 7.23. Flex-Dock: towards flexible docking predictions using metaheuristics optimisation methods

**Participants:** Emilie Neveu, Sergei Grudinin, Alexandre Hoffman, Angelo Migliosi, Xavier Besseron, Grégoire Danoy, Pascal Bouvry.

Docking numerical methods are used to predict the preferred location of one molecule with respect to the second when bound to each other. This is particularly useful for the design of drugs that inhibit the effects of viruses or bacteria. However proteins change their conformation upon binding and searching for flexible conformations involves enormous degrees of freedom and complex physics. Thus, the prediction of realistic interactions with full flexibility of the two partners is an intractable global optimisation problem.
There are currently several algorithms that produce high quality predictions of molecular complexes [43]. But very few manages to deal with the flexibility of the proteins. A common method is to refine the most probable predicted rigid complexes with a scoring allowing for flexibility [81]. Here, we want to tackle flexibility and sampling all together. Exhaustive search methods, which were by now the most accurate optimisation method for relatively small molecules [53] will be too time-consuming when it comes to large proteins. There is a strong need to explore and define new optimisation algorithms such as metaheuristic ones that can deal with several local minima and a large minima and a large search space. The main goal of this project is to define the problem and find for the optimisation method that will potentially give better results than the actual reference, SwarmDock [54].
We worked on a first comparison of several evolutionary-based algorithms (Genetic Algorithm [40], Differential Evolution [76], Particle Swarm Optimisation [46]) using rigid proteins only and on the use of multi-objective algorithms when the proteins are flexible.

To take into account flexibility, we approximate large-scale deformations of each proteins using an elastic network model combined with a low-frequency approximation called normal mode analysis such as in [81] or in [54]. Combined with the rigid transformation between the two proteins, it defines a complete while reduced set of degrees of freedom to search for.

The scoring function has to discriminate correct conformations from impossible ones. Our scoring is the main difference with SwarmDock. It takes into account the energy gained by docking using the precise knowledge-based potentials derived in [58], whereas only a simple physics-based energy is used in SwarmDock. We also want to explore another scoring that will also add the energetic cost of each moves of the proteins. To do so, we started to develop multi-objective algorithms. Combined with a Pareto Front analysis, this will help us to validate the scoring and to compare different evolutionary-based algorithms.Tests will be directly made on the Protein-Protein Benchmark [42] so that we can compare with other docking methods.

## 7.24. FastRMS: rapid determination of RMSDs corresponding to macromolecular rigid body motions, adding flexibility via collective motions

**Participants:** Sergei Grudinin, Petr Popov, Emilie Neveu.

Computing the root mean sum of squared deviations (RMSDs) between two sets of coordinates each describing a different conformation of a macromolecule is a necessary step in many structural bioinformatics and molecular modelling technics to assess structural predictions [43], identify biding sites [49] or structurally classify proteins. A straightforward and universally-used method determines the RMSD with a computational complexity proportional to the number of atoms in the molecule. We recently presented RigidRMSD, a fast algorithm that determines RMSDs corresponding to a set of rigid body motions of a macromolecule in constant time with respect to the number of atoms in the molecule [57]. Here, we extend it to proteins with flexibility modelled with collective motion such as an elastic network model combined with normal mode analysis.

With these new assumptions, the complexity of the algorithm depends linearly or quadratically with the number of collective motion vectors selected to approximate the flexibility. The typical number of vectors needed to have accurate flexible movements being much lower than the number of atoms composing the molecules, we prove our algorithm is still faster than the common method. Our algorithm is particularly useful for rigid body modelling applications such as rigid body docking procedures allowing for flexibility via collective motions: clustering, high-throughput analysis and simulation results [49], [26], [59]. A C++ implementation of our algorithm will be soon available at http://nano-d.inrialpes.fr/software/RigidRMSD.

## 7.25. SAM : Spherical Polar Fourier Assembly of Protein Complexes with Arbitrary Point Group Symmetry

**Participants:** David W. Ritchie, Sergei Grudinin.

We presented a novel FFT-based *ab inito* docking algorithm called "SAM" for building perfectly symmetrical models of protein complexes with arbitrary point group symmetry. The basic approach uses a novel and very fast 1D symmetry-constrained spherical polar Fourier search to assemble cyclic $C_n$ systems from a given protein monomer. Structures with higher order ($D_n, T, O$, and $I$) point group symmetries may be built using a subsequent symmetry-constrained Fourier domain search to assemble trimeric sub-units. Our results show that the SAM algorithm can correctly assemble monomers of up to around 500 residues to produce a near-native complex structure with the given point group symmetry in 17 out of 18 test cases. The SAM program may be downloaded for academic use at *http://sam.loria.fr/*.

## 7.26. KSENIA : Knowledge of Native Protein-Protein Interfaces is Sufficient to Construct Predictive Models for the Selection of Binding Candidates

**Participants:** Petr Popov, Sergei Grudinin.

Selection of putative binding poses is a challenging part of virtual screening for protein-protein interactions. Predictive models to filter out binding candidates with the highest binding affinities comprise scoring functions that assign a score to each binding pose. Existing scoring functions are typically deduced collecting statistical information about interfaces of native conformations of protein complexes along with interfaces of a large generated set of non-native conformations. However, the obtained scoring functions become biased toward the method used to generate the non-native conformations, i.e. they may not recognize near-native interfaces generated with a different method.

Present study demonstrates that knowledge of only native protein-protein interfaces is sufficient to construct well-discriminative predictive models for the selection of binding candidates. Here, we introduce a new scoring method that comprises a knowledge-based potential called *KSENIA* deduced from the structural information about the native interfaces of 844 crystallographic protein-protein complexes. We derive *KSENIA* using convex optimization with a training set composed of native protein complexes and their near-native conformations that are obtained using deformations along the low-frequency normal modes. As a result, our knowledge-based potential has only marginal bias toward a method to generate putative binding poses. Furthermore, *KSENIA* is smooth by construction, which allows to use it along with a rigid-body optimization to refine the binding poses. Using several test benchmarks we demonstrate that our method discriminates well native and near-native conformations of protein complexes from the non-native ones. Our methodology can be easily adapted to the recognition of other types of molecular interactions, such as protein-ligand, protein-RNA, etc. *KSENIA* will be made publicly available as a part of the SAMSON software platform at https://www.samson-connect.net.

## 7.27. Predicting Binding Poses and Affinities in the CSAR 2013–2014 Docking Exercises Using the Knowledge-Based Convex-PL Potential

**Participants:** Sergei Grudinin, Petr Popov, Emilie Neveu, Georgy Cheremovskiy.

The 2013–2014 CSAR docking exercise was the opportunity to assess the performance of the novel knowledge- based potential we are developing, named Convex-PL. The data used to derive the potential consists only of structural information from protein-ligand interfaces found in the PDBBind database. As expected, our potential proved to be very efficient in the near- native pose detection exercises, where we correctly predicted two near-native poses in the 2013 exercise and also ranked 22 near- native poses first and 2 second in the 2014 exercise. Somewhat more surprisingly, we obtained a fair performance in some of the CSAR affinity ranking exercises, where the Spearman correlation coefficients between our predictions and the experiments are greater than 0.5 for several protein–ligand sets. Nonetheless, affinity prediction exercises turned out to be a challenge, and significant progress in the development of our method is needed before we can successfully predict binding constants.

## 7.28. Prediction of homo- and hetero-protein complexes by ab-initio and template-based docking: a CASP-CAPRI experiment

**Participants:** Sergei Grudinin, Petr Popov, Emilie Neveu.

We present the results for CAPRI Round 30, the first joint CASP-CAPRI experiment, which brought together experts from the protein structure prediction and protein-protein docking communities. The Round comprised 25 targets from amongst those submitted for the CASP11 prediction experiment of 2014. The targets included mostly homodimers, a few homotetramers, and two heterodimers, and comprised protein chains that could readily be modeled using templates from the Protein Data Bank. On average 24 CAPRI groups and 7 CASP groups submitted docking predictions for each target, and 12 CAPRI groups per target participated in the CAPRI scoring experiment. In total more than 9500 models were assessed against the 3D structures of the corresponding target complexes. Results show that the prediction of homodimer assemblies by homology modeling techniques and docking calculations is quite successful for targets featuring large enough subunit interfaces to represent stable associations. Targets with ambiguous or inaccurate oligomeric state assignments, often featuring crystal contact-sized interfaces, represented a confounding factor. For those,

a much poorer prediction performance was achieved, while nonetheless often providing helpful clues on the correct oligomeric state of the protein. The prediction performance was very poor for genuine tetrameric targets, where the inaccuracy of the homology-built subunit models and the smaller pair-wise interfaces severely limited the ability to derive the correct assembly mode. Our analysis also shows that docking procedures offer a clear advantage over standard homology modeling techniques and that highly accurate models of the protein components are not always required to identify their association modes with acceptable accuracy.

Most of the targets in Round 30 of CAPRI were homodimers and homotetramers, thus it was a good opportunity to test our novel symmetry assembling docking method. To do so, we imposed C2 symmetry constraints for all the homodimers and we imposed C4 and D2 symmetry constraints for all the homotetramers from the target complexes. Below, we present the new fast multi-resolution method for docking both symmetric and non-symmetric protein complexes that was used in Round 30 of CAPRI. First, the structures of the individual subunits were taken from the stage two predictions of the CASP10 assessment experiment. More precisely, starting from 150 available CASP 3D models of monomers, we predicted models of symmetric multimers using the novel symmetry docking method, which performs symmetry-induced protein docking using the shape-complementarity scoring function computed as spherical polar Fourier correlations. Specifically, this method performs exhaustive search over the available (four in case of cyclic symmetries or six otherwise) degrees of freedom for the given point group symmetry type. For the targets of Round 30 of CAPRI we imposed three types of symmetry, C2, C4, and D2. For the case of heterodimers, we used the standard Hex docking method.

For the input of the docking methods, we generated the scaffolds of initial models of monomers by cutting-off the side chains. More specifically, we mutated all side-chains except for the glycines to alanines. Compared to the standard all-atom rigid-body docking methods, we expect the scaffold docking approach to produce binding poses that are less sensitive to the flexibility of the side-chains. We clustered the solutions with the threshold ligand-RMSD value of 8 Åusing the RigidRMSD library. Finally, we ranked the clusters by the value of the best score and kept 50 best clusters for the refinement stage. In total, for each target we proceeded to the refinement with 7,500 modeled structures of protein complexes.

On the next step, we optimized each putative binding interface of the all-atom representation of a protein complex by means of a rigid-body first-order minimization scheme. Specifically, after each rigid-body minimization step we proceeded with the optimization of side-chains described by the rotameric representation using the SCWRL4 package. We computed the interactions between the subunits in a protein complex using the novel reference state-free knowledge-based scoring function KSENIA, which is smooth by construction and is thus very suitable for a gradient-based minimization protocol. Finally, we ranked the predictions by the value of the KSENIA potential of the optimized structure and selected ten best candidates for the submission.

## 7.29. Convex relaxation for non-convex quadratic optimization problems with applications to side-chain prediction in protein structures

**Participants:** Aleksandr Katrutsa, Sergei Grudinin.

The side-chain prediction problem is the major part of the more general protein structure prediction problem, which is very important for drug design and in the prediction of stable protein mutations. Formally, the side-chain prediction problem states in the form of discrete quadratic optimization problem with an indefinite matrix in the quadratic term,

$$\mathbf{x}^\top \mathbf{Q} \mathbf{x} + \mathbf{b}^\top \mathbf{x} \to \min_{\mathbf{x} \in \{0,1\}^n} \qquad (2)$$

This problem is NP-hard, so to get a good approximation solution we used convex semidefinite relaxation with different types of constraints. This approach is the powerful optimization technique that helps to reformulate the initial non-convex problem as a convex one and sometimes even gives the exact solution. The important step is to operate with precise energy function, which is used to compute the energy of different interactions

in proteins. To obtain this, we used the machine-learning procedure, which extracts the parameter vector for the potential from the training set of protein structures. After the training step, we used this vector to compute the energy of a protein and to find the side-chains corresponding to the minimal total energy of the protein. The current accuracy in side-chain prediction is about $80\%$, which is achieved using the spectrum relaxation of the matrix in the quadratic term. Also, this approach is very fast, precisely, it requires less than 1 second per protein to predict the positions of its side-chains.

## 7.30. Critical assessment of protein-ligand docking methods via Drug Design Data Resource (D3R) Challenge 2015

**Participants:** Andreas Eisenbarth, Sergei Grudinin.

We participated at the Drug Design Data Resource (D3R) Challenge 2015. In the challenge, we were given protein structures and sets of ligand molecules in order to detect the putative binding poses. The aim was to find the energetically most favourable pose of each ligand relative to a protein. To do so, we first performed the docking simulations using the state-of-the-art software AutoDock Vina, then explored sets of parameters that produced chemically reasonable poses, and finally did the re-scoring using the ComvexPL potential. Later, we critically examined AutoDock Vina sampling method and detected points where it can be improved and also assessed the integration of our inhouse developed ConvexPL scoring algorithm.

## 7.31. Towards the development of FFT-accelerated flexible fitting methods

**Participants:** Alexandre Hoffmann, Valerie Perrier, Sergei Grudinin.

We studied a set of new methods for non-rigid molecular fitting. The problem can be formulated as follows : Let $\mathcal{P}_1$ and $\mathcal{P}_2$ be two molecular structures (e.g. proteins). We are given $d_1 : \mathbb{R}^3 \mapsto \mathbb{R}$, the electron density of $\mathcal{P}_1$ and $(Y_k \in \mathbb{R}^3)_{k=1\cdots N_{atoms}}$, the average positions of the atoms of $\mathcal{P}_2$. Assuming we can generate an artificial electron density $d_2 : \mathbb{R}^3 \mapsto \mathbb{R}$ from $(Y_k \in \mathbb{R}^3)_{k=1\cdots N_{atoms}}$, our problem is to find a transformation of the atoms $T : \mathbb{R}^{3N_{atoms}} \mapsto \mathbb{R}^{3N_{atoms}}$ that minimizes the $L^2$ distance between $d_1$ and $d_2$.

In image processing, this problem is usually solved using the optimal transport theory, but this method assumes that both densities have the same $L^2$ norm, which is not necessarily the case for the fitting problem. To solve this problem, one instead starts by splitting $T$ into a rigid transformation $T_{rigid}$ (which is a combination of translation and rotation) and a flexible transformation $T_{flex}$. Two classes of methods have been developed to find $T_{rigid}$ :

- the first one uses optimization techniques such as gradient descent,
- the second one uses the Fast Fourier Transform (FFT) to compute the Cross Correlation Function (CCF) of $d_1$ and $d_2$.

We have developed several algorithms based on the FFT to find $T_{rigid}$ and we have developed two algorithms for flexible molecular fitting that are based on convex and non-convex optimization and the trust region methods. Our tests demonstrate that while one method gives good results for small deformations, the other gives good results for bigger deformations.

We have been also improving the current NMA method (which is essentially a model reduction technique), that is used in other tools such as the flexible fitting to small angle scattering profiles. Finally, we started the development of a method for a harder fitting/docking problem in which only electron density would be known. The basic idea would be to find the $C^1$-diffeomorphism $T : \mathbb{R}^3 \mapsto \mathbb{R}^3$ that minimizes the $L^2$ distance between $d_1$ and $d_2$.

We developed several stand-alone C++ libraries to solve some of our problems including:

- a non-convex optimization library,
- a normal mode analysis library,
- a fitting library that implements our new methods.

<span style="color:red">NECS Project-Team</span>

# 7. New Results

## 7.1. Network systems and graph analysis

### 7.1.1. Distributed estimation of graph Laplacian eigenvalues

**Participants:** A. Kibangou [Contact person], T.-M. D. Tran.

Linear average-consensus is a well-known iterative protocol allowing agents to converge to the average of initial values by taking suitable convex combinations of the messages received from neighbors. From the recent literature, it is known that, after a finite time, some consecutive measurements of a state of the consensus dynamical system can be used to compute the exact average of the initial condition. In [23], we have shown that these measurements can also be used for estimating the Laplacian eigenvalues of the graph representing the network. As recently shown in the literature, by solving the factorization of the averaging matrix, the Laplacian eigenvalues can be inferred. In our paper, the problem is posed as a constrained consensus problem. A first formulation (direct approach) yields a non-convex optimization problem, which we solve in a distributed way using Lagrange multipliers. A second formulation (indirect approach) is obtained after a suitable re-parameterization. The problem is then convex and is solved by using the distributed subgradient algorithm and the alternating direction method of multipliers (ADMM). The proposed algorithms allow estimating the actual Laplacian eigenvalues with high accuracy. However, they face numerical instability when considering very large graphs.

### 7.1.2. Distributed solution to the network reconstruction problem

**Participants:** A. Kibangou [Contact person], T.-M. D. Tran.

We address the problem of reconstructing the network topology from data propagated through the network by means of a linear average-consensus protocol. In [34], we propose a new method based on the distributed estimation of graph Laplacian spectral properties. Precisely, the identification of the network topology is implemented by estimating both eigenvalues and eigenvectors of the consensus matrix, which is related to the graph Laplacian matrix. Having already solved in [23] the problem of estimating the eigenvalues (see paragraph above), in this paper we focus on the eigenvectors. We show how the topology can be reconstructed in presence of anonymous nodes, i.e., nodes that do not disclose their ID. Actually, in presence of anonymous nodes, eigenvectors are estimated up to a permutation of rows; the obtained graph is then isomorphic to the original one. Moreoved, under some observability assumption on the consensus dynamical system (if the graph is node-observable or neighborhood-observable from the node of interest) and if all the entries of the initial condition of the network state are distinct, then the node can exactly reconstruct the network topology. If the entries of the initial condition of the network state are independently generated from a continuous probability distribution, then the node can reconstruct the network topology almost surely. The main assumption in this work is: all eigenvalues are distinct, that is the case of most random graphs. Future works encompass the design of the network reconstruction protocol that deals with spectrums in which the multiplicities of the eigenvalues can be higher than 1 and also directed graphs. In addition, numerical issues for large graphs are to be considered for making the proposed method scalable.

## 7.2. Sensor networks: estimation and data fusion

### 7.2.1. Multisensor data fusion for attitude estimation

**Participants:** H. Fourati [Contact person], A. Kibangou, A. Makni, T. Michel, P. Geneves [Tyrex, Inria], N. Layaida [Tyrex, Inria].

Multisensor data fusion has gained in importance over the last decades and found applications in an impressive variety of areas within diverse disciplines: navigation, sensor networks, intelligent transportation systems, security, medical diagnosis, biometrics, environmental monitoring, remote sensing, measurements, robotics, and so forth. Different concepts, techniques, and architectures have been developed to optimize the overall system output in applications for which sensor fusion might be useful and enables development of concrete solutions. These concepts and ideas are treated in the book [35], as a response to the great interest and strong activities in the field of multisensor data fusion during the last few years, both in theoretical and practical aspects.

In the team, we have carried out works related to attitude estimation for pedestrian navigation purpose.

In [32], we investigated a new modeling and filtering approach for rigid body attitude estimation. In contrast to the current state-of-the-art. where the process model is driven by gyroscope measurements, we propose an alternative modeling formulation where the process model is fed by the magnetometer measurements. The resulting dynamic model takes the form of a descriptor system, also known as singular system. Based on this model and using the quaternion formalism we derive a recursive filter whose performance is validated through numerical and experimental tests.

In [20], we focused on two main challenges. The first one concerns the attitude estimation during dynamic cases, in which external acceleration occurs. In order to compensate for such external acceleration, we design a quaternion-based adaptive Kalman filter q-AKF. Precisely, a smart detector is designed to decide whether the body is in static or dynamic case. Then, the covariance matrix of the external acceleration is estimated to tune the filter gain. The second challenge is related to the energy consumption issue of gyroscope. In order to ensure a longer battery life for the Inertial Measurement Units, we study the way to reduce the gyro measurements acquisition by switching on/off the sensor while maintaining an acceptable attitude estimation. The switching policy is based on the designed detector. The efficiency of the proposed scheme is evaluated by means of numerical simulations and experimental tests.

In [33], we investigated the precision of attitude estimation solutions in the context of Pedestrian Dead-Reckoning (PDR) with commodity smartphones and inertial/magnetic sensors by carrying out a concise comparison of various methods. We conducted an experimental study with a precise ground truth obtained with a motion capture system. We precisely quantified the error in attitude estimation obtained with each filter which combines a 3-axis accelerometer, a 3-axis magnetometer and a 3-axis gyroscope measurements.

### 7.2.2. *Sensor placement of unreliable sensors*

**Participants:** F. Garin [Contact person], P. Frasca [U. Twente], B. Gerencsér [U. Catholique de Louvain], J. Hendrickx [U. Catholique de Louvain].

We consider problems in which sensors have to be deployed in a given environment in such a way to provide good coverage of it. It is clear that sensor failures may deteriorate the performance of the resulting sensor network. Then, it is also natural to ask if taking into account such uncertainties changes the coverage optimization problem and leads to a different optimal solution. For simplicity, we start considering a one-dimensional problem, where sensors are to be placed on a line in such a way to optimize the disk-coverage cost. The optimal solution for reliable sensors is simply an equally-spaced configuration of the sensors. If we allow that the sensors may fail to take or communicate their measurements, this solution may instead not be optimal. In our work, we assume that sensor can fail, independently and with a same failure probability, and we aim to minimize, in expectation, the largest distance between a point in the environment and an active sensor. Our first result states that the problem at hand is equivalent to a linear program, albeit with a number of variables growing exponentially with the number of sensors. This fact allows for a computational solution that is tractable if the number of sensors is not large. Secondly, we show that for large number of sensors n, the cost of the equispaced placement is asymptotically optimal, i.e., the ratio between its cost and the optimal cost tends to 1 when n grows. By contrast, we show in that a random sensor placement has an expected cost which is larger. This work is described in the paper [18].

# 7.3. Control design and networked control

### 7.3.1. *Control design for hydro-electric power-plants*

**Participants:** C. Canudas de Wit [Contact person], S. Gerwig, F. Garin, B. Sari [Alstom].

We have a collaboration with Alstom on collaborative and resilient control of hydro-electric power-plants, with the CIFRE PhD thesis of Simon Gerwig. The first goal of this research is to improve performance of a hydro-electric power-plant outside its design operation conditions, by cancellation of oscillations that occur in such an operation range. Indeed, current operation of power-plants often requires to operate on a variety of conditions, often different from the ones initially considered when designing the plant. At off-design operation pressure, the hydraulic turbine exhibits a vortex rope below the runner. This vortex generates pressure fluctuations after the turbine and can excite the hydraulic pipes. Indeed the water is compressible and the pipe walls elastic, so the system can oscillate. The goal is to damp these pressure oscillations as they create vibrations in the system and can lead to damages. Our first contribution has been to model the effect of the vortex rope on the hydraulic system as an external perturbation source acting on pipes. The pipes themselves are described with equations taking into account water compressibility and pipe-wall elasticity. The resulting model is nonlinear with hyperbolic functions in the equations (analogous to high-frequency transmission lines), from which we obtain a suitably linearized model.

### 7.3.2. *Collaborative source seeking*

**Participants:** C. Canudas de Wit [Contact person], R. Fabbiano, F. Garin.

The problem of source localization consists in finding, with one or several agents possibly cooperating with each other, the point or the spatial region from which a quantity of interest is being emitted. Source-seeking agents can be fixed sensors, that collect and exchange some information about the signal field and try to identify the position of the source (or the smallest region in which it is included), or moving devices equipped with one or more sensors, that physically reach the source in an individual or cooperative way. This is particularly difficult when the agents have limited or no position information and GPS navigation is not available, as in underwater navigation or in cave exploration: for instance, source localization is relevant to many applications of vapor emitting sources such as explosive detection, drug detection, sensing leakage or hazardous chemicals, pollution sensing and environmental studies. Other fields of interest are sound source localization, heat source localization and vent sources in underwater field. Techniques present in literature either are based on a specific knowledge of the solution of the diffusion process, or make use of an extremum-seeking approach, exciting the system with a periodic signal so as to explore the field and collect enough information to reconstruct the gradient of the quantity of interest. Our approach lies in the computation of derivatives (potentially of any order) from Poisson integrals that, for isotropic diffusive source in steady-state, whose solution satisfies the Laplace equation, allows for a gradient search with a small computation load (derivatives are computed by integrals) and without requiring any knowledge of the closed-form solution, avoiding in the same time extremum-seeking oscillations; this has the additional advantage of an intrinsic high-frequency filtering, that makes the method robust to measurement noise. We also propose a distributed version of this algorithm, where agents communicate in order to reconstruct gradient information from local pointwise measurements, and a control law combines the two objectives of formation control (to have a circular formation, so that measurements are taken around circle) and gradient ascent (so as to move towards the source); differently from previous literature, the moving agents do not need to know their absolute position, but only relative bearing angle of their neighbours. This work is the topic of the Ph.D. thesis of Ruggero Fabbiano [12].

### 7.3.3. *Synchronization of heterogenous networks*

**Participants:** E. Lovisari [Contact person], C.-Y. Kao [National Sun Yat-Sen University, Taiwan].

Syncronization of agents in large-scale networks is studied in [19]. Each agent is modeled as a Single Input Single Output operator composed of the series of a common Linear Time-Invariant system and a possibly nonlinear perturbation. Interconnection is represented via a graph whose edges model communication channels between agents, in turn modeled as a nominal component and a possibly nonlinear perturbation. Two agents are synchronized if their outputs are the same, possibly time-varying signal. The main result provides synchronization certificates based on the Robust Control Technique of Integral Quadratic Constraints. Exploitation of graph structures allows then to reduce the computational burden of the certificate in a way that scales with the dimension of the network. This provides framework which unifies and extendes several results already presented in the literature.

### 7.3.4. *Observer-based FDI scheme for switched systems with sensor faults*

**Participants:**  H. Fourati [Contact person], D. E. C. Belkhiat [U. Setif], D. Jabri [U. Setif].

The Fault Detection and Isolation (FDI) problem for a class of Switched Linear Systems (SLS) subject to sensor faults and unknown bounded Disturbances is proposed in [24]. The main work is based on the design of a generalized switched observer scheme. The FDI problems have been solved by using a robust control techniques. A suitable trade-off between the robustness to disturbances and the sensitivity to sensor faults was obtained. The main results are reformulated by using Linear Matrix Inequality (LMI) formulation. An example is included to illustrate the efficiency of the proposed approach.

## 7.4. Transportation networks and vehicular systems

### 7.4.1. *Traffic estimation: sensors placement and data fusion*

**Participants:**  C. Canudas de Wit [Contact person], E. Lovisari, A. Kibangou.

Ability to reconstruct the state of a transportation network is of paramount importance. Indeed, such an information is used to forecast traffic evolution, to inform drivers in real-time through navigation systems, to provide statistical information to public authorities to detect in a timely fashion accidents and predict hazardous scenarios, and finally to compute controls and to actuate the network through traffic lights, ramp metering, or adaptive speed limits.

A primary source of information on the state of the network are fixed traffic detectors, namely, devices able to measure density, flow and average speed of vehicles crossing the section of the road where they are placed. We have addressed the Optimal Sensor Placement problem [31], namely, the problem of finding the best physical location for sensors. This is based on a trade-off of two contrasting objectives: the first, to maximize the performance of state reconstruction; the second, to minimize the total economic cost of the network. To simplify the setting, we consider the related problem of reconstruction in a static setting, by considering as performance metric the error covariance of an estimator of the cumulative flows in the network over a long period of time. Since the resulting trade-off problem remains a combinatorial problem, we relax it using a method that we call Virtual Variance algorithm, based on the idea to associate to each sensor a virtual variance, which is large when the sensor is not needed for good reconstruction of the flow vector. The only input that the algorithm needs is an estimate of the matrix of splitting ratios and the nominal variance of each sensor. Since in real application a pre-existing sensor network is often unavailable, possible alternatives are field surveys with operators visually counting vehicles, as commonly done for calibration of traffic software, or temporary non-invasive equipment such as radar traffic detectors.

In addition to fixed traffic detectors, the spread of wireless devices allows new sensing and communication capabilities. In particular, for the traffic application, any vehicle equipped with a GPS device can act as a probe in the traffic and provide Floating Car Data (FCD). If a non negligible fraction of vehicles acts as probe, the collected data provides an estimate of the evolution of speed in the network. Due to privacy reasons, single vehicles traces are usually not directly used, but rather aggregated as average speed of vehicles in segments of road. Advanced methodologies, such as the one used by INRIX, ensure a very fine spatial partition of the network, with segments as short as 250 meters (see the INRIX official website http://www.inrix.com/xd-traffic). Compared to fixed sensors, this technology is less precise, but since it exploits existing communication

systems it is relatively less expensive and already covers all major traffic networks. In our work [30], [29], we propose an algorithm that aims at reconstructing the traffic density by fusing fixed sensors measurements and Floating Car Data. We employ a macroscopic model, partitioning the network in cells and assigning to each cell a density of vehicles. The latter evolves dynamically according to a first order mass-conservation law. Our approach inherits from the CTM the cell-based topology, but we do not directly employ the resulting dynamical model. Instead, inflows and outflows are estimated on the basis of the available flow measurements only, and speed measurements are employed to compute a pseudo-measurement of the density. These quantities are the inputs for the density observer. In addition, we propose a gradient descent method to calibrate the Fundamental Diagram, and we implement the proposed solution using real fixed sensor measurements from the Grenoble Traffic Lab [14] and speed FCD measurements provided by INRIX, one of the most well known traffic solutions companies.

### 7.4.2. *Traffic forecasting*

**Participants:** A. Kibangou [Contact person], C. Canudas de Wit, H. Fourati, A. Ladino Lopez.

Traffic forecasting is one of the most desired tools for traffic management, requested by operators and commuters. In the era of data deluge in which we are, measurements collected by sensors are important sources of information that require analysis, classification and processing in order to detect patterns and behaviours that can be exploited for traffic prediction ([30], [37]). The collected information can be classified by clustering algorithms such as K-means; each cluster collects traffic patterns, which in some cases characterize typical regimes such as congestion. Based on clustered data, we have first developed forecasting schemes based on adaptive Kalman filtering [14]. These schemes were designed for specific origin-destination (OD) pairs, assuming availability of meaurements whatever the time instants. Recently, within the PhD thesis in progress of Andres Ladino Lopez, we considered a network-oriented forecasting scheme, where travel time measurements are assumed to be available only for a few sets of OD pairs and sporadically (missing data), but forecasting is to be achieved for all the OD pairs of the network. To reduce the dimensionality of the problem, we actually predicted the travel time for the internal state of the network. In addition, since travel time measurements for all the OD pairs cannot be available all the time, we faced a missing data problem. To overcome this issue, we resorted to a data imputation based on a dictionary learning approach. From the imputed data, a clusterization was achieved, defining different clusters characterized by a centroid containing the mean of the data and a given dispersion around it. The evolution of the centroid can be used as future observation, herein called pseudo-observation, that can feed a Kalman filter. Therefore the prediction problem was solved as a filtering one. However, the main question was, how to associate the current day data to a specific cluster, since we didn't know its future? To solve this issue, we run Kalman filters for each cluster and then made the fusion of the obtained forecasts.

### 7.4.3. *Traffic control*

**Participants:** C. Canudas de Wit [Contact person], F. Garin, D. Pisarski, P. Grandinetti, E. Lovisari, G. Como [U. Lund], K. Savla [U. of Southern California].

The activities of the team on traffic control can be organized in three parts: freeway traffic control, urban control, and analysis and control of monotone flows.

First, we have studied optimal balancing of vehicle density in the freeway traffic. The optimization is performed in a distributed manner by utilizing the controllability properties of the freeway network represented by the Cell Transmission Model. By using these properties, we identify the subsystems to be controlled by local ramp meters. The optimization problem is then formulated as a non-cooperative Nash game that is solved by decomposing it into a set of two-players hierarchical and competitive games. The process of optimization employs the communication channels matching the switching structure of system interconnectivity. By defining the internal model for the boundary flows, local optimal control problems are efficiently solved by utilizing the method of Linear Quadratic Regulator. The developed control strategy is tested via numerical simulations in two scenarios for uniformly congested and transient traffic. This work is described in the paper [21].

Second, we have considered optimal or near-optimal operation of traffic lights in an urban area. The goal is on-line optimization of traffic light 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, while also ensuring good servicing of demands of incoming cars in the network from other areas. One way to address the complexity of the resulting optimization problem is to use a simplified averaged model for the traffic variables, and to optimize only the duty-cyles of traffic lights, i.e., the fractions of green time. This, together with a one-step optimization horizon, allows us to turn the problem into a simple linear program [27]. Another approach is to include as optimization variables both duty-cycles and phases of the traffic lights. We show how to turn the resulting problem into a mixed-integer linear program (MILP). Then, to overcome its complexity, we propose a sub-optimal distributed solution, while the global MILP can be used off-line for performance comparison [28].

Third, stability and throughput properties of monotone dynamical flow networks are studied in [15]. Vehicular density on the cells of the networks evolves according to laws that deterministically split the traffic flow at each intersection as a function of the density of other cells around the intersection. By exploiting the theory of monotone operators it is proven that under certain condition the system achieves an equilibrium that maximizes the throughput of the network, namely, if the inflow is smaller than the network capacity, then asymptotically the total outflow matches the total inflow, otherwise the total outflow matches the network capacity. In [25] a different traffic model is employed which uses demand and supply functions to relate density and flows of the network. The Social Optimum Dynamic Traffic Assignment, which is an optimal control problem with cost corresponding to the total travel time of vehicles in the network, is solved making use of ramp metering and speed limits. The optimization is shown to be a convex optimization problem, making its numerical solution feasible by employing well known tools.

### 7.4.4. *Energy-aware control of communicating vehicles*
**Participants:**  C. Canudas de Wit [Contact person], G. de Nunzio.

The research in this domain focuses mainly on efficient traffic energy consumption and has been carried out at two levels. Strategies for both the vehicles-side and the infrastructure-side eco-management have been proposed or extended. As for the vehicle-side control of communicating vehicles, assuming I2V communication, and therefore full knowledge of the traffic lights timings, the goal is to analyze the driving horizon and compute an energy-efficient speed advisory for the driver. As in previous works, stops at a red traffic light are to be avoided. The novelty of our approach is summarized as follows. Given a set of green traffic light phases, there exist different driving profiles to reach a given destination at a given final time in compliance with traffic lights constraints (i.e. always catching the green light) and city speed limits. The presented strategy is capable of an a priori identification of the most energy-efficient velocity trajectory, by approximating the available paths and their energy cost with an oriented weighted graph. The computational complexity of the graph creation has been reduced in this work from exponential [26] to polynomial, thanks to the introduction of the line graph. The computation time has been consequently significantly reduced. Only after this preliminary stage of path selection, a formal optimization problem is solved in order to calculate the optimal arrival times at each intersection, by explicitly minimizing the energy consumption of the vehicle. This approach qualifies as a pre-trip eco-driving ADAS, since the speed advisory is provided to the driver at the beginning of the driver horizon. However, given the very little computation time required by the algorithm, it may be employed online thus enabling in-trip assistance features. This allows to respond dynamically to traffic perturbations and/or deviations from the speed advisory, and to increase the robustness and the applicability of the strategy in a realistic environment. Simulations in a microscopic traffic simulator demonstrate that the proposed strategy is able to deal online with perturbations coming from traffic and to reduce the overall energy consumption without affecting travel time [16].

At a lower level, the eco-driving from the vehicle perspective has been also addressed in a comprehensive analysis of the optimal driving strategy for different types of powertrains [22].

As for the infrastructure-side eco-management, this year's research focused on extending the results published in [26]. The two-way arterial bandwidth maximization problem is addressed with a particular focus on the

benefits induced by the speed advisory, and on reducing energy consumption. The problem with internal offsets constraints presents difficulties that make necessary the formulation of the problem as an MILP. The first contribution of our work lies in the addition of terms representing traffic energy consumption and network travel time to the objective function of the two-way arterial bandwidth maximization. The segment speeds, as additional control action, allow to reach higher theoretical bandwidths but might induce driving discomfort and higher energy consumption if the variability of the recommended speeds is too high. Furthermore, optimal solutions with low speeds and high travel time are to be avoided, in trade-off with the energy consumption. The second contribution is given by the extensive evaluation of the benefits of bandwidth maximization via a microscopic traffic simulator. Bandwidth is a theoretical quantity and a correlation with known traffic performance metrics needs to be established in order to justify its use. The combined control of offsets and speed advisory is shown to have a large impact on energy consumption without affecting the travel time. Lastly, an analysis of the traffic performance at different levels of traffic demands has been conducted, testing both under-saturated traffic conditions with the existence of a green wave, and saturated conditions. The goal of this analysis is to identify the best operation conditions of the presented approach, assess the performance degradation with traffic load, and, most importantly, propose a demand-dependent optimization. Several strategies were compared to the presented one in order to assess its performance. This work has been submitted for review to the IEEE Transactions on Control Systems Technology.

Finally, a detailed description of the proposed strategies and the achieved results in the domain of the energy-aware traffic management in urban networks can be found in the dissertation [11].

<p style="text-align:center"><span style="color:red">**AIRSEA Team**</span></p>

# 7. New Results

## 7.1. Modeling for Oceanic and Atmospheric flows

### 7.1.1. Coupling Methods for Oceanic and Atmospheric Models

**Participants:** Eric Blayo, Mehdi-Pierre Daou, Laurent Debreu, Florian Lemarié, Charles Pelletier, Antoine Rousseau.

*7.1.1.1. Coupling heterogeneous models in hydrodynamics*

The coupling of models of different kinds is gaining more and more attention, due in particular to a need for more global modeling systems encompassing different disciplines (e.g. multi-physics) and different approaches (e.g. multi-scale, nesting). In order to develop such complex systems, it is generally more pragmatic to assemble different modeling units inside a user friendly modelling software platform rather than to develop new complex global models.

In the context of hydrodynamics, global modeling systems have to couple models of different dimensions (1D, 2D or 3D) and representing different physics (Navier-Stokes, hydrostatic Navier-Stokes, shallow water...). We have been developing coupling approaches for several years, based on so-called Schwarz algorithms. Our recent contributions address the development of absorbing boundary conditions for Navier-Stokes equations [1], and of interface conditions for coupling hydrostatic and nonhydrostatic Navier-Stokes flows [2]. In the context of our partnership with with ARTELIA Group (PhD thesis of Medhi Pierre Daou), implementations of Schwarz coupling algorithms have been performed for hydrodynamics industrial codes (Mascaret, Telemac and OpenFoam), using the PALM coupling software. A first implementation has been realized in an academic test case, and a second one is presently under implementation in a much more realistic context.

*7.1.1.2. Ocean-atmosphere coupling*

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 [12]. For the last few years we have been actively working on the analysis of Schwarz waveform relaxation to apply this type of iterative coupling method to air-sea coupling [59], [60], [58]. In the context of the simulation of tropical cyclone, sensitivity tests to the coupling method have been carried out using ensemble simulations (through perturbations of the coupling frequency and initial conditions). We showed that the use of the Schwarz iterative coupling methods leads to a significantly reduced spread in the ensemble results (in terms of cyclone trajectory and intensity), thus suggesting that a source of error is removed w.r.t coupling methods en vogue in existing coupled models [61].

Motivated by this encouraging result, our activities over the last year can be divided into three topics

1. *Stability and consistency analysis of existing coupling methods*: in [12] 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.

2. *Study of physics-dynamics coupling*: during the PhD-thesis of Charles Pelletier (funded by Inria) the scope is on including the formulation of physical parameterizations in the theoretical analysis of the coupling. The first months of this Ph-D were dedicated to the study of the parameterization schemes to compute air-sea fluxes. A thorough sensitivity analysis showed that several parameters within existing schemes have no influence on the resulting fluxes. A simplified scheme retaining most the complexity of complicated parameterizations has thus been designed. This new scheme has also the advantage to be more adequate to conduct the mathematical analysis of the coupling.

3. *Design of a coupled single column model*: in order to focus on specific problems of ocean-atmosphere coupling, a work on simplified equation sets has been started. The aim is to implement a one-dimensional (in the vertical direction) coupled model with physical parameterizations representative of those used in realistic models. Thanks to this simplified coupled model the objective is to develop a benchmark suite for coupled models evaluation.

These three topics are addressed through strong collaborations between the applied mathematics and the climate community As an illustration, the PhD-thesis of Charles Pelletier is in collaboration with the LSCE (Laboratoire des Sciences du Climat et de l'Environnement).

Moreover a PPR (*Projet à partenariat renforcé*) called SIMBAD (SIMplified Boundary Atmospheric layer moDel for ocean modeling purposes) is funded by Mercator-Ocean for the next three years (from march 2015 to march 2018). The aim of this project in collaboration with Meteo-France, Ifremer, LMD, and LOCEAN is to derive a metamodel to force high-resolution oceanic operational models for which the use of a full atmospheric model is not possible due to a prohibitive computational cost.

### 7.1.1.3. 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, R. Pellerej started a PhD on that topic late 2014. So far, three general data assimilation algorithms, based on variational data assimilation techniques, have been developed and applied to a simple coupled problem. 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. Preliminary results shows significant improvement compared to the usual approach on this simple system.

## 7.1.2. Numerical Schemes for Ocean Modelling

**Participants:** Eric Blayo, Laurent Debreu, Florian Lemarié.

In 2015, we worked on the stability constraints for oceanic numerical models ([13]). The idea is to carry a deep analysis of these constraints in order to propose new time stepping algorithms for ocean models. Except for vertical diffusion (and possibly the external mode and bottom drag), oceanic models usually rely on explicit time-stepping algorithms subject to Courant-Friedrichs-Lewy (CFL) stability criteria. Implicit methods could be unconditionally stable, but an algebraic system must be solved at each time step and other considerations such as accuracy and efficiency are less straightforward to achieve. Depending on the target application, the process limiting the maximum allowed time-step is generally different. In this paper, we introduce offline diagnostics to predict stability limits associated with internal gravity waves, advection, diffusion, and rotation. This suite of diagnostics is applied to a set of global, regional and coastal numerical simulations with several horizontal/vertical resolutions and different numerical models. We show that, for resolutions finer that $1/2°$, models with an Eulerian vertical coordinate are generally constrained by vertical advection in a few hot spots and that numerics must be extremely robust to changes in Courant number. Based on those results, we review the stability and accuracy of existing numerical kernels in vogue in primitive equations oceanic models with a focus on advective processes and the dynamics of internal waves. We emphasize the additional value of studying the numerical kernel of oceanic models in the light of coupled space-time approaches instead of studying the time schemes independently from spatial discretizations. From this study, we suggest some guidelines for the development of temporal schemes in future generation multi-purpose oceanic models.

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). In [16], we present a Baroclinic Jet test case set in a zonally reentrant channel that provides a controllable test of a model capacity at resolving submesoscale dynamics. We compare simulations from two models, ROMS and NEMO, at different mesh sizes (from 20 to 2 km). Through a spectral decomposition of kinetic energy and its budget terms, we identify the characteristics of numerical dissipation and effective resolution. It shows that numerical dissipation appears in different parts of a model, especially in spatial advection-diffusion schemes for momentum equations (KE dissipation) and tracer equations (APE dissipation) and in the time stepping algorithms. Effective resolution, defined by scale-selective dissipation, is inadequate to qualify traditional ocean models with low-order spatial and temporal filters, even at high grid resolution. High- order methods are better suited to the concept and probably unavoidable. Fourth-order filters are suited only for grid resolutions less than a few kilometers and momentum advection schemes of even higher-order may be justified. The upgrade of time stepping algorithms (from filtered Leapfrog), a cumbersome task in a model, appears critical from our results, not just as a matter of model solution quality but also of computational efficiency (extended stability range of predictor-corrector schemes). Effective resolution is also shaken by the need for non scale-selective barotropic mode filters and requires carefully addressing the issue of mode splitting errors. Possibly the most surprising result is that submesoscale energy production is largely affected by spurious diapycnal mixing (APE dissipation). This result justifies renewed efforts in reducing tracer mixing errors and poses again the question of how much vertical diffusion is at work in the real ocean.

### 7.1.3. *Better Parameterization of the Coastline for Ocean Models*

**Participants:** Eric Blayo, Eugene Kazantsev, Florian Lemarié, Pierre Marchand.

We aim at the development of finer approximations of lateral boundaries and boundary conditions for NEMO, by investigating and comparing analytical and optimal control approaches.

Regarding the analytical approach, we focused on a 2D shallow water formulation, and revisited the properties of the energy and enstrophy conserving schemes in the presence of a coastline. This led us to highlight a number of problems with the enstrophy conserving scheme (sensitivity to the choice of a slip or a noslip boundary condition, non conservation of the enstrophy, numerical instability). We also proposed a corrected scheme near the boundary for the continuity equation and new values for ghost points derived from the energy conservation in order for the energy conserving scheme to take into account a coastline with some inclination with regard to the numerical grid. We also investigated the viscous case, and proposed an implementation of slip and no slip boundary conditions for the viscous term in such a case of an inclined coastline.

These results are under comparison with the optimal control approach 7.3.2 realised for the Nemo model in a similar configuration.

## 7.2. Model reduction / multiscale algorithms

### 7.2.1. *Intrusive sensitivity analysis, reduced models*

**Participants:** Maëlle Nodet, Clémentine Prieur.

Another point developed in the team for sensitivity analysis is model reduction. To be more precise regarding model reduction, the aim is to reduce the number of unknown variables (to be computed by the model), using a well chosen basis. Instead of discretizing the model over a huge grid (with millions of points), the state vector of the model is projected on the subspace spanned by this basis (of a far lesser dimension). The choice of the basis is of course crucial and implies the success or failure of the reduced model. Various model reduction methods offer various choices of basis functions. A well-known method is called "proper orthogonal decomposition" or "principal component analysis". More recent and sophisticated methods also exist and may be studied, depending on the needs raised by the theoretical study. Model reduction is a natural

way to overcome difficulties due to huge computational times due to discretizations on fine grids. In [55], the authors present a reduced basis offline/online procedure for viscous Burgers initial boundary value problem, enabling efficient approximate computation of the solutions of this equation for parametrized viscosity and initial and boundary value data. This procedure comes with a fast-evaluated rigorous error bound certifying the approximation procedure. The numerical experiments in the paper show significant computational savings, as well as efficiency of the error bound.

When a metamodel is used (for example reduced basis metamodel, but also kriging, regression, ...) for estimating sensitivity indices by Monte Carlo type estimation, a twofold error appears: a sampling error and a metamodel error. Deriving confidence intervals taking into account these two sources of uncertainties is of great interest. We obtained results particularly well fitted for reduced basis metamodels [56]. In [54], the authors provide asymptotic confidence intervals in the double limit where the sample size goes to infinity and the metamodel converges to the true model. These results were also adapted to problems related to more general models such as Shallow-Water equations, in the context of the control of an open channel [8].

Let us come back to the output of interest. Is it possible to get better error certification when the output is specified. A work in this sense has been accepted, dealing with goal oriented uncertainties assessment [7].

A collaboration has been started with Christophe Prieur (Gipsa-Lab) on the very challenging issue of sensitivity of a controlled system to its control parameters [8].

### 7.2.2. *Multigrid Methods for Variational Data Assimilation.*

**Participants:** Laurent Debreu, François-Xavier Le Dimet, Arthur Vidard.

In order to lower the computational cost of the variational data assimilation process, we investigate the use of multigrid methods to solve the associated optimal control system. On a linear advection equation, we study the impact of the regularization term on the optimal control and the impact of discretization errors on the efficiency of the coarse grid correction step. We show that even if the optimal control problem leads to the solution of an elliptic system, numerical errors introduced by the discretization can alter the success of the multigrid methods. The view of the multigrid iteration as a preconditioner for a Krylov optimization method leads to a more robust algorithm. A scale dependent weighting of the multigrid preconditioner and the usual background error covariance matrix based preconditioner is proposed and brings significant improvements. This work is summarized in ([5]).

## 7.3. Dealing with uncertainties

### 7.3.1. *Sensitivity Analysis for Forecasting Ocean Models*

**Participants:** Eric Blayo, Laurent Gilquin, Céline Helbert, François-Xavier Le Dimet, Elise Arnaud, Simon Nanty, Maëlle Nodet, Clémentine Prieur, Laurence Viry, Federico Zertuche.

#### 7.3.1.1. *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.

*7.3.1.2. Estimating sensitivity indices*

A first task is to develop tools for estimated sensitivity indices. In variance-based sensitivity analysis, a classical tool is the method of Sobol' [68] which allows to compute Sobol' indices using Monte Carlo integration. One of the main drawbacks of this approach is that the estimation of Sobol' indices requires the use of several samples. For example, in a $d$-dimensional space, the estimation of all the first-order Sobol' indices requires $d + 1$ samples. Some interesting combinatorial results have been introduced to weaken this defect, in particular by Saltelli [66] and more recently by Owen [64] but the quantities they estimate still require $O(d)$ samples.

In a recent work [71] we introduce a new approach to estimate all first-order Sobol' indices by using only two samples based on replicated latin hypercubes and all second-order Sobol' indices by using only two samples based on replicated randomized orthogonal arrays. We establish theoretical properties of such a method for the first-order Sobol' indices and discuss the generalization to higher-order indices. As an illustration, we propose to apply this new approach to a marine ecosystem model of the Ligurian sea (northwestern Mediterranean) in order to study the relative importance of its several parameters. The calibration process of this kind of chemical simulators is well-known to be quite intricate, and a rigorous and robust — i.e. valid without strong regularity assumptions — sensitivity analysis, as the method of Sobol' provides, could be of great help. The computations are performed by using CIGRI, the middleware used on the grid of the Grenoble University High Performance Computing (HPC) center. We are also applying these estimates to calibrate integrated land use transport models. As for these models, some groups of inputs are correlated, Laurent Gilquin extended the approach based on replicated designs for the estimation of grouped Sobol' indices [6].

We can now wonder what are the asymptotic properties of these new estimators, or also of more classical ones. In [54], the authors deal with asymptotic properties of the estimators. In [52], the authors establish also a multivariate central limit theorem and non asymptotic properties.

*7.3.1.3. Sensitivity analysis with dependent inputs*

An important challenge for stochastic sensitivity analysis is to develop methodologies which work for dependent inputs. For the moment, there does not exist conclusive results in that direction. Our aim is to define an analogue of Hoeffding decomposition [53] in the case where input parameters are correlated. Clémentine Prieur supervised Gaëlle Chastaing's PhD thesis on the topic (defended in September 2013) [44]. We obtained first results [45], deriving a general functional ANOVA for dependent inputs, allowing defining new variance based sensitivity indices for correlated inputs. We then adapted various algorithms for the estimation of these new indices. These algorithms make the assumption that among the potential interactions, only few are significant. Two papers have been recently accepted [43], [46]. We also considered (see the paragraph 7.3.1 ) the estimation of groups Sobol' indices, with a procedure based on replicated designs. These indices provide information at the level of groups, and not at a finer level, but their interpretation is still rigorous.

Céline Helbert and Clémentine Prieur supervised the PhD thesis of Simon Nanty (funded by CEA Cadarache, and defended in October, 2015). The subject of the thesis is the analysis of uncertainties for numerical codes with temporal and spatio-temporal input variables, with application to safety and impact calculation studies. This study implied functional dependent inputs. A first step was the modeling of these inputs, and a paper has been submitted [63]. The whole methodology proposed during the PhD is under advanced revision [36].

*7.3.1.4. Multy-fidelity modeling for risk analysis*

Federico Zertuche's PhD concerns the modeling and prediction of a digital output from a computer code when multiple levels of fidelity of the code are available. A low-fidelity output can be obtained, for example on a coarse mesh. It is cheaper, but also much less accurate than a high-fidelity output obtained on a fine mesh. In this context, we propose new approaches to relieve some restrictive assumptions of existing methods ( [57], [65]): a new estimation method of the classical cokriging model when designs are not nested and a nonparametric modeling of the relationship between low-fidelity and high-fidelity levels. The PhD takes place in the REDICE consortium and in close link with industry. The first part of the thesis was also dedicated to the development of a case study in fluid mechanics with CEA in the context of the study of a nuclear reactor.

The second part of the thesis was dedicated to the development of a new sequential approach based on a course to fine wavelets algorithm. Federico Zertuche presented his work at the annual meeting of the GDR Mascot Num in 2014 [72].

*7.3.1.5. Data assimilation and second order sensitivity analysis*

Basically, in the deterministic approach, a sensitivity analysis is the evaluation of a functional depending on the state of the system and of parameters. Therefore it is natural to introduce an adjoint model. In the framework of variational data assimilation the link between all the ingredients (observations, parameters and other inputs of the model is done through the optimality system (O.S.), therefore a sensitivity will be estimated by deriving the O.S. leading to a second order adjoint. This is done in the paper [15] in which a full second order analysis is carried out on a model of the Black Sea.

This methodology has been applied to

- **Oil Spill**. These last years have known several disasters produced by wrecking of ships and drifting platforms with severe consequences on the physical and biological environments. In order to minimize the impact of these oil spills its necessary to predict the evolution of oil spot. Some basic models are available and some satellites provide images on the evolution of oil spots. Clearly this topic is a combination of the two previous one: data assimilation for pollution and assimilation of images. A theoretical framework has been developed with Dr. Tran Thu Ha (iMech).

- **Data Assimilation in Supercavitation (with iMech)**. Some self propelled submarine devices can reach a high speed thanks to phenomenon of supercavitation: an air bubble is created on the nose of the device and reduces drag forces. Some models of supercavitation already exist but are working on two applications of variational methods to supercavitation:

  - Parameter identification : the models have some parameters that can not be directly measured. From observations we retrieve the unknown parameters using a classical formalism of inverse problems.
  - Shape Optimization. The question is to determine an optimum design of the shape of the engine in order to reach a maximum speed.

## 7.3.2. *Optimal Control of Boundary Conditions*

**Participants:** Christine Kazantsev, Eugene Kazantsev.

A variational data assimilation technique is applied to the identification of the optimal boundary conditions for a simplified configuration of the NEMO model. A rectangular box model placed in mid-latitudes, and subject to the classical single or double gyre wind forcing, is studied. The model grid can be rotated on a desired angle around the center of the rectangle in order to simulate the boundary approximated by a staircase-like coastlines. The solution of the model on the grid aligned with the box borders was used as a reference solution and as artificial observational data. It is shown in [9], [10] that optimal boundary has a rather complicated geometry which is neither a staircase, nor a straight line. The boundary conditions found in the data assimilation procedure bring the solution toward the reference solution allowing to correct the influence of the rotated grid (see fig. 1 ).

Adjoint models, necessary to variational data assimilation, have been produced by the TAPENADE software, developed by the SCIPORT team. This software is shown to be able to produce the adjoint code that can be used in data assimilation after a memory usage optimization.

## 7.3.3. *Non-Parametric Estimation for Kinetic Diffusions*

**Participants:** Clémentine Prieur, Jose Raphael Leon Ramos.

This research is the subject of a collaboration with Venezuela and is partly funded by an ECOS Nord project.
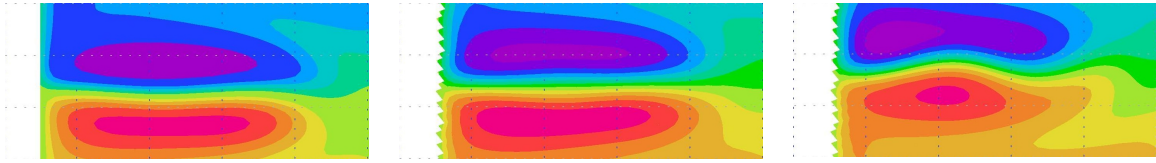
*Figure 1. Sea surface elevation: reference solution on the aligned grid (left), solutions on the $30°$ rotated grid with optimal (center) and conventioal (right) boundary conditions.*

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 [40] and when only partial observational data are available. A paper on the non parametric estimation of the drift has been accepted recently [41] (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 [41], in collaboration with J.R. Leon (Caracas, Venezuela) and P. Cattiaux (Toulouse). Recursive estimators have been also proposed by the same authors in [42], also recently accepted.20

Note that Professor Jose R. Leon (Caracas, Venezuela) is now funded by an international Inria Chair and will spend one year in our team, allowing to collaborate further on parameter estimation.

### 7.3.4. *Multivariate Risk Indicators*

**Participants:**  Clémentine Prieur, Patricia Tencaliec.

Studying risks in a spatio-temporal context is a very broad field of research and one that lies at the heart of current concerns at a number of levels (hydrological risk, nuclear risk, financial risk etc.). Stochastic tools for risk analysis must be able to provide a means of determining both the intensity and probability of occurrence of damaging events such as e.g. extreme floods, earthquakes or avalanches. It is important to be able to develop effective methodologies to prevent natural hazards, including e.g. the construction of barrages.

Different risk measures have been proposed in the one-dimensional framework . The most classical ones are the return level (equivalent to the Value at Risk in finance), or the mean excess function (equivalent to the Conditional Tail Expectation CTE). However, most of the time there are multiple risk factors, whose dependence structure has to be taken into account when designing suitable risk estimators. Relatively recent regulation (such as Basel II for banks or Solvency II for insurance) has been a strong driver for the development of realistic spatio-temporal dependence models, as well as for the development of multivariate risk measurements that effectively account for these dependencies.

We refer to [47] for a review of recent extensions of the notion of return level to the multivariate framework. In the context of environmental risk, [67] proposed a generalization of the concept of return period in dimension greater than or equal to two. Michele et al. proposed in a recent study [48] to take into account the duration and not only the intensity of an event for designing what they call the dynamic return period. However, few studies address the issues of statistical inference in the multivariate context. In [49], [51], we proposed non parametric estimators of a multivariate extension of the CTE. As might be expected, the properties of these estimators deteriorate when considering extreme risk levels. In collaboration with Elena Di Bernardino (CNAM, Paris), Clémentine Prieur is working on the extrapolation of the above results to extreme risk levels.

Elena Di Bernardino, Véronique Maume-Deschamps (Univ. Lyon 1) and Clémentine Prieur also derived an estimator for bivariate tail [50]. The study of tail behavior is of great importance to assess risk.

With Anne-Catherine Favre (LTHE, Grenoble), Clémentine Prieur supervises the PhD thesis of Patricia Tencaliec. We are working on risk assessment, concerning flood data for the Durance drainage basin (France). The PhD thesis started in October 2013. A first paper on data reconstruction has been accepted [18]. It was a necessary step as the initial series contained many missing data.

## 7.4. Assimilation of Images

**Participants:** François-Xavier Le Dimet, Maëlle Nodet, Arthur Vidard, Nelson Feyeux, Vincent Chabot, Nicolas Papadakis.

### 7.4.1. 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.

MOISE was the leader of the ANR ADDISA project dedicated to the assimilation of images, and is a member of its follow-up GeoFluids (along with EPI FLUMINANCE and CLIME, and LMD, IFREMER and Météo-France) that ended in 2013.

During the ADDISA project we developed Direct Image Sequences Assimilation (DISA) and proposed a new scheme for the regularization of optical flow problems [69], which was recently extended [17]. Thanks to the nonlinear brightness assumption, we proposed an algorithm to estimate the motion between two images, based on the minimization of a nonlinear cost function. We proved its efficiency and robustness on simulated and experimental geophysical flows [38]. As part of the ANR project GeoFluids, we are investigating new ways to define distance between a couple of images. One idea is to compare the gradient of the images rather than the actual value of the pixels. This leads to promising results. Another idea, currently under investigation, consists in comparing main structures within each image. This can be done using, for example, a wavelet representation of images. Both approaches have been compared, in particular their relative merits in dealing with observation errors, in a paper published early 2015 [4] and presented in several international conferences [21], [28].

In recent developments [11] we have also used "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.

### 7.4.2. Optimal transport for image assimilation

Within the optimal transport project TOMMI funded by the ANR white program (started mid 2011), a new optimization scheme based on proximal splitting method has been proposed to solve the dynamic optimal transport problem. We investigate the use of optimal transport based distances for data assimilation. The study is still under investigation in the framework of N. Feyeux's PhD, but preliminary encouraging results have already been presented in [20] and an article is in preparation on this topic.

## 7.5. Tracking of Mesoscale Convective Systems

**Participant:** Clémentine Prieur.

We are interested in the tracking of mesoscale convective systems.A particular region of interest is West Africa. Data and hydrological expertise is provided by T. Vischel and T. Lebel (LTHE, Grenoble).

A first approach involves adapting the multiple hypothesis tracking (MHT) model originally designed by the NCAR (National Centre for Atmospheric Research) for tracking storms [70] to the data for West Africa. With A. Makris (working on a post-doctoral position), we proposed a Bayesian approach [62], which consists in considering that the state at time t is composed on one hand by the events (birth, death, splitting, merging) and on the other hand by the targets' attributes (positions, velocities, sizes, ... ). The model decomposes the state into two sub-states: the events and the targets positions/attributes. The events are updated first and are conditioned to the previous targets sub-state. Then given the new events the target substate is updated. A simulation study allowed to verify that this approach improves the frequentist approach by Storlie et al. (2009). It has been tested on simulations [62] and investigated in the specific context of real data on West Africa [35]. Using PHD (probability hypothesis density) filters adapted to our problem, generalizing recent developments in particle filtering for spatio-temporal branching processes (e.g. [39]) could be an interesting alternative to explore. The idea of a dynamic, stochastic tracking model should then provide the base for generating rainfall scenarios over a relatively vast area of West Africa in order to identify the main sources of variability in the monsoon phenomenon.

## 7.6. Land Use and Transport Models Calibration

**Participants:** Thomas Capelle, Laurent Gilquin, 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. This work led to conference papers [24], [23] and a two published journal paper [3], [6]

<span style="color:red">**BEAGLE Project-Team**</span>

# 7. New Results

## 7.1. Subspace Clustering Using Evolvable Genome Structure

We have developed an evolutionary algorithm to tackle the subspace clustering problem. Subspace clustering is recognized as more difficult than standard clustering since it requires to identify not only the clusters but also the various subspaces where the clusters hold. We propose to tackle this problem with a bio-inspired algorithm that includes many bio-like features like variable genome length and organization, functional and non-functional elements, and variation operators including chromosomal rearrangements. These features give the algorithm a large degree of freedom to achieve subspace clustering with satisfying results on a reference benchmark with respect to state of the art methods. One of the main advantages of the approach is that it needs only one subspace clustering ad-hoc parameter: the maximal number of clusters. This is a single and intuitive parameter that sets the maximal level of details of the clustering, while other algorithms require more complicated parameter space exploration. The other parameters of the algorithm are related to the evolution strategy (population size, mutation rate, ...) and for them we use a single setting that turns out to be effective on all the datasets of the benchmark.

This work has been presented at the main conference for genetic & evolutionary computation, GECCO [31], where it received the best paper award and during the EvoEvo Workshop of ECAL 2015 [35].

## 7.2. Epigenetic inheritance speeds up evolution of artificial organisms

DNA is not the sole medium by which parents transmit information to their offspring. Epigenetic inheritance, in particular, is based on the partial transmission of the cellular state of the parental cell to its descendants. Although the reality of epigenetic inheritance is now firmly established, whether it has an influence on the long term evolutionary process is still subject to debate. To address this question, we used the RAevol extention of the Aevol simulator developed in the team, and defined 4 scenarios with static or dynamic environments and with or without epigenetic inheritance. Simulations in dynamic environments show that protein inheritance indeed increases the rate of evolution on the long term. But they also show that it impedes evolution in its very first stages. This negative effect can be explained by instabilities generated by the interference between the two inheritance mediums. On the opposite, the long term gain can be explained by protein inheritance reducing the constraints on the genetic regulation network.

This work has been published in the article [33].

## 7.3. In silico evolution improves statistical models of genome dynamics

Using Aevol, we have proved that statistical frameworks published in the last twenty years for inferring evolutionary genome rearrangements are flawed in two ways. First, they mistranslated a null hypothesis on a uniform breakage model, and second, they assumed that genomic breakable regions are known *a priori*. We propose ways to correct these flaws by combining mathematical approaches, simulations, observations and validation on real genomic data. The results will be of interest for an audience from evolutionary biology, computational biology, bioinformatics and mathematics. We successively show that:

- a truly uniform hypothesis on rearrangement breakages leads to a model with an equilibrium intergene size distribution that fits the measured one on diverse genomes,
- estimations based on the flawed uniform breakage model completely fail on simulations with the truly uniform model,
- coherently with previous studies the flawed, and to a lesser extent, the truly uniform model are rejected on amniote genomes if breakable regions are identified with intergenic regions,
- co-estimating the number of breakable regions with the rearrangement distance gives coherent values on amniote genomes.

A paper reporting these results has been submitted by Priscila Biller, Carole Knibbe and Eric Tannier.

## 7.4. Temperature-induced variation in gene expression burst size in metazoan cells

Gene expression is an inherently stochastic process, owing to its dynamic molecular nature. Protein amount distributions, which can be acquired by cytometry using a reporter gene, can inform about the mechanisms of the underlying microscopic molecular system. By using different clones of chicken erythroid progenitor cells harboring different integration sites of a CMV-driven mCherry protein, we investigated the dynamical behavior of such distributions. We show that, on short term, clone distributions can be quickly regenerated from small population samples with a high accuracy. On longer term, on the contrary, we show variations manifested by correlated fluctuation in the Mean Fluorescence Intensity. In search for a possible cause of this correlation, we demonstrate that in response to small temperature variations cells are able to adjust their gene expression rate: a modest (2 °C) increase in external temperature induces a significant down regulation of mean expression values, with a reverse effect observed when the temperature is decreased. Using a two-state model of gene expression we further demonstrate that temperature acts by modifying the size of transcription bursts, while the burst frequency of the investigated promoter is less systematically affected. For the first time, we report that transcription burst size is a key parameter for gene expression that metazoan cells from homeotherm animals can modify in response to an external thermal stimulus.
This work has been published in the article [11].

## 7.5. Deciphering the signalling networks of synaptic plasticity

Synaptic plasticity, i.e. adaptive modifications of synaptic strength between two neurons depending on their activity, is a main substrate for learning and memory. Experimentally, synaptic plasticity is commonly assessed using prolonged electrical stimulations. Since learning can arise from few or even a single trial, synaptic strength is expected to adapt rapidly. However, whether synaptic plasticity occurs in response to limited event occurrences remains elusive. To address this question, we started a collaboration with Laurent Venance Lab (experimental neuroscience, College de France, Paris). Combining experimental and modelling approaches, we investigated whether a low number of stimulations can induce plasticity in a major synaptic learning rule, spike-timing-dependent plasticity (STDP). It is known that 100 stimulations induce bidirectional STDP, i.e. spike-timing-dependent potentiation (tLTP) and depression (tLTD) at most central synapses. In rodent striatum, we found that tLTD progressively disappears when the number of stimulations is decreased (below 50 pairings) whereas tLTP displays a biphasic profile: tLTP is observed for 75-100 stimulations, absent for 25-50 stimulations and re-emerges for 5-10 stimulations. This tLTP, induced by very few stimulations (5-10) depends on the endocannabinoid (eCB) system. The eCB system has recently emerged as a pivotal pathway for synaptic plasticity because of its widely characterized ability to depress synaptic transmission on short- and long-term scales. Our result therefore indicate that eCBs also mediate potentiation of the synapse. To understand how eCB signaling may support such bidirectionality, we combined electrophysiology experiments with mathematical modeling. Our model describes the temporal kinetics of the biochemical species involved in a first signaling pathway leading from NMDAR to calmodulin and CaMKII with that of a a second, distinct one that assembles mGluR and cytosolic calcium to eCB production and the resulting activation of CB1R. This demonstrated that STDP outcome is controlled by eCB levels and dynamics: prolonged and moderate levels of eCB lead to eCB-mediated long-term depression (eCB- tLTD) while short and large eCB transients produce eCB-mediated long-term potentiation (eCB-tLTP). Therefore, just like neurotransmitters glutamate or GABA, eCB forms a bidirectional system to encode learning and memory.
For reasons of publication strategy, our first co-publication on the subject presents our major experimental results [16]. A second article, featuring both experimental and modelling results, explains how the underlying signalling network can support the observed bidirectionality and is under submission.

## 7.6. Anomalous diffusion as an age-structured renewal process

Continuous-time random walks (CTRW) are one of the main mechanisms that are recurrently evoked to explain the emergence of subdiffusion in cells. CTRW were introduced fifty years ago as a generalisation of random walks, where the residence time (the time between two consecutive jumps) is a random variable. If the expectation of the residence time is defined, for instance when it is dirac-distributed or decays exponentially fast, one recovers "normal" Brownian motion. However, when the residence time expectation diverges, the CTRW describes a subdiffusive behavior. The classical approach to CTRW yields a non-Markovian (mean-field) transport equation, which is a serious obstacle when one wants to couple subdiffusion with (bio)chemical reaction. We took an alternative approach to CTRW that maintains the Markovian property of the transport equation at the price of a supplementary independent variable. We associate each random walker with an age $a$, that is the time elapsed since its last jump and describe the subdiffusive CTRW using an age-structured partial differential equations with age renewal upon each walker jump. In the spatially-homogeneous (zero-dimensional) case, we follow the evolution in time of the age distribution. An approach inspired by relative entropy techniques allows us to obtain quantitative explicit rates for the convergence of the age distribution to a self-similar profile, which corresponds to convergence to a stationary profile for the rescaled variables. An important difficulty arises from the fact that the equation in self-similar variables is not autonomous and we do not have a specific analytical solution. Therefore, in order to quantify the latter convergence, we estimate attraction to a time-dependent "pseudo-equilibrium", which in turn converges to the stationary profile.
The corresponding article is currently in press [38].

## 7.7. IGF-I signalling in neural stem cells during neurogenesis and aging

Downregulation of insulin-like growth factor (IGF) pathways prolongs lifespan in various species, including mammals. Still, the cellular mechanisms by which IGF signaling controls the aging trajectory of individual organs are largely unknown. Z. Chaker, in M. Holzenberg Lab (Centre de Recherche Saint-Antoine, Paris), asked whether suppression of IGF-I receptor (IGF-1R) in adult stem cells preserves long-term cell replacement, and whether this may prevent age-related functional decline in a regenerating tissue. Using neurogenesis as a paradigm, we showed that conditional knockout of IGF-1R specifically in adult neural stem cells maintained youthful characteristics of olfactory bulb neurogenesis within an aging brain. This in turn resulted in neuro-anatomical changes that improved olfactory function. To help interpret these results, we developed a mathematical model of stem cell differentiation using ordinary differential equations with time-dependent growth, division and death rates (to account for aging) and optimizing at each time step the amount of IGF-IR to maximize an experimentally-derived tissue efficiency criterion. The model predicts that decreased stimulation of growth in adults is indeed optimal for tissue aging. Thus, inhibiting growth and longevity gene IGF-1R in adult stem cells induced a gain-of-function phenotype during aging, marked by optimized management of cell renewal, and enhanced olfactory sensory function.
This work has been published in the article [14].

## 7.8. A novel model for leptin resistance

Leptin is a major hormone that regulates food intake and appetite in most mammals. Leptin increase in the blood tends to decrease the food intake and leptin is produced in proportion with fat depot. Leptin is therefore a simple probe that feed backs energy reserve to the brain and maintains a constant weight. It is a central hormone for this balance because KO mice without the leptin gene are quickly extremely obese. Also obese people (and animal) tend to have high concentration of leptin suggesting that after a certain point the brain ignores the leptin signal. We developed a mathematical model that explores this resistance developed by neural cells to leptin. This model predicts leptin resistance if food intake is artificially increased and predict a pathway to obesity by such mechanism. This work has been published by H. Soula (Beagle) in collaboration with F. Crauste (Dracula) with co-supervised PhD student Marine Jacquier[19].

## 7.9. Without eye contact, birds are Markovian!

Any social birds rely on acoustic messages to organize their daily activity (such as parenting and food foraging). In many occasions, birds are within earshot but not in visual contact and therefore should rely only on acoustic channel for this communication. In collaboration with the University of Saint-Etienne, we developed automatic extraction scripts that can detect birds vocalizations in a protocol of meeting with decreasing distance and with or without visual contact modality. Our worked showed that without visual contact birds are more synchronized and their vocal dynamics cannot be distinguished from a two state Markov chain. This markov property vanishes as soon as visual contact is restablished. This work has been published in the main ethology journal: Animal Behaviour[23].

<p style="text-align:center; color:red;">**DRACULA Project-Team**</p>

# 6. New Results

## 6.1. Implication of the autologous immune system in BCR-ABL transcript variations in chronic myelogenous leukemia patients treated with Imatinib

Imatinib (IM) and other tyrosine kinase inhibitors (TKI) have improved treatment of chronic myelogenous leukemia (CML); however, most patients are not cured. Deeper mechanistic understanding may improve TKI combination therapies to better control the residual leukemic cell population. In analyzing our patients' data, we found that many patients who otherwise responded well to IM therapy still showed variations in their BCR-ABL transcripts. To investigate this phenomenon, we applied a mathematical model (see [14]) that integrates CML and an autologous immune response to the patients' data. We define an immune window, or a range of leukemic loads for which the autologous immune system induces an improved response. Our modeling results in [14], suggest that, at diagnosis, a patient's leukemic load is able to partially or fully suppress the autologous immune response developed in a majority of patients, towards the CML clone(s). IM therapy drives the leukemic population into the "immune window", allowing the patient's autologous immune cells to expand and eventually mount an efficient recognition of the residual leukemic burden. This response drives the leukemic load below this immune window, allowing the leukemic population to partially recover until another weaker immune response is initiated. Thus, the autologous immune response may explain the oscillations in BCR-ABL transcripts regularly observed in patients on IM.

## 6.2. Predicting pathogen-specific CD8 T cell immune responses from a modeling approach

The primary CD8 T cell immune response constitutes a major mechanism to fight an infection by intra-cellular pathogens. We aim at assessing whether pathogen-specific dynamical parameters of the CD8 T cell response can be identified, based on measurements of CD8 T cell counts, using a modeling approach. We generated experimental data consisting in CD8 T cell counts kinetics during the response to three different live intra-cellular pathogens: two viruses (influenza, vaccinia) injected intranasally, and one bacteria (Listeria monocytogenes) injected intravenously. All pathogens harbor the same antigen (NP68), but differ in their interaction with the host. In parallel, we developed in [16] a mathematical model describing the evolution of CD8 T cell counts and pathogen amount during an immune response. This model is characterized by 9 parameters and includes relevant feedback controls. The model outputs were compared with the three data series and an exhaustive estimation of the parameter values was performed. By focusing on the ability of the model to fit experimental data and to produce a CD8 T cell population mainly composed of memory cells at the end of the response, critical parameters were identified. We show that a small number of parameters $(2 - 4)$ define the main features of the CD8 T cell immune response and are characteristic of a given pathogen. Among these parameters, two are related to the effector CD8 T cell mediated control of cell and pathogen death. The parameter associated with memory cell death is shown to play no relevant role during the main phases of the CD8 T cell response, yet it becomes essential when looking at the predictions of the model several months after the infection.

## 6.3. Dynamics of cell generation and turnover in the human heart

The contribution of cell generation to physiological heart growth and maintenance in humans has been difficult to establish and has remained controversial. We report in [8] that the full complement of cardiomyocytes is established perinataly and remains stable over the human lifespan, whereas the numbers of both endothelial and mesenchymal cells increase substantially from birth to early adulthood. Analysis of the integration of

nuclear bomb test-derived $14C$ revealed a high turnover rate of endothelial cells throughout life ($> 15\%$ per year) and more limited renewal of mesenchymal cells ($< 4\%$ per year in adulthood). Cardiomyocyte exchange is highest in early childhood and decreases gradually throughout life to $< 1\%$ per year in adulthood, with similar turnover rates in the major subdivisions of the myocardium. We provide an integrated model of cell generation and turnover in the human heart.

## 6.4. Travelling waves of cell differentiation

The paper [7] is devoted to modelling of cell differentiation in an initially homogeneous cell population. The mechanism which provides coexistence of two cell lineages in the initially homogeneous cell population is suggested. If cell differentiation is initiated locally in space in the population of undifferentiated cells, it can propagate as a travelling wave converting undifferentiated cells into differentiated ones. We suggest a model of this process which takes into account intracellular regulation, extracellular regulation and different cell types. They include undifferentiated cells and two types of differentiated cells. When a cell differentiates, its choice between two types of differentiated cells is determined by the concentrations of intracellular proteins. Differentiated cells can either stimulate differentiation into their own cell lineage or into another cell lineage. In the case of the positive feedback, only one lineage of differentiated cells will finally appear. In the case of negative feedback, both of them can coexist. In this case a periodic spatial pattern emerges behind the wave.

## 6.5. Pattern regeneration based on cell memory

In [24], we present a new model of the cellular dynamics that enable regeneration of complex biological morphologies. Biological cell structures are considered as an ensemble of mathematical points on the plane. Each cell produces a signal which propagates in space and is received by other cells. The total signal received by each cell forms a signal distribution defined on the cell structure. This distribution characterizes the geometry of the cell structure. If a part of this structure is removed, the remaining cells have two signals. They keep the value of the signal which they had before the amputation (memory), and they receive a new signal produced after the amputation. Regeneration of the cell structure is stimulated by the difference between the old and the new signals. It is stopped when the two signals coincide. The algorithm of regeneration contains certain rules which are essential for its functioning, being the first quantitative model of cellular memory that implements regeneration of complex patterns to a specific target morphology. Correct regeneration depends on the form and the size of the cell structure, as well as on some parameters of regeneration.

## 6.6. Target morphology and cell memory

Despite the growing body of work on molecular components required for regenerative repair, we still lack a deep understanding of the ability of some animal species to regenerate their appropriate complex anatomical structure following damage. A key question is how regenerating systems know when to stop growth and remodeling – what mechanisms implement recognition of correct morphology that signals a stop condition? In [11], we review two conceptual models of pattern regeneration that implement a kind of pattern memory. In the first one, all cells communicate with each other and keep the value of the total signal received from the other cells. If a part of the pattern is amputated, the signal distribution changes. The difference from the original signal distribution stimulates cell proliferation and leads to pattern regeneration, in effect implementing an error minimization process that uses signaling memory to achieve pattern correction. In the second model, we consider a more complex pattern organization with different cell types. Each tissue contains a central (coordinator) cell that controls the tissue and communicates with the other central cells. Each of them keeps memory about the signals received from other central cells. The values of these signals depend on the mutual cell location, and the memory allows regeneration of the structure when it is modified. The purpose of these models is to suggest possible mechanisms of pattern regeneration operating on the basis of cell memory which are compatible with diverse molecular implementation mechanisms within specific organisms.

## 6.7. Transplanted bone marrow-derived cells contribute to human adipogenesis

Because human white adipocytes display a high turnover throughout adulthood, a continuous supply of precursor cells is required to maintain adipogenesis. Bone marrow (BM)-derived progenitor cells may contribute to mammalian adipogenesis; however, results in animal models are conflicting. In [22], we demonstrate in 65 subjects who underwent allogeneic BM or peripheral blood stem cell (PBSC) transplantation that, over the entire lifespan, BM/PBSC-derived progenitor cells contribute $10\%$ to the subcutaneous adipocyte population. While this is independent of gender, age, and different transplantation-related parameters, body fat mass exerts a strong influence, with up to $2.5$-fold increased donor cell contribution in obese individuals. Exome and whole-genome sequencing of single adipocytes suggests that BM/PBSC-derived progenitors contribute to adipose tissue via both differentiation and cell fusion. Thus, at least in the setting of transplantation, BM serves as a reservoir for adipocyte progenitors, particularly in obese subjects.

## 6.8. Modelling of platelet–fibrin clot formation in flow

The paper [23] is devoted to mathematical modelling of clot growth in blood flow. Great complexity of the hemostatic system dictates the need of usage of the mathematical models to understand its functioning in the normal and especially in pathological situations. In this work we investigate the interaction of blood flow, platelet aggregation and plasma coagulation. We develop a hybrid DPD–PDE model where dissipative particle dynamics (DPD) is used to model plasma flow and platelets, while the regulatory network of plasma coagulation is described by a system of partial differential equations. Modelling results confirm the potency of the scenario of clot growth where at the first stage of clot formation platelets form an aggregate due to weak inter-platelet connections and then due to their activation. This enables the formation of the fibrin net in the centre of the platelet aggregate where the flow velocity is significantly reduced. The fibrin net reinforces the clot and allows its further growth. When the clot becomes sufficiently large, it stops growing due to the narrowed vessel and the increase of flow shear rate at the surface of the clot. Its outer part is detached by the flow revealing the inner part covered by fibrin. This fibrin cap does not allow new platelets to attach at the high shear rate, and the clot stops growing. Dependence of the final clot size on wall shear rate and on other parameters is studied.

## 6.9. Conceptual model of morphogenesis and regeneration

The paper [24] is devoted to computer modelling of the development and regeneration of multicellular biological structures. Some species (e.g. planaria and salamanders) are able to regenerate parts of their body after amputation damage, but the global rules governing cooperative cell behaviour during morphogenesis are not known. Here, we consider a simplified model organism, which consists of tissues formed around special cells that can be interpreted as stemcells. We assume that stem cells communicate with each other by a set of signals, and that the values of these signals depend on the distance between cells. Thus the signal distribution characterizes location of stem cells. If the signal distribution is changed, then the difference between the initial and the current signal distribution affects the behaviour of stem cells—e.g. as a result of an amputation of a part of tissue the signal distribution changes which stimulates stem cells to migrate to new locations, appropriate for regeneration of the proper pattern. Moreover, as stem cells divide and form tissues around them, they control the form and the size of regenerating tissues. This two-level organization of the model organism, with global regulation of stem cells and local regulation of tissues, allows its reproducible development and regeneration.

## 6.10. Delay differential-difference system for hematopoietic stem cell dynamics

We investigate in [2] and [3] a mathematical model of hematopoietic stem cell dynamics. We take two cell populations into account, quiescent and proliferating one, and we note the difference between dividing cells that enter directly to the quiescent phase and dividing cells that return to the proliferating phase to divide again. The resulting mathematical model is a system of two age-structured partial differential equations. By

integrating this system over age and using the characteristics method, we reduce it to a delay differential-difference system, and we investigate the existence and stability of the steady states. We give sufficient conditions for boundedness and unboundedness properties for the solutions of this system. By constructing a Lyapunov function, the trivial steady state, describing cell's dying out, is proven to be globally asymptotically stable when it is the only equilibrium. The stability analysis of the unique positive steady state, the most biologically meaningful one, and the existence of a Hopf bifurcation allow the determination of a stability area, which is related to a delay-dependent characteristic equation. Numerical simulations illustrate our results on the asymptotic behavior of the steady states and show very rich dynamics of this model. This study may be helpful in understanding the uncontrolled proliferation of blood cells in some hematological disorders.

## 6.11. Discrete limit and monotonicity properties of the Floquet eigenvalue in an age structured cell division cycle model

We consider in [19] a cell population described by an age-structured partial differential equation with time periodic coefficients. We assume that division only occurs after a minimal age (majority) and within certain time intervals. We study the asymptotic behavior of the dominant Floquet eigenvalue, or Perron-Frobenius eigenvalue, representing the growth rate, as a function of the majority age, when the division rate tends to infinity (divisions become instantaneous). We show that the dominant Floquet eigenvalue converges to a staircase function with an infinite number of steps, determined by a discrete dynamical system. As an intermediate result, we give a structural condition which guarantees that the dominant Floquet eigenvalue is a nondecreasing function of the division rate. We also give a counter example showing that the latter monotonicity property does not hold in general.

## 6.12. Optimal linear stability condition for scalar differential equations with distributed delay

Linear scalar differential equations with distributed delays appear in the study of the local stability of nonlinear differential equations with feedback, which are common in biology and physics. Negative feedback loops tend to promote oscillations around steady states, and their stability depends on the particular shape of the delay distribution. Since in applications the mean delay is often the only reliable information available about the distribution, it is desirable to find conditions for stability that are independent from the shape of the distribution. We show in [9] that for a given mean delay, the linear equation with distributed delay is asymptotically stable if the associated differential equation with a discrete delay is asymptotically stable. We illustrate this criterion on a compartment model of hematopoietic cell dynamics to obtain sufficient conditions for stability.

## 6.13. A mathematical model of leptin resistance

Obesity is often associated with leptin resistance, which leads to a physiological system with high leptin concentration but unable to respond to leptin signals and to regulate food intake. We propose in [20] a mathematical model of the leptin-leptin receptors system, based on the assumption that leptin is a regulator of its own receptor activity, and investigate its qualitative behavior. Based on current knowledge and previous models developed for body weight dynamics in rodents, the model includes the dynamics of leptin, leptin receptors and the regulation of food intake and body weight. It displays two stable equilibria, one representing a healthy state and the other one an obese and leptin resistant state. We show that a constant leptin injection can lead to leptin resistance and that a temporal variation in some parameter values influencing food intake can induce a change of equilibrium and a pathway to leptin resistance and obesity.

<div align="center">

## ERABLE Project-Team

</div>

# 6. New Results

## 6.1. General comments

We present in this section the main results obtained in 2015. Some were already in preparation or submitted at the end of 2014. It will be indicated whenever this is the case.

We tried to organise the results following four of the five main axes of research of the team. Clearly, in some cases, a result obtained overlaps more than one axis. We chose the one that could be seen as the main concerned by such results. As concerns the Axis "Going towards control", a work is in preparation that fits it. It will be presented in 2016.

We did not indicate here the results on more theoretical aspects of computer science if it did not seem for now that they could be relevant in contexts related to computational biology. Actually, we do believe those on scheduling (by, among others, A. Marchetti-Spaccamela and/or L. Stougie) [3], [38], [39], [10], [31], [23], [44], [43] or even one result related to context-free grammars (by, among others, P. Crescenzi) [11] could in the future become relevant for the life sciences (biology or ecology). However, we preferred for now to only indicate the theoretical results related to problems closely resembling questions that have already been addressed by us in computational biology.

Notice that such CS results concern not only cross-fertilising issues among different computational approaches, and we therefore extended the title of this axis for the purpose of presenting such results, for now purely theoretical.

A few other results are not mentioned either, not because the corresponding work is not important, but because it was likewise more specialised, or the work represented a survey.

## 6.2. Identifying the molecular elements

### Genomic / NGS data management

Next-generation sequencing (NGS) technology has led the life sciences into the big data era. Today, sequencing genomes takes little time and cost, but yields terabytes of data to be stored and analysed. The biologists are often exposed to excessively time consuming and error-prone data management and analysis hurdles. We therefore proposed a database management system (DBMS) based approach to accelerate and substantially simplify genome sequence analysis [9]. To that aim, we extended MONETDB, an open-source column-based DBMS (urlhttps://www.monetdb.org), with a BAM module, which enables easy, flexible, and rapid management and analysis of sequence alignment data stored as Sequence Alignment/Map (SAM/BAM) files. The main features of MONETDB/BAM were described using a case study on Ebola virus.

We also designed and realised a knowledge base for collecting, elaborating, and extracting analytical results of genomic, proteomic, biochemical, morphological investigations from animal models of cerebral stroke [45]. Data analysis techniques are tailored to make the data available for processing and correlation, in order to increase the predictive value of the preclinical data, to perform bio-simulation studies, and to support both academic and industrial research in the area of cerebral stroke therapy. The low reliability of animal models in replicating the human disease is one of the most serious problems in the field of medical and pharmaceutical research about stroke. The standard models for the study of ischaemic stroke are often poorly predictive as they simulate only partially the human disease. This work aims therefore at investigating animal models with diseases typically associated with the onset of stroke in human patients. A first statistical analysis of the retrieved information led to the validation of our animal models and suggested a predictive and translational value for parameters related to a specific model. In particular, concerning gene expression data, we applied a data analysis pipeline that initially takes into account an initial set of 64,000 genes and brought down the focus on a few tens of them.

**NGS data analysis**

The problem of enumerating bubbles with length constraints in directed graphs arises in transcriptomics where the question is to identify all alternative splicing events present in a sample of mRNAs sequenced by RNA?seq. We presented a new algorithm for enumerating bubbles with length constraints in weighted directed graphs [30]. This is the first polynomial delay algorithm for this problem and we showed that in practice, it is faster than previous approaches. This settled one of the main open questions from previous literature. Moreover, the new algorithm allows us to deal with larger instances and possibly detect longer alternative splicing events.

We also developed CIDANE, a novel framework for genome-based transcript reconstruction and quantification from RNA-seq reads [37]. CIDANE assembles transcripts with significantly higher sensitivity and precision than existing tools, while competing in speed with the fastest methods. In addition to reconstructing transcripts *ab initio*, the algorithm also allows to make use of the growing annotation of known splice sites, transcription start and end sites, or full-length transcripts, which are available for most model organisms. CIDANE supports the integrated analysis of RNA-seq and additional gene-boundary data and recovers splice junctions that are invisible to other methods.

SNPs (Single Nucleotide Polymorphisms) are genetic markers used in many areas of biology. Their precise identification is a prerequisite for association studies, which associate genotypes to phenotypes. Methods are currently developed for model species, but rely on the availability of a (good) reference genome, and cannot be applied to non-model species. They are also mostly tailored for whole genome (re-)sequencing experiments, whereas in many cases, transcriptome sequencing can be used as a cheaper alternative which already enables to identify SNPs located in transcribed regions. We proposed a method that identifies, quantifies and annotates SNPs without any reference genome, using RNA-seq data only. Individuals can be pooled prior to sequencing, if not enough material is available for sequencing from one individual. This pooling strategy still enables to allelotype loci and to associate them to phenotypes. Using human RNA-seq data, we first compared the performance of our algorithm, KISSSPLICE, with GATK, a well established method that requires a reference genome. We showed that both methods perform similarly in terms of precision and recall. We then validated experimentally the predictions of our method using RNA-seq data from two non-model species. The method can be used for any species to annotate SNPs and to predict their impact on proteins. It can further be used to assess variants that are associated to a particular phenotype within a population, when replicates are provided for each biological condition. This work was submitted at the end of 2015.

**Sequence alignment (full genomes or NGS data)**

Sequence comparison is a fundamental step in many important tasks related to biology. Traditional algorithms for measuring approximation in sequence comparison are based on the notions of distance or similarity, and are generally computed through sequence alignment techniques. As circular genome structure is a common phenomenon in nature, a caveat of specialised alignment techniques for circular sequence comparison is that they are computationally expensive, requiring from super-quadratic to cubic time in the length of the sequences. We introduced a new distance measure based on $q$-grams, and showed how it can be computed efficiently for circular sequence comparison [41]. Experimental results, using real and synthetic data, demonstrated orders-of-magnitude superiority of our approach in terms of efficiency, while maintaining an accuracy very competitive to the state of the art.

Burrows-Wheeler Transform (BWT) has been successfully used to reduce the memory requirement for sequence alignment. We improved on previous results related to the problem of computing the Burrows-Wheeler Transform (BWT) using small additional space [12]. Our in-place algorithm does not need the explicit storage for the suffix sort array and the output array, as typically required in such previous work. It relies on the combinatorial properties of the BWT, and runs in $O(n^2)$ time in the comparison model using $O(1)$ extra memory cells, apart from the array of $n$ cells storing the $n$ characters of the input text. We then discussed the time-space trade-off when $O(k\sigma k)$ extra memory cells are allowed with $\sigma k$ distinct characters, providing an $O((n^2/k + n)\log?k)$-time algorithm to obtain (and invert) the BWT. In real systems where the alphabet size is a constant, for any arbitrarily small $\epsilon > 0$, the BWT of a text of $n$ bytes can be computed in $O(n\sigma^{-1}\log n)$ time using just $\sigma n$ extra bytes.

**Genome assembly problems**

The human genome is diploid, which requires assigning heterozygous single nucleotide polymorphisms (SNPs) to the two copies of the genome. The resulting haplotypes, lists of SNPs belonging to each copy, are crucial for downstream analyses in population genetics. Currently, statistical approaches, which are oblivious to direct read information, constitute the state-of-the-art. Haplotype assembly, which addresses phasing directly from sequencing reads, suffers from the fact that sequencing reads of the current generation are too short to serve the purposes of genome-wide phasing. While future-technology sequencing reads will contain sufficient amounts of SNPs per read for phasing, they are also likely to suffer from higher sequencing error rates. Currently, no haplotype assembly approaches exist that allow for taking both increasing read length and sequencing error information into account. We developed WHATSHAP, the first approach that yields provably optimal solutions to the weighted minimum error correction problem in runtime linear in the number of SNPs [25]. WHATSHAP is a fixed parameter tractable (FPT) approach with coverage as the parameter. We demonstrated that WHATSHAP can handle datasets of coverage up to 20x, and that 15x are generally enough for reliably phasing long reads, even at significantly elevated sequencing error rates. We also find that the switch and flip error rates of the haplotypes we output are favourable when comparing them with state-of-the-art statistical phasers. By using novel combinatorial properties of Minimum Error Correction (MEC) instances, we were then able to provide new results on the fixed-parameter tractability and approximability of MEC [35]. In particular, we showed that MEC is in FPT when parameterised by the number of corrections, and, on "gapless" instances, it is in FPT also when parameterised by the length of the fragments, whereas the result known in the literature forces the reconstruction of complementary haplotypes. We then showed that MEC cannot be approximated within any constant factor while it is approximable within factor $O(\log nm)$ where $nm$ is the size of the input. Finally, we provided a practical 2-approximation algorithm for the Binary MEC, a variant of MEC that has been applied in the framework of clustering binary data. Finally, by exploiting a feature of future-generation technologies – the uniform distribution of sequencing errors – we designed an exact algorithm, called HAPCOL, that is exponential in the maximum number of corrections for each SNP position and that minimises the overall error-correction score [26]. We performed an experimental analysis, comparing HAPCOL with the current state-of-the-art combinatorial methods both on real and simulated data. On a standard benchmark of real data, we showed that HAPCOL is competitive with state-of-the-art methods, improving the accuracy and the number of phased positions. Furthermore, experiments on realistically-simulated datasets revealed that HAPCOL requires significantly less computing resources, especially memory. Thanks to its computational efficiency, HAPCOL can overcome the limits of previous approaches, allowing to phase datasets with higher coverage and without the traditional all-heterozygous assumption.

Completing the genome sequence of an organism is an important task in comparative, functional and structural genomics. However, this remains a challenging issue from both a computational and an experimental viewpoint. Genome scaffolding (*i.e.* the process of ordering and orientating contigs) of de novo assemblies usually represents the first step in most genome finishing pipelines. We developed MEDUSA (Multi-Draft based Scaffolder), an algorithm for genome scaffolding [6]. MEDUSA exploits information obtained from a set of (draft or closed) genomes from related organisms to determine the correct order and orientation of the contigs. MEDUSA formalises the scaffolding problem by means of a combinatorial optimisation formulation on graphs and implements an efficient constant factor approximation algorithm to solve it. In contrast to currently used scaffolders, it does not require either prior knowledge on the microorganisms dataset under analysis (*e.g.* their phylogenetic relationships) or the availability of paired end read libraries. This makes usability and running time two additional important features of our method. Moreover, benchmarks and tests on real bacterial datasets showed that MEDUSA is highly accurate and, in most cases, outperforms traditional scaffolders. The possibility to use MEDUSA on eukaryotic datasets has also been evaluated, leading to interesting results. medusa/releases.

**Genome annotation problems**
Repetitive DNA, including transposable elements (TEs), is found throughout eukaryotic genomes. Annotating and assembling the "repeatome" during genome-wide analysis often poses a challenge. To address this problem, we developed DNAPIPETE – a new pipeline that uses a sample of raw genomic reads [20]. It produces precise estimates of repeated DNA content and TE consensus sequences, as well as the relative ages of TE families. We showed that DNAPIPETE performs well using very low coverage sequencing in different

genomes, losing accuracy only with old TE families. We applied this pipeline to the genome of the Asian tiger mosquito *Aedes albopictus*, an invasive species of human health interest, for which the genome size is estimated to be over 1 Gbp. Using DNAPIPETE, we showed that this species harbours a large (50% of the genome) and potentially active repeatome with an overall TE class and order composition similar to that of *Aedes aegypti*, the yellow fever mosquito. However, intra-order dynamics showed clear distinctions between the two species, with differences at the TE family level. Our pipeline's ability to manage the repeatome annotation problem will make it helpful for new or ongoing assembly projects, and our results will benefit future genomic studies of *A. albopictus*.

On another topic, we developed a reliable, robust, and much faster method for the prediction of pre-miRNAs [22]. With this method, we aimed mainly at two goals: efficiency and flexibility. Efficiency was made possible by means of a quadratic algorithm. Since the majority of the predictors use a cubic algorithm to verify the pre-miRNA hairpin structure, they may take too long when the input is large. Flexibility relies on two aspects, the input type and the organism clade. MIRINHO can receive as input both a genome sequence and small RNA sequencing (sRNA-seq) data of both animal and plant species. To change from one clade to another, it suffices to change the lengths of the stem-arms and of the terminal loop. Concerning the prediction of plant miRNAs, because their pre-miRNAs are longer, the methods for extracting the hairpin secondary structure are not as accurate as for shorter sequences. With MIRINHO, we also addressed this problem, which enabled to provide premiRNA secondary structures more similar to the ones in MIRBASE than the other available methods. MIRINHO served also as the basis to two other issues we addressed. The first issue led to the treatment and analysis of sRNA-seq data of *Acyrthosiphon pisum*, the pea aphid. The goal was to identify the miRNAs that are expressed during the four developmental stages of this species, allowing further biological conclusions concerning the regulatory system of such an organism. For this analysis, we developed a whole pipeline, called MIRINHOPIPE, at the end of which MIRINHO was aggregated. A paper is currently being prepared that presents this work.

## 6.3. Inferring and analysing the networks of molecular elements

**Protein structure comparison**

We proposed a new distance measure for comparing two protein structures based on their contact map representations [1]. We showed that our novel measure, which we refer to as the maximum contact map overlap (max-CMO) metric, satisfies all properties of a metric on the space of protein representations. Having a metric in that space allows one to avoid pairwise comparisons on the entire database and, thus, to significantly accelerate exploring the protein space compared to no-metric spaces. We showed on a gold standard superfamily classification benchmark set of 6759 proteins that our exact k-nearest neighbour ($k - NN$) scheme classifies up to 224 out of 236 queries correctly and on a larger, extended version of the benchmark with 60; 850 additional structures, up to 1361 out of 1369 queries. Our $k - NN$ classification thus provides a promising approach for the automatic classification of protein structures based on flexible contact map overlap alignments.

**Metabolic network analysis**

Flux balance analysis (FBA) is one of the most often applied methods on genome-scale metabolic networks. Although FBA uniquely determines the optimal yield, the pathway that achieves this is usually not unique. The analysis of the optimal-yield flux space has been an open challenge. Flux variability analysis is only capturing some properties of the flux space, while elementary mode analysis is intractable due to the enormous number of elementary modes. However, it had been previously found that the space of optimal-yield fluxes decomposes into flux modules. These decompositions allow a much easier but still comprehensive analysis of the optimal-yield flux space. Using the mathematical definition of module introduced by Müller and Bockmayr in 2013, we discovered that flux modularity is rather a local than a global property which opened connections to matroid theory [28]. Specifically, we showed that our modules correspond one-to-one to so-called separators of an appropriate matroid. Employing efficient algorithms developed in matroid theory we are now able to compute the decomposition into modules in a few seconds for genome-scale networks. Using that every module can be represented by one reaction that corresponds to its function, we also presented a method that uses this

decomposition to visualise the interplay of modules. We expect the new method to replace flux variability analysis in the pipelines for metabolic networks.

**Integrated network analysis**

Data on molecular interactions is increasing at a tremendous pace. Since biological functionality primarily operates at the network level, there is a clear need for topology-aware comparison methods. We developed one such method for global network alignment that is fast and robust and can flexibly deal with various scoring schemes taking both node-to-node correspondences as well as network topologies into account [18]. We exploited that network alignment is a special case of the well-studied quadratic assignment problem (QAP). We focused on sparse network alignment, where each node can be mapped only to a typically small subset of nodes in the other network. This corresponds to a QAP instance with a symmetric and sparse weight matrix. We obtained strong upper and lower bounds for the problem by improving a Lagrangian relaxation approach and introduce the open source software tool NATALIE 2.0, a publicly available implementation of our method (https://github.com/ls-cwi/natalie). In an extensive computational study on protein interaction networks for six different species, we found that our new method outperforms alternative established and recent state-of-the-art methods.

Integrative network analysis methods provide robust interpretations of differential high-throughput molecular profile measurements. They are often used in a biomedical context-to generate novel hypotheses about the underlying cellular processes or to derive biomarkers for classification and subtyping. The underlying molecular profiles are frequently measured and validated on animal or cellular models. Therefore the results are not immediately transferable to human. In particular, this is also the case in a study of the recently discovered interleukin-17 producing helper T cells (Th17), which are fundamental for anti-microbial immunity but also known to contribute to autoimmune diseases. We proposed a mathematical model for finding active subnetwork modules that are conserved between two species [19]. These are sets of genes, one for each species, which (1) induce a connected subnetwork in a species-specific interaction network, (2) show overall differential behaviour and (iii) contain a large number of orthologous genes. We proposed a flexible notion of conservation, which turns out to be crucial for the quality of the resulting modules in terms of biological interpretability. We developed an algorithm that finds provably optimal or near-optimal conserved active modules in our model. We applied our algorithm to understand the mechanisms underlying Th17 T cell differentiation in both mouse and human. As a main biological result, we found that the key regulation of Th17 differentiation is conserved between human and mouse.

## 6.4. Modelling and analysing a network of individuals, or a network of individuals' networks

**Computationally investigating co-phylogenetic reconstructions and co-evolution**

Despite an increasingly vast literature on co-phylogenetic reconstructions for studying host-symbiont associations, understanding the common evolutionary history of such systems remains a problem that is far from being solved. Most algorithms for host-symbiont reconciliation use an event-based model, where the events include in general (a subset of) co-speciation, duplication, loss, and host-switch. All known parsimonious event-based methods then assign a cost to each type of event in order to find a reconstruction of minimum cost. The main problem with this approach is that the cost of the events strongly influences the reconciliation obtained.To deal with this problem, we developed an algorithm, called COALA, for estimating the frequency of the events based on an approximate Bayesian computation approach [4]. The benefits of this method are twofold: (1) it provides more confidence in the set of costs to be used in a reconciliation, and (2) it allows estimation of the frequency of the events in cases where the dataset consists of trees with a large number of taxa. We evaluated our method on simulated and on biological datasets. We showed that in both cases, for the same pair of host and parasite trees, different sets of frequencies for the events lead to equally probable solutions. Moreover, often these solutions differ greatly in terms of the number of inferred events. It appears crucial to take this into account before attempting any further biological interpretation of such reconciliations. More generally, we also showed that the set of frequencies can vary widely depending on the input host and parasite trees. Indiscriminately applying a standard vector of costs may thus not be a good strategy. This work had been indicated as submitted in 2014.

Once such a cost vector has been inferred, one can proceed analysing the possible co-evolution of host-symbiont associations, phylogenetic tree reconciliation is the approach of choice for investigating the co-evolution of sets of organisms such as hosts and parasites. It consists in a mapping between the parasite tree and the host tree using event-based maximum parsimony. Given a cost model for the events, many optimal reconciliations are however possible. Only two algorithms existed that attempted such enumeration; in one case not all possible solutions are produced while in the other not all cost vectors are currently handled. We developed a polynomial-delay algorithm, EUCALYPT, for enumerating all optimal reconciliations that address these two issues [15]. We showed that in general many solutions exist. We gave an example where, for two pairs of host-parasite trees having each less than 41 leaves, the number of solutions is 5120, even when only time-feasible ones are kept. To facilitate their interpretation, those solutions are also classified in terms of how many of each event they contain. The number of different classes of solutions may thus be notably smaller than the number of solutions, yet they may remain high enough, in particular for the cases where losses have cost 0. In fact, depending on the cost vector, both numbers of solutions and of classes thereof may increase considerably (for the same instance, to respectively 4080384 and 275). To further deal with this problem, we introduced and analysed a restricted version where host-switches are allowed to happen only between species that are within some fixed distance along the host tree. This restriction allowed us to reduce the number of time-feasible solutions while preserving the same optimal cost, as well as to find time-feasible solutions with a cost close to the optimal in the cases where no time-feasible solution is found. This work had been indicated as submitted in 2014.

**Evolution and metabolic complementation of organisms leaving inside the cells of another (endosymbionts)**

Insect cells host many endosymbiotic bacteria, which are in general classified according to their importance for the host: "primary" symbionts are by definition mandatory and synthesise essential nutrients for the insects that feed on poor or unbalanced food sources, while "secondary" symbionts are optional and use mutualistic strategies and/or manipulation of reproduction to invade and persist within insect populations.

*Hamiltonella defensa* is a secondary endosymbiont that established two distinct associations with phloemophagous insects. In aphids, it protects the host against parasitoid attacks. Its ability to infect many host tissues, notably the hemolymph, could promote its contact with parasitoid eggs. Despite this protective phenotype, the high costs associated with its presence within the host prevent its fixation in the population. In the whitefly *Bemisia tabaci* however, this symbiont is found only in cells specialised in hosting endosymbionts, the bacteriocytes. In these cells, it cohabits with other symbiotic species, such as the primary symbiont *Portiera aleyrodidarum*, a proximity that favours potential exchanges between the two symbionts. It is fixed in populations of *B. tabaci*, which suggests an important role for the consortium, probably nutritious.

We studied the specificities of each of these systems [27]. First, in the bacteriocytes of *B. tabaci*, we identified a partitioning of the synthetic capacities of two endosymbionts, *H. defensa* and *P. aleyrodidarum*, in addition to a potential metabolic complementation between the symbionts and their host for the synthesis of essential amino acids. We proposed a key nutritive role for *H. defensa*, which would indicate a transition to a mandatory status in relation to the host and would explain its fixation in the population.

We also focused on the genomic evolution of the genus *Hamiltonella*, by comparing the strains infecting *B. tabaci* with a strain infecting the aphids [29]. We highlighted the specialisation of the symbionts to their hosts, and found that the genomes of the endosymbionts reflected their respective ecology. The aphid strain thus possesses many virulence factors and is associated with two partners, a bacteriophage and a recombination plasmid. These systems, inactive in the symbiont of *B. tabaci*, are directly related to the protection against and arms race with parasitoids. Conversely, the presumed avirulence of whitefly endosymbionts is consistent with their nutritional phenotype and a transition to a mandatory status to the host.

Finally, we studied the phenomenon of "accelerated mutation rate" in *H. defensa*, compared to its sister species *Regiella insecticola*, which is also a clade of protective endosymbionts of aphids. After excluding the assumption that the transition to the intracellular life occurred independently in the two lineages, we tried to establish a link between these differences in terms of evolvability in the endosymbionts and of their gene contents, particularly for genes involved in ecology and DNA repair. All the results obtained have provided

insight into the evolution of the species *H. defensa*, since the last ancestor to the present species, by establishing a link between bacterial.

These results were part of the PhD of Pierre-Antoine Rollat-Farnier, co-supervised by Laurence Mouton (LBBE, UMR5558), Marie-France Sagot (Inria and LBBE, UMR5558) and Fabrice Vavre (LBBE, UMR5558) and defended on November 24th, 2014. The results had been indicated as submitted in 2014.

**Insights on the virulence of swine respiratory tract mycoplasmas through genome-scale metabolic modelling**

The respiratory tract of swines is colonised by several bacteria among which are three *Mycoplasma* species: *Mycoplasma flocculare*, *Mycoplasma hyopneumoniae* and *Mycoplasma hyorhinis*. While colonisation by *M. flocculare* was shown to be virtually asymptomatic, *M. hyopneumoniae* is known to be the causative agent of enzootic pneumonia and *M. hyorhinis* to be present in cases of pneumonia, polyserositis and arthritis. Nonetheless, the elevated genomic resemblance among these three mycoplasmas combined with their different levels of pathogenicity is an indication that they have unknown mechanisms of virulence and differential expression. We performed whole-genome metabolic network reconstructions for these three mycoplasmas and were able to show that overall they have similar metabolic capabilities. The metabolic differences that were observed include a wider range of carbohydrate uptake in *M. hyorhinis*, which in turn may also explain why this species is a widely known contaminant in cell cultures. Moreover, the myo-inositol catabolism is exclusive to *M. hyopneumoniae* and may be an important trait for virulence. However, the most important difference seems to be related to glycerol conversion to dihydroxyacetone-phosphate, which produces toxic hydrogen peroxide. This activity, missing only in *M. flocculare*, may be directly involved in cytotoxicity, as already been described for two lung pathogenic mycoplasmas, namely *Mycoplasma pneumoniae* in human and *Mycoplasma mycoides* subsp. mycoides in ruminants. Metabolomic data suggest that even though these mycoplasmas are extremely similar in terms of their genome and metabolism, different products and reaction rates may be the result of differential expression in each of them. We were able to infer from the reconstructed networks that the lack of pathogenicity of *M. flocculare* if compared to the highly pathogenic *M. hyopneumoniae* may be related to its incapacity to produce cytotoxic hydrogen peroxide. Moreover, the ability of *M. hyorhinis* to grow in diverse sites and even in different hosts may be a reflection of its enhanced and wider carbohydrate uptake. Altogether, the metabolic differences highlighted in silico and in vitro provide important insights to the different levels of pathogenicity observed in each of the studied species.

These results were part of the PhD of Mariana Galvão Ferrarini, co-supervised by Arnaldo Zaha (Federal University of Rio Grande do Sul and Marie-France Sagot (Inria and LBBE, UMR5558). and defended on December 10th, 2015. These results have been submitted to a journal. The PhD manuscript will be made available in HAL in early 2016.

# 6.5. Cross-fertilising different computational approaches

**Tree matching**

We considered the following problem related to tree matching, that we called the Tree-Constrained Bipartite Matching problem. Given a bipartite graph $G = (V_1, V_2, E)$ with edge weights $w : E \rightarrow R^+$, a rooted tree $T_1$ on the set $V_1$ and a rooted tree $T_2$ on the set $V_2$, find a maximum weight matching $M$ in $G$, such that none of the matched nodes is an ancestor of another matched node in either of the trees [8]. This generalisation of the classical bipartite matching problem appears, for example, in the computational analysis of live cell video data. We showed that the problem is APX-hard and thus, unless P = NP, disproved a previous claim that it is solvable in polynomial time. Furthermore, we gave a 2-approximation algorithm based on a combination of the local ratio technique and a careful use of the structure of basic feasible solutions of a natural LP-relaxation, which we also show to have an integrality gap of $2 - o(1)$. We then considered a natural generalisation of the problem, where trees are replaced by partially ordered sets (posets). We showed that the local ratio technique gives a $2k\sigma$- approximation for the $k$-dimensional matching generalisation of the problem, in which the maximum number of incomparable elements below (or above) any given element in each poset is bounded by $\sigma$. We finally gave an almost matching integrality gap example, and an inapproximability result showing that the dependence on $\sigma$ is most likely unavoidable.

**Graph measures**

We proposed a new algorithm that computes the radius and the diameter of a weakly connected digraph $G = (V, E)$, by finding bounds through heuristics and improving them until they are validated [5]. Although the worst-case running time is $O(|V||E|)$, we experimentally showed that it performs much better in the case of real-world networks, finding the radius and diameter values after 10-100 BFSs instead of $|V|$ BFSs (independently of the value of |V|), and thus having running time $O(|E|)$ in practice. As far as we know, this is the first algorithm able to compute the diameter of weakly connected digraphs, apart from the naive algorithm, which runs in time $O(|V||E|)$ performing a BFS from each node. In the particular cases of strongly connected directed or connected undirected graphs, we compared our algorithm with known approaches by performing experiments on a dataset composed by several real-world networks of different kinds. These experiments showed that, despite its generality, the new algorithm outperforms all previous methods, both in the radius and in the diameter computation, both in the directed and in the undirected case, both in average running time and in robustness. Finally, as an application example, we used the new algorithm to determine the solvability over time of the "Six Degrees of Kevin Bacon" game, and of the "Six Degrees of Wikipedia" game. As a consequence, we computed for the first time the exact value of the radius and the diameter of the whole Wikipedia digraph.

The closeness and the betweenness centralities are two well-known measures of importance of a vertex within a given complex network. Having high closeness or betweenness centrality can have positive impact on the vertex itself: hence, we considered the problem of determining how much a vertex can increase its centrality by creating a limited amount of new edges incident to it [40]. We first proved that this problem does not admit a polynomial-time approximation scheme (unless P=NP), and we then proposed a simple greedy approximation algorithm (with an almost tight approximation ratio), whose performance is then tested on synthetic graphs and real-world networks.

The (Gromov) hyperbolicity is a topological property of a graph, which has been recently applied in several different contexts, such as the design of routing schemes, network security, computational biology, the analysis of graph algorithms, and the classification of complex networks. Computing the hyperbolicity of a graph can be very time consuming: indeed, the best available algorithm has running-time $O(n^{3.69})$, which is clearly prohibitive for big graphs. We provided a new and more efficient algorithm: although its worst-case complexity is $O(n^4)$, in practice it is much faster, allowing, for the first time, the computation of the hyperbolicity of graphs with up to 200,000 nodes [36]. We experimentally showed that the new algorithm drastically outperforms the best previously available algorithms, by analyzing a big dataset of real-world networks. Finally, we applied the new algorithm to compute the hyperbolicity of random graphs generated with the Erdös-Renyi model, the Chung-Lu model, and the Configuration Model.

**Hypergraph problems**

It had been previously proved independently and with different techniques that there exists an incremental output polynomial algorithm for the enumeration of the minimal edge dominating sets in graphs, *i.e.*, minimal dominating sets in line graphs. We provided the first polynomial delay and polynomial space algorithm for the problem [42]. We proposed a new technique to enlarge the applicability of Berge's algorithm that is based on skipping hard parts of the enumeration by introducing a new search strategy. The new search strategy is given by a strong use of the structure of line graphs.

We also studied some average properties of hypergraphs and the average complexity of algorithms applied to hypergraphs under different probabilistic models [14]. Our approach is both theoretical and experimental since our goal is to obtain a random model that is able to capture the real-data complexity. Starting from a model that generalizes the Erdös-Renyi model and we obtain asymptotic estimations on the average number of transversals, irredundants and minimal transversals in a random hypergraph. We use those results to obtain an upper bound on the average complexity of algorithms to generate the minimal transversals of a hypergraph. Then we make our random model more complex in order to bring it closer to real-data and identify cases where the average number of minimal transversals is at most polynomial, quasi-polynomial or exponential.

The hypergraph transversal problem has been intensively studied, both from a theoretical and a practical point of view. In particular, its incremental complexity is known to be quasi-polynomial in general and polynomial for bounded hypergraphs. Recent applications in computational biology however require to solve a generalisation of this problem, that we call bi-objective transversal problem. The instance is in this case composed of a pair of hypergraphs $(A, B)$, and the aim is to enumerate minimal sets which hit all the hyperedges of $A$ while intersecting a minimal set of hyperedges of $B$. We formalised this problem and related it to the enumeration of minimal hitting sets of bundles [32]. We showed cases when under degree or dimension contraints, these problems remain NP-hard, and gave a polynomial algorithm for the case when $A$ has bounded dimension, by building a hypergraph whose transversals are exactly the hitting sets of bundles.

<center>**IBIS Project-Team**</center>

# 6. New Results

## 6.1. Inference of bacterial regulatory networks from 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 ). In order to fully exploit this technology, we need methods to rapidly construct reporter genes, both on plasmids and on the chromosome, mathematical models to infer biologically relevant quantities from the primary data, and computer tools to achieve this in an efficient and user-friendly manner. For instance, in a typical microplate experiment, 96 cultures are followed in parallel, over several hours, resulting in 10,000-100,000 measurements of absorbance and fluorescence and luminescence intensities. Over the past few years, we put into place an experimental platform and data analysis software, notably the WELLREADER program (Section 5.4 ), to allow biologists to make the most out of the information contained in reporter gene expression data. An invited review on the analysis of fluorescent reporter gene data was published in the proceedings of the Third International Workshop on Hybrid Systems Biology (HSB 14) [25].

Valentin Zulkower, in the framework of his PhD thesis, has developed novel methods for the analysis of reporter gene data, based on the use of regularized linear inversion. This allows a range of estimation problems in the analysis of reporter gene data, notably the inference of growth rate, promoter activity, and protein concentration profiles, to be solved in a mathematically sound and practical manner. We have evaluated the validity of the approach using *in-silico* simulation studies, and observed that the methods are more robust and less biased than indirect approaches usually encountered in the experimental literature based on smoothing and subsequent processing of the primary data, like in WELLREADER. We have applied the methods to the analysis of fluorescent reporter gene data acquired in kinetic experiments with *Escherichia coli*. The methods were shown capable of reliably reconstructing time-course profiles of growth rate, promoter activity, and protein concentration from weak and noisy signals at low population volumes. Moreover, they captured critical features of those profiles, notably rapid changes in gene expression during growth transitions. The linear inversion methods have been implemented in the Python package WELLFARE, and integrated by Michel Page in the web application WELLINVERTER (Section 5.3 ). This work was presented at the major bioinformatics conference ISMB/ECCB 2015 and published in the special issue of *Bioinformatics* associated with the conference [24]. The Institut Français de Bioinformatique (IFB) accepted a proposal to extend WellInverter into a scalable and user-friendly web service providing a guaranteed quality of service, in terms of availability and response time. This web service will be deployed on the IFB platform and accompanied by extensive user documentation, online help, and a tutorial.

Over the years, the above tools have been used in several studies in IBIS directed at the experimental mapping of gene regulatory networks in *E. coli*. An example is the motility network of *E. coli*, studied by Diana Stefan in the context of her PhD thesis. The main thrust of this work lies in clarifying and solving methodological issues in the automated inference of quantitative models of gene regulatory networks from time-series gene expression data, also called reverse engineering in the bioinformatics literature. The application of existing reverse engineering methods is commonly based on implicit assumptions on the biological processes under study. First, the measurements of mRNA abundance obtained in transcriptomics experiments are taken to be representative of protein concentrations. Second, the observed changes in gene expression are assumed to be solely due to transcription factors and other specific regulators, while changes in the activity of the gene expression machinery and other global physiological effects are neglected. While convenient in practice, these assumptions are often not valid and bias the reverse engineering process. In her PhD thesis, Diana Stefan systematically investigated, using a combination of models and experiments, the importance of this bias and possible corrections. She measured with the help of fluorescent reporter genes the activity of genes involved in the FliA-FlgM module of the *E. coli* motility network. From these data, protein concentrations and global physiological effects were estimated by means of kinetic models of gene expression. The results indicate

that correcting for the bias of commonly-made assumptions improves the quality of the models inferred from the data. Moreover, it was shown by simulation that these improvements are expected to be even stronger for systems in which protein concentrations have longer half-lives and the activity of the gene expression machinery varies more strongly across conditions than in the FliA-FlgM module. The approach proposed in this study is broadly applicable when using time-series transcriptome data to learn about the structure and dynamics of regulatory networks. The paper describing the work was published in *PLoS Computational Biology* [23].

In addition to reporter gene data, a variety of other experimental data can be used for the mapping of gene regulatory networks. For example, using Chromatin Immunoprecipitation-sequencing (ChIP-seq) experiments, Stéphan Lacour and colleagues have identified a large number of target promoters of the sigma factor $\sigma^S$ during the transition from exponential to stationary phase. Sigma factors are accesory subunits of RNA polymerase, allowing the recognition of specific promoter sequences by the transcriptional machinery, and $\sigma^S$ is known to specifically accumulate in a variety of stress conditions. The study, published in *Scientific Reports* [21], has confirmed the importance of $\sigma^S$ for redirecting RNA polymerase to promoters that drive the expression of genes necessary for the survival of *E. coli* after nutrient exhaustion. Furthermore, the results highlight the role of $\sigma^S$ in the regulation of several noncoding RNAs.

## 6.2. Models of carbon metabolism in bacteria

All free-living bacteria have to adapt to a changing environment. Specific regulatory systems respond to particular stresses, but the most common decision bacteria have to make is the choice between alternative carbon sources, each sustaining a specific, maximal growth rate. Many bacteria have evolved a strategy that consists in utilizing carbon sources sequentially, in general favouring carbon sources that sustain a higher growth rate. As long as a preferred carbon source is present in sufficient amounts, the synthesis of enzymes necessary for the uptake and metabolism of less favourable carbon sources is repressed. This phenomenon is called Carbon Catabolite Repression (CCR) and the most salient manifestation of this regulatory choice is diauxic growth, a phenomenon discovered by Jacques Monod more than 70 years ago. Although this system is one of the paradigms of the regulation of gene expression in bacteria, the underlying mechanisms remain controversial. CCR involves the coordination of different subsystems of the cell - responsible for the uptake of carbon sources, their breakdown for the production of energy and precursors, and the conversion of the latter to biomass.

The complexity of this integrated system, with regulatory mechanisms cutting across metabolism, gene expression, signaling and subject to global physical and physiological constraints, has motivated important modeling efforts over the past four decades, especially in the enterobacterium Escherichia coli. Different hypotheses concerning the dynamic functioning of the system have been explored by a variety of modeling approaches. In an article in *Trends in Microbiology* [19], which was initiated during the sabbatical of Andreas Kremling in Grenoble in 2013, we have reviewed these studies and summarized their contributions to the quantitative understanding of CCR, focusing on diauxic growth in *E. coli*. Moreover, we have proposed a highly simplified representation of diauxic growth that makes it possible to bring out the salient features of the models proposed in the literature and confront and compare the explanations they provide. In parallel, specific aspects of CCR, in particular a better understanding of the role of the signalling molecule cyclic adenosine monophosphate (cAMP) in the dynamic regulation of promoters during growth transitions in *E. coli*, have been studied in the context of the PhD thesis of Valentin Zulkower, using both models and experimental data.

Beside CCR and the multiple regulatory systems controlling the metabolism of *E. coli*, the involvement of post-transcriptional regulation is uncertain. The post-transcriptional factor CsrA is stated as being the only regulator essential for the use of glycolytic substrates, but its impact on the functioning of central carbon metabolism has not been demonstrated. In the framework of the PhD thesis of Manon Morin, supported by a Contrat Jeune Scientifique INRA-Inria, the collaboration of Delphine Ropers, Muriel Cocaign-Bousquet and Brice Enjalbert from LISBP at INSA Toulouse has resulted in a multi-scale analysis of a wild-type strain and its isogenic mutant attenuated for CsrA. A variety of experimental data has been acquired for these two strains in relevant conditions, including growth parameters, gene expression levels, metabolite pools, enzyme

activities and metabolic fluxes. Data integration, metabolic flux analysis and regulation analysis revealed the pivotal role of post-transcriptional regulation for reshaping carbon metabolism. In particular, the work has shed light on *csrA* essentiality and has provided an explanation for the glucose-phosphate stress observed in the mutant strain. A paper summarizing the work has been submitted for publication in a microbiology journal.

## 6.3. 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. 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 intrinsic noise models from population snapshot data, on the one hand, and the inference of models focusing on cellular variability within isogenic populations from fluorescence microscopy gene expression profiles, on the other hand. Along with modelling and inference comes analysis of the inferred models in various respects, notably in terms of identifiability, single-cell state estimation and control. Other problems related with single-cell modelling and extracellular variability are considered in eukaryotic cells through external collaborations.

In the context of the response of yeast cells to osmotic shocks, in collaboration with the LIFEWARE project-team and colleagues from Université Paris Descartes and University of Pavia (Italy), Eugenio Cinquemani has investigated the use of mixed effects-modelling and identification techniques to characterize individual cell dynamics in isogenic cell populations. Mixed-effects models are hierarchical models where parametric response profiles of individuals is subject to inter-individual parameter variability following a common population distribution. Starting from identification approaches in pharmacokinetics, we have developed and applied inference methods to microfluidics data, with a focus on the response of budding yeast to osmotic shocks. First results presented at conference in 2013 and the identification and validation work performed with Andres Gonzales, who visited IBIS for a few months in 2014 during his PhD at the University of Pavia, have been finalized into a journal article recently accepted for publication in *PLoS Computational Biology* [20].

Started with a study of the arabinose uptake dynamics in *E. coli*, work on identification and state estimation for single-cell intrinsic noise models of gene networks has focused on the reconstruction of promoter activity profiles from fluorescent reporter data. In the single-cell stochastic context, given population snapshots of fluorescence levels at subsequent experimental instants, the problem becomes that of inferring promoter activity statistics over a cell population such as mean, variance or even higher-order moments from analogous statistics of the reporter output. This nontrivial extension of the deterministic deconvolution of promoter activity from population-average data requires knowledge of the stochastic reporter dynamics and of the relation between promoter and fluorescence statistical moments. In two conference papers, Eugenio Cinquemani investigated identifiability and identification of the kinetic parameters of the stochastic reporter dynamics [28] and proposed parametric and nonparametric methods for the reconstruction of the desired promoter activity statistics [27], [28], demonstrating their effectiveness *in silico*. Further developments of these methods and application to experimental data for addressing relevant biological questions will be the subject of future journal publications.

In parallel, collaboration of Eugenio Cinquemani with Marianna Rapsomaniki, post-doctoral researcher at at IBM Zurich Research Lab (Switzerland), Zoi Lygerou at the University of Patras (Greece) and John Lygeros at ETH Zurich (Switzerland) has been devoted to the analysis of data from Fluorescence Recovery After Photobleaching (FRAP) experiments and the inference of kinetic parameters of protein dynamics in single

eukaryotic cells. As an alternative to current approximate analytical methods, we have explored inference methods based on simulation of biological processes in realistic environments at a particle level. We introduced and demonstrated a new method for the inference of kinetic parameters of protein dynamics, where a limited number of *in-silico* FRAP experiments is used to construct a mapping from FRAP recovery curves to the parameters sought. Parameter estimates from experimental data are then computed by applying the mapping to the observed recovery curves, at virtually no additional price for any number of experiments, along with the application of a bootstrap procedure for determining identifiability of the parameters and confidence intervals for their estimates. After validation on synthetic data, the method was successfully applied to the analysis of the nuclear proteins Cdt1, PCNA and GFPnls in mammalian cells, also sheding light on cell-to-cell variability of the protein kinetics. Method and results have been published in *Bioinformatics* this year [22].

## 6.4. 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. The objective of the RESET project, supported in the framework of the Programme d'Investissements d'Avenir (Section 8.2 ), is to develop novel strategies to limit or completely stop microbial growth and to explore biotechnological applications of these approaches.

A foundation result for growth control in bacteria was published in the journal *Molecular Systems Biology* this year [18]. In this publication, which is based on the PhD thesis of Jérôme Izard and post-doctoral work of Cindy Gomez Balderas, we describe an engineered *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. We have shown that our synthetic growth switch functions in a medium-independent and reversible way, and we have provided evidence that the switching phenotype arises from the ultrasensitive response of the growth rate to the concentration of RNA polymerase. In parallel, Delphine Ropers in collaboration with Jean-Luc Gouzé and Stefano Casagrande of the BIOCORE team are developing a quantitative model of the gene expression machinery to account for this surprising observation.

The publication in *Molecular Systems Biology* also presents a biotechnological application of the growth switch in which both the wild-type *E. coli* strain and our modified strain are 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. Remarkably, without any further optimization, the improved yield is close to the theoretical maximum computed from a flux balance model of *E. coli* metabolism. The synthetic growth switch is thus a promising tool for gaining a better understanding of bacterial physiology and for applications in synthetic biology and biotechnology. We submitted a patent for such applications at the European Patent Office.

Whereas the synthetic growth switch has been designed for biotechnological purposes, the question can be asked how resource allocation is organized in wild-type strains that have naturally evolved. Recent work has shown that coarse-grained models of resource allocation can account for a number of empirical regularities relating the the macromolecular composition of the cell to the growth rate. Some of these models hypothesize control strategies enabling microorganisms to optimize growth. While these studies focus on steady-state growth, such conditions are rarely found in natural habitats, where microorganisms are continually challenged by environmental fluctuations. The aim of the PhD thesis of Nils Giordano is to extend the study of microbial growth strategies to dynamical environments, using a self-replicator model. In a recently submitted paper, we have formulated dynamical growth maximization as an optimal control problem that can be solved using Pontryagin's Maximum Principle. We compare this theoretical gold standard with different possible

implementations of growth control in bacterial cells. This study has been carried out in collaboration with Jean-Luc Gouzé and Francis Mairet of the BIOCORE project-team.

# NUMED Project-Team  (section vide)

<p style="text-align:center"><span style="color:red">**STEEP Project-Team**</span></p>

# 6. New Results

## 6.1. Methods for the calibration of LUTI models

The setting up of a LUTI model requires, like most numerical models, at least one phase of parameter estimation. This is concisely referred to here as calibration, although the calibration of a LUTI model also entails other aspects such as the definition of spatial zones, of economic sectors, etc. The TRANUS LUTI model plus software, like many other existing models, come along with a relatively simple calibration methodology. Most LUTI models indeed perform parameter estimation in a piecewise fashion, by sequentially estimating subsets of parameters. While this reduces the mathematical and computational complexity of calibration, neglecting the interactions across different modules and their parameters, may result in a significant loss of a model?s quality. A second issue is that TRANUS, like several other LUTI softwares, employs rudimentary numerical routines for parameter estimation. We aim at reducing these weaknesses.

In 2014, we had obtained first results along these lines: parameter estimation of the so-called shadow prices (specific parameters of the TRANUS model) was posed as optimization problem and several solution procedures were developed which were based on "unwinding" the dynamics of the model, making the problem amenable to standard numerical optimisation techniques.

The work continued throughout 2015, along different directions. First, the calibration was extended to handle several different parameter types simultaneously (shadow prices as well as the so-called substitution parameters, which are notoriously difficult to estimate) [7]. Such a simultaneous estimation of different parameter sets seems to be rare in LUTI practice.

Second, we proposed a methodology for assessing properties (convergence, accuracy) of our (and other) LUTI calibration methods [6]. This consists in generating synthetic data, starting from a model calibrated on observed data, such that the synthetic data are completely consistent, i.e. there are a set of model parameters that exactly reproduce these data (which is not the case with the observed data). The ground truth model parameters are then easily used to assess calibration parameters. Such a methodology, akin to twin experiments in data assimilation, seems to be novel for LUTI research.

Third, LUTI models are usually calibrated on a base year or period, and used in a prospective manner (via simulated "predictions" for future periods). As with any numerical model, it is wise to make sure that a calibrated LUTI model does not overfit the observations used for calibration; otherwise, its "predictions" may be grossly erroneous. Potential overfitting does not seem to have been deeply studied in the LUTI literature. We have made an initial investigation by calibrating different versions of a TRANUS model, varying the number of shadow prices used as parameters in the model (there is, by default, one shadow price per combination of geographical zone of the study area and economic sector) [6]. For instance, after an initial calibration using all shadow prices, we then dropped the two third smallest of them and re-calibrated the model using the remaining third. The goodness-of-fit to observations was worse by only 3%. In line with well-known principles of model selection (Occam's razor), this may suggest that it is preferable to use the model with fewer parameters when doing predictions. This is still work in progress; showing its relevance is planned to be studied by a similar methodology as above, using simulated twin experiments.

This work is done in collaboration with Arthur Vidard from the AIRSEA Inria project-team and Brian Morton from the University of North Carolina at Chapel Hill.

## 6.2. Estimation of Sobol' indices combining nested designs and replication method

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 becomes independent of the input space dimension. The generalization to closed second-order Sobol' indices relies on the replication of randomized orthogonal arrays. However the replication method still requires a large number of model evaluations. By rendering this method iterative, the required number of evaluations can be controlled. The estimation procedure is therefore stopped when the convergence of estimates is considered reached. The key feature of this approach is the construction of nested designs. For the estimation of first-order indices, we exploit a nested Latin Hypercube already introduced in the litterature. For the estimation of closed second-order indices, two methods are proposed to construct a nested orthogonal array. One of the two leads to a partition of the coordinate space over a Galois field.

This work has been done in collaboration with Laurent Gilquin and Clementine Prieur (members of Moise Team), and belongs to the work program of CiTIES project. The proposed procedure will be soon applied to study the sensitivity of TRANUS model.

## 6.3. Environmental pressures associated with material flows

This work is the follow-up of a previous study dedicated to material flow analysis of the French cereal supply chain at various spatial levels [12]. The goal was twofold:

- trace the flows to their initial geographic origin or final destination,
- couple material flows with a series of environmental pressures associated to them.

For the first goal, we used an Absorbing Markov Chains model where transient states represent raw or semi-products and absorbing states correspond to final consumption products. For the second goal, we used pressure ratios for environmental pressures most relevant to cereals, namely energy use, GHG emissions, land use, use of pesticides and blue water footprint. The model is based on physical supply and use tables and distinguishes between 21 industries, 22 products, 38 regions of various spatial resolution (22 French regions, 10 countries, 6 continents) and 4 modes of transport. Illustrative examples were taken in order to demonstrate the versatility of the results produced, for instance: What is the fate/supply area of a region's production/consumption? What are the production and consumption footprint of a region? These results are designed to be a first step towards scenario analysis for decision-aiding that would also include socioeconomic indicators [13]

## 6.4. Material flows of the French forest-wood supply chain

The methodology developed in Courtonne et al. [12] on the case of the cereal supply chain was adapted to the French forest-wood supply chain in collaboration with the Laboratoire d'Economie Forestière. Supply chain flows were estimated both at the national and regional scale for wood harvest, addition to stock, production, imports and exports of construction wood, industrial wood and energy wood. These results can be a basis to analyze potential value losses throughout the supply chain, for instance exports of raw materials instead of local transformation. They can also be use to study the competitive use of wood for energy, industry and construction/furnitures, which is a question of growing importance in the context of energy transition.

## 6.5. Land Use/Land Cover Change (LUCC) Modelling and Ecosytem Services

The ESNET project (EcoSystem services NETworks) is a collaboration lead by LECA (Laboratoire d'ECologie Alpine, UJF) that aims at characterizing the ecosystem services of the Grenoble urban region (about 2/3 of the Isere département) at the 2040 horizon under various constraints of urban policy planning, changes in agricultural and forest management, and climate change impact on ecosystems.

The cartographic effort of the project has been hosted at Inria, and has produced in 2014 three very detailed maps of land use and land cover at the 15m resolution over the whole study area, in 1998, 2003 and 2009, respectively. An extensive analysis of the patterns of landscape change has been performed from these data, with special emphasis on urban sprawl and the associated loss of arable land. This work has been submitted for publication very recently.

A second related piece of work has been produced, both from this cartographic source and more specific remote sensing data. The objective was to characterize in detail the cultural successions and patterns of the study area, in order to produce fine scale maps of associated ecosystem services. This work has just been submitted for publication at the time of writing.

Finally, the scenarios of future land use and land cover that have been elaborated for this project have all been projected at the 2040 horizon at the 15m scale with a well-known LUCC modelling environment (Dinamica) for urban changes, and from in-project models for the other types of land use and cover. A third article bearing on these scenarios and their LUCC modelling is in preparation.

As an aside of this land use/cover modelling effort, the STEEP team has been involved in two of the most detailed ecosystem service models developed for the project: one for the analysis of crop production and associated nitrogen cycle assessment — with the final aim to constrain both the production services and water quality issue related to nitrogen loading — and one on "recreational services". Our involvement in these models was directly related to the acquired expertise in land use modelling.

In the process of this modelling exercise, the STEEP team has acquired an in-depth knowledge and expertise of LUCC models. As a consequence, various theoretical flaws have been identified in the theoretical foundations of such models. An important by-product of the ESNET project is therefore a series of articles in preparation in the team, whose aim is to address and correct these flaws in a very general way; it is hoped that LUCC theory will be put on a more serious theoretical footing as a result of this methodological work, which should be submitted for publication in 2016 for the most part. Another but more limited methodological contribution bears on the development of error models for landscape metrics, another important methodological blind-spot in the specialized literature.

## 6.6. A benchmarking tool to assess the compatibility of the INDCs with the 2°C long-term target

Climate negotiations related to global warming are another important issue of sustainable development. In this framework that is place at international scale we have developed a benchmarking tool which allows to assess the compatibility of the Intended Nationally-Determined Contributions (INDCs) given by all states for the Conference COP21, with the 2°C long-term target. This benchmarking tool has been designed via an adaptation of REDEM model and algorithm we developed in 2014 with EDDEN laboratory. This tool has been used by the "*Groupe Interdisciplinaire sur les Contributions Nationales*" (GICN) which has been mandated by french ministry of Sustainable Development to prepare the climate change conference COP21 at Paris.

<h1 style="text-align:center; color:red">AVALON Project-Team</h1>

# 7. New Results

## 7.1. Energy Efficiency of Large Scale Distributed Systems

**Participants:** Laurent Lefèvre, Daniel Balouek Thomert, Eddy Caron, Radu Carpa, Marcos Dias de Assunçao, Jean-Patrick Gelas, Olivier Glück, Jean-Christophe Mignot, Violaine Villebonnet.

### 7.1.1. Energy efficient Core Networks

This work [8], [43] seeks to improve the energy efficiency of backbone networks by providing an intra-domain Software Defined Network (SDN) approach to selectively turn off a subset of links. To do this, we designed an energy-aware traffic engineering technique for reducing energy consumption in backbone networks. Energy-efficient traffic engineering was analysed in previous work, but none addressed implementation challenges of their solutions. We showed that ignoring to test the feasibility of techniques can lead to bad estimations and unstable solutions. We proposed the STREETE framework (SegmenT Routing based Energy Efficient Traffic Engineering) that represents an online method to switch some links off/on dynamically according to the network load. We have implemented a working prototype in the OMNET++ simulator. Networks are progressively using centralised architecture, and SDN is increasingly utilised in data centre networks. We believe that SDN may be extended to backbone networks. The implemented solution shows that SDN may also be a good means for reducing the energy consumption of network devices. Compared to previous work, in this work we used the SPRING protocol to improve the stability of energy-efficient traffic engineering solutions. To the best of our knowledge, this is the first work proposing the use of SPRING to improve the energy efficiency of backbone networks. The flexibility of this routing protocol is well suited to frequent route changes that happen when we switch links off and on. Moreover, this protocol can be easily applied to SDN solutions. Using simulations, we showed that as much as 44% of links can be switched off to save energy in real backbone networks. Even greedy techniques can easily approach the maximum reduction in the amount of energy consumed. In fact, the bottleneck in terms of energy efficiency in energy-aware traffic engineering is the connectivity constraint. We performed a stress test of our solution under rapidly increasing traffic patterns and showed that more work must be done in the domain of switching links back on: a field which has received little attention from the research community.

### 7.1.2. Energy proportionality in HPC systems

Energy savings are among the most important topics concerning Cloud and HPC infrastructures nowadays. Servers consume a large amount of energy, even when their computing power is not fully utilized. These static costs represent quite a concern, mostly because many datacenter managers are over-provisioning their infrastructures compared to the actual needs. This results in a high part of wasted power consumption. In this work [19], [47] , we proposed the BML ("Big, Medium, Little") infrastructure, composed of heterogeneous architectures, and a scheduling framework dealing with energy proportionality. We introduce heterogeneous power processors inside datacenters as a way to reduce energy consumption when processing variable workloads. Our framework brings an intelligent utilization of the infrastructure by dynamically executing applications on the architecture that suits their needs, while minimizing energy consumption. Our first validation process focuses on distributed stateless web servers scenario and we analyze the energy savings achieved through energy proportionality. This research activity is performed with the collaboration of Sepia Team (IRIT, Toulouse) through the co-advising of Violaine Villebonnet.

### 7.1.3. Energy-Aware Server Provisioning

Several approaches to reduce the power consump- tion of datacenters have been described in the literature, most of which aim to improve energy efficiency by trading off performance for reducing power consumption. However, these approaches do not always provide means for administrators and users to specify how they want to explore such trade-offs. This work [27] provides techniques for assigning jobs to distributed resources, exploring energy efficient resource provisioning. We use middleware-level mechanisms to adapt resource allocation according to energy-related events and user-defined rules. A proposed framework enables developers, users and system ad- ministrators to specify and explore energy efficiency and perfor- mance trade-offs without detailed knowledge of the underlying hardware platform. Evaluation of the proposed solution under three scheduling policies shows gains of 25% in energy-efficiency with minimal impact on the overall application performance. We also evaluate reactivity in the adaptive resource provisioning This approach has been applied in the Nuage research project [26].

### 7.1.4. Virtual Home Gateway

About 80-90% of the energy in today's wireline networks is consumed in the access network, , including about 10 to 30W per user being dissipated mostly by the customer premises equipment (CPE). Home gateway is a popular equipment deployed at the end of networks and supporting a set of heterogeneous services (data, phone, television, multimedia, security services). These gateways and associated services can be difficult to deploy and maintain for customers. These gateways are difficult to manage for network operators and consume a lot of energy. We explore the technical solutions to reduce the complexity and energy impact of such equipments by moving services to some external dedicated and shared facilities of network operators. This result is a joint work between Avalon team (J.P. Gelas, L. Lefevre) and Addis Abeba University (M. Tsibie and T. Assefa). This research has been demonstrated in the GreenTouch final celebration event in New York (June 2015).

## 7.2. MPI Application and Storage System Simulation

**Participants:** Frédéric Suter, Laurent Pouilloux.

### 7.2.1. Scalable Off-line Simulation of MPI Applications

Analyzing and understanding the performance behavior of parallel applications on parallel computing plat-forms is a long-standing concern in the High Performance Computing community. When the targeted plat-forms are not available, simulation is a reasonable approach to obtain objective performance indicators and explore various hypothetical scenarios. In the context of applications implemented with the Message Passing Interface, two simulation methods have been proposed, on-line simulation and off-line simulation, both with their own drawbacks and advantages.

We proposed in [9] an off-line simulation framework, i.e., one that simulates the execution of an application based on event traces obtained from an actual execution. The main novelty of this work, when compared to previously proposed off-line simulators, is that traces that drive the simulation can be acquired on large, distributed, heterogeneous, and non-dedicated platforms. As a result the scalability of trace acquisition is increased, which is achieved by enforcing that traces contain no time-related information. Moreover, our framework is based on an state-of-the-art scalable, fast, and validated simulation kernel.

Such off-line analysis faces scalability issues for acquiring, storing, or replaying large event traces. Then, in [10], we combined our framework with another, specialized in the production of compact traces, to capitalize on their respective strengths while alleviating several of their limitations. We showed that the combined framework affords levels of scalability that are beyond that achievable by either one of the two individual frameworks.

### 7.2.2. Simulation of Storage Elements

Storage is a essential component of distributed computing infrastructures, *i.e.,* clusters, grids, clouds, data centers, or supercomputers, to cope with the tremendous increase in scientific data production and the ever-growing need for data analysis and preservation. Understanding the performance of a storage subsystem or dimensioning it properly is an important concern for which simulation can help by allowing for fast, fully repeatable, and configurable experiments for arbitrary hypothetical scenarios. However, most simulation frameworks tailored for the study of distributed systems offer no or little abstractions or models of storage resources.

In [34], we detailed the extension of SimGrid with storage simulation capacities. We first defined the required abstractions and propose a new API to handle storage components and their contents in SimGrid-based simulators. Then we characterized the performance of the fundamental storage component that are disks and derive models of these resources. Finally we listed several concrete use cases of storage simulations in clusters, grids, clouds, and data centers for which the proposed extension would be beneficial.

## 7.3. MapReduce Computations on Hybrid Distributed Computations Infrastructures

**Participants:** Gilles Fedak, Julio Anjos, Anthony Simonet.

In this section we report on our efforts to provide MapReduce Computing environments on Hybrid infrastructures, i.e composed of Desktop Grids and Cloud computing environments.

Cloud computing has increasingly been used as a platform for running large business and data processing applications. Although clouds have become extremely popular, when it comes to data processing, their use incurs high costs. Conversely, Desktop Grids, have been used in a wide range of projects, and are able to take advantage of the large number of resources provided by volunteers, free of charge. Merging cloud computing and desktop grids into a hybrid infrastructure can provide a feasible low-cost solution for big data analysis. Although frameworks like MapReduce have been devised to exploit commodity hardware, their use in a hybrid infrastructure raise some challenges due to their large resource heterogeneity and high churn rate.

### 7.3.1. BIGhybrid - A Toolkit for Simulating MapReduce in Hybrid Infrastructures

In [20], we introduced BIGhybrid, a toolkit that is used to simulate MapReduce in hybrid environments. Its main goal is to provide a framework for developers and system designers that can enable them to address the issues of Hybrid MapReduce. In this paper, we described the framework which simulates the assembly of two existing middleware: BitDew- MapReduce for Desktop Grids and Hadoop-BlobSeer for Cloud Computing. The experimental results that are included in this work demonstrate the feasibility of our approach.

### 7.3.2. HybridMR: a New Approach for Hybrid MapReduce Combining Desktop Grid and Cloud Infrastructures

In [18], we proposed a novel MapReduce computation model in hybrid computing environment called HybridMR. Using this model, high performance cluster nodes and heterogeneous desktop PCs in Internet or Intranet can be integrated to form a hybrid computing environment. In this way, the computation and storage capability of large-scale desktop PCs can be fully utilized to process large-scale datasets. HybridMR relies on a hybrid distributed file system called HybridDFS, and a time-out method has been used in HybridDFS to prevent volatility of desktop PCs, and file replication mechanism is used to realize reliable storage. A new node priority-based fair scheduling (NPBFS) algorithm has been developed in HybridMR to achieve both data storage balance and job assignment balance by assigning each node a priority through quantifying CPU speed, memory size and I/O bandwidth. Performance evaluation results showed that the proposed hybrid computation model not only achieves reliable MapReduce computation, reduces task response time and improves the performance of MapReduce, but also reduces the computation cost and achieves a greener computing mode.

### 7.3.3. $D^3$ -MapReduce: Towards MapReduce for Distributed and Dynamic Data Sets

So far MapReduce has been mostly designed for batch processing of bulk data. The ambition of $D^3$-MapReduce, presented in [32], is to extend the MapReduce programming model and propose efficient implementation of this model to: i) cope with distributed data sets, i.e. that span over multiple distributed infrastructures or stored on network of loosely connected devices; ii) cope with dynamic data sets, i.e. which dynamically change over time or can be either incomplete or partially available. In this paper, we draw the path towards this ambitious goal. Our approach leverages Data Life Cycle as a key concept to provide MapReduce for distributed and dynamic data sets on heterogeneous and distributed infrastructures. We first reported on our attempts at implementing the MapReduce programming model for Hybrid Distributed Computing Infrastructures (Hybrid DCIs). We present the architecture of the prototype based on BitDew, a middleware for large scale data management, and Active Data, a programming model for data life cycle management. Second, we outlined the challenges in term of methodology and present our approaches based on simulation and emulation on the Grid'5000 experimental testbed. We conducted performance evaluations and compare our prototype with Hadoop, the industry reference MapReduce implementation. We presented our work in progress on dynamic data sets that has lead us to implement an incremental MapReduce framework. Finally, we discussed our achievements and outline the challenges that remain to be addressed before obtaining a complete D 3-MapReduce environment.

### 7.3.4. Availability and Network-Aware MapReduce Task Scheduling over the Internet.

MapReduce offers an ease-of-use programming paradigm for processing large datasets. In our previous work, we have designed a MapReduce framework called BitDew-MapReduce for desktop grid and volunteer computing environment, that allows nonexpert users to run data-intensive MapReduce jobs on top of volunteer resources over the Internet. However, network distance and resource availability have great impact on MapReduce applications running over the Internet. To address this, an availability and network-aware MapReduce framework over the Internet is proposed in [38]. Simulation results show that the MapReduce job response time could be decreased by 27.15%, thanks to Naive Bayes Classifier-based availability prediction and landmark-based network estimation.

## 7.4. Managing Big Data Life Cycle

**Participants:**  Gilles Fedak, Anthony Simonet.

### 7.4.1. Active Data - Enabling Smart Data Life Cycle Management for Large Distributed Scientific Data Sets

The Big Data challenge consists in managing, storing, analyzing and visualizing these huge and ever growing data sets to extract sense and knowledge. As the volume of data grows exponentially, the management of these data becomes more complex in proportion. A key point is to handle the complexity of the data life cycle, i.e. the various operations performed on data: transfer, archiving, replication, deletion, etc. Indeed, data-intensive applications span over a large variety of devices and e-infrastructures which implies that many systems are involved in data management and processing. In [17], we proposed Active Data, a programming model to automate and improve the expressiveness of data management applications. We first define the concept of data life cycle and introduce a formal model that allows to expose data life cycle across heterogeneous systems and infrastructures. The Active Data programming model allows code execution at each stage of the data life cycle: routines provided by programmers are executed when a set of events (creation, replication, transfer, deletion) happen to any data. We implement and evaluate the model with four use cases: a storage cache to Amazon-S3, a cooperative sensor network, an incremental implementation of the MapReduce programming model and automated data provenance tracking across heterogeneous systems. Altogether, these scenarios illustrate the adequateness of the model to program applications that manage distributed and dynamic data sets. We also show that applications that do not leverage on data life cycle can still benefit from Active Data to improve their performances.

### 7.4.2. *Using Active Data to Provide Smart Data Surveillance to E-Science Users*

Modern scientific experiments often involve multiple storage and computing platforms, software tools, and analysis scripts. The resulting heterogeneous environments make data management operations challenging, the significant number of events and the absence of data integration makes it difficult to track data provenance, manage sophisticated analysis processes, and recover from unexpected situations. Current approaches often require costly human intervention and are inherently error prone. The difficulties inherent in managing and manipulating such large and highly distributed datasets also limits automated sharing and collaboration. In [37], we study a real world e-Science application involving terabytes of data, using three different analysis and storage platforms, and a number of applications and analysis processes. We demonstrate that using a specialized data life cycle and programming model, Active Data, we can easily implement global progress monitoring, and sharing, recover from unexpected events, and automate a range of tasks.

### 7.4.3. *SMART: An Application Framework for Real Time Big Data Analysis on Heterogeneous Cloud Environments.*

The amount of data that human activities generate poses a challenge to current computer systems. Big data processing techniques are evolving to address this challenge, with analysis increasingly being performed using cloud-based systems. Emerging services, however, require additional enhancements in order to ensure their applicability to highly dynamic and heterogeneous environments and facilitate their use by Small & Medium-sized Enterprises (SMEs). Observing this landscape in emerging computing system development, this work presents Small & Medium-sized Enterprise Data Analytic in Real Time (SMART) for addressing some of the issues in providing compute service solutions for SMEs. SMART offers a framework for efficient development of Big Data analysis services suitable to small and medium-sized organizations, considering very heterogeneous data sources, from wireless sensor networks to data warehouses, focusing on service composability for a number of domains. In [62], we presented the basis of this proposal and preliminary results on exploring application deployment on hybrid infrastructure.

## 7.5. Desktop Grid Computing

**Participants:** Gilles Fedak, Anthony Simonet.

### 7.5.1. *Multi-Criteria and Satisfaction Oriented Scheduling for Hybrid Distributed Computing Infrastructures*

Assembling and simultaneously using different types of distributed computing infrastructures (DCI) like Grids and Clouds is an increasingly common situation. Because infrastructures are characterized by different attributes such as price, performance, trust, greenness, the task scheduling problem becomes more complex and challenging. In [15], we presented the design for a fault-tolerant and trust-aware scheduler, which allows to execute Bag-of-Tasks applications on elastic and hybrid DCI, following user-defined scheduling strategies. Our approach, named Promethee scheduler, combines a pull-based scheduler with multi-criteria Promethee decision making algorithm. Because multi-criteria scheduling leads to the multiplication of the possible scheduling strategies, we proposed SOFT, a methodology that allows to find the optimal scheduling strategies given a set of application requirements. The validation of this method is performed with a simulator that fully implements the Promethee scheduler and recreates an hybrid DCI environment including Internet Desktop Grid, Cloud and Best Effort Grid based on real failure traces. A set of experiments shows that the Promethee scheduler is able to maximize user satisfaction expressed accordingly to three distinct criteria: price, expected completion time and trust, while maximizing the infrastructure useful employment from the resources owner point of view. Finally, we present an optimization which bounds the computation time of the Promethee algorithm, making realistic the possible integration of the scheduler to a wide range of resource management software.

### 7.5.2. *Synergy of Volunteer Measurements and Volunteer Computing for Effective Data Collecting, Processing, Simulating and Analyzing on a Worldwide Scale*

The paper [31] concerns the hype idea of Citizen Science and the related paradigm shift: to go from the passive "volunteer computing" to other volunteer actions like "volunteer measurements" under guidance of scientists. They can be carried out by ordinary people with standard computing gadgets (smartphone, tablet, etc.) and the various standard sensors in them. Here the special attention is paid to the system of volunteer scientific measurements to study air showers caused by cosmic rays. The technical implementation is based on integration of data about registered night flashes (by radiometric software) in shielded camera chip, synchronized time and GPS-data in ordinary gadgets: to identify night air showers of elementary particles; to analyze the frequency and to map the distribution of air showers in the densely populated cities. The project currently includes the students of the National Technical University of Ukraine KPI, which are compactly located in Kyiv city and contribute their volunteer measurements. The technology would be very effective for other applications also, especially if it will be automated (e.g., on the basis of XtremWeb or/and BOINC technologies for distributed computing) and used in some small area with many volunteers, e.g. in local communities (Corporative/Community Crowd Computing).

### 7.5.3. *Towards an Environment for doing Data Science that runs in Browsers*

In [25], we proposed a path for doing Data Science using browsers as computing and data nodes. This novel idea is motivated by the cross-fertilized fields of desktop grid computing, data management in grids and clouds, Web technologies such as Nosql tools, models of interactions and programming models in grids, cloud and Web technologies. We propose a methodology for the modeling, analyzing, implemention and simulation of a prototype able to run a MapReduce job in browsers. This work allows to better understand how to envision the big picture of Data Science in the context of the Javascript language for programming the middleware, the interactions between components and browsers as the operating system. We explain what types of applications may be impacted by this novel approach and, from a general point of view how a formal modeling of the interactions serves as a general guidelines for the implementation. Formal modeling in our methodology is a necessary condition but it is not sufficient. We also make round-trips between the modeling and the Javascript or used tools to enrich the interaction model that is the key point, or to put more details into the implementation. It is the first time to the best of our knowledge that Data Science is operating in the context of browsers that exchange codes and data for solving computational and data intensive programs. Computational and data intensive terms should be understand according to the context of applications that we think to be suitable for our system.

### 7.5.4. *E-Fast & CloudPower: Towards High Performance Technical Analysis for Small Investors.*

About 80% of the financial market investors fail, the main reason for this being their poor investment decisions. Without advanced financial analysis tools and the knowledge to interpret the analysis, the investors can easily make irrational investment decisions. Moreover, in- vestors are challenged by the dynamism of the market and a relatively large number of indicators that must be computed. In this paper we propose E-Fast, an innovative approach for on-line technical analysis for helping small investors to obtain a greater efficiency on the market by increasing their knowledge. The E-Fast technical analysis platform proto- type relies on High Performance Computing (HPC), allowing to rapidly develop and extensively validate the most sophisticated finance analysis algorithms. In [36], we aim at demonstrating that the E-Fast im- plementation, based on the CloudPower HPC infrastructure, is able to provide small investors a realistic, low-cost and secure service that would otherwise be available only to the large financial institutions. We describe the architecture of our system and provide design insights. We present the results obtained with a real service implementation based on the Exponential Moving Average computational method, using CloudPower and Grid5000 for the computations' acceleration. We also elaborate a set of interesting challenges emerging from this work, as next steps towards high performance technical analysis for small investors.

## 7.6. HPC Component Model

**Participants:** Hélène Coullon, Vincent Lanore, Christian Perez, Jérôme Richard.

### 7.6.1. 3D FFT and $L^2C$

We have completed the work started in 2014. To harness the computing power of supercomputers, HPC application algorithms have to be adapted to the underlying hardware. This is a costly and complex process which requires handling many algorithm variants. In [23], we studied the ability of the component model $L^2C$ to express and handle the variability of HPC applications. The goal is to ease application adaptation. Analysis and experiments are done on a 3D-FFT use case. Results show that $L^2C$, and components in general, offer a generic and simple handling of 3D-FFT variants while obtaining performance close to well-known libraries

### 7.6.2. Multi-Stencil DSL in $L^2C$

As high performance architectures evolve continuously to be more powerful, such architectures also usually become more difficult to use efficiently. As a scientist is not a low level and high performance programming expert, Domain Specific Languages (DSLs) are a promising solution to automatically and efficiently write high performance codes. However, if DSLs ease programming for scientists, maintainability and portability issues are transferred from scientists to DSL de- signers. This work [44] has dealt with an approach to improve maintainability and programming productivity of DSLs through the generation of a component-based parallel runtime. To study it, we have designed a DSL for multi-stencil programs, that is evaluated on a real-case of shallow water equations implemented with $L^2C$.

### 7.6.3. Reconfigurable HPC component model

High-performance applications whose structure changes dynamically during execution are extremely complex to develop, maintain and adapt to new hardware. Such applications would greatly benefit from easy reuse and separation of concerns which are typical advantages of component models. Unfortunately, no existing component model is both HPC-ready (in terms of scalability and overhead) and able to easily handle dynamic reconfiguration. In [33], we aimed at addressing performance, scalability and programmability by separating locking and synchronization concerns from reconfiguration code. To this end, we propose directMOD, a component model which provides on one hand a flexible mechanism to lock subassemblies with a very small overhead and high scalability, and on the other hand a set of well-defined mechanisms to easily plug various independently-written reconfiguration components to lockable subassemblies. We evaluate both the model itself and a C++/MPI implementation called directL2C .

### 7.6.4. Towards a Task-Component Model

In [24], we propose a first model that aims at combining both component models and task based models such as StarPU. Component models bring many good software enginering properties such as code re-use while task based models seems to be very efficient to exploit recent hardware such as SMP, manycore, or GPGPUs. This work evaluates a proof-of-concepts only considering SMP nodes.

## 7.7. Security for Virtualization and Clouds

**Participants:** Eddy Caron, Arnaud Lefray.

### 7.7.1. Security and placement

We have proposed a solution for placement-based security and client-centric security. Even with perfect information flow control mechanisms, virtualized environments are still sensitive to silent information leakage, that is covert channels, due to shared hardware ressources. We have proposed a fine-grained placement based on the client's security properties to tackle this issue. The client submits an application i.e., a graph of VMs, and information flow rules defining the acceptable risk. Due to the lack of usable covert channel metric to qualify an acceptable risk, we have proposed a new information leakage metric. As covert channels exploit microarchitecture flaws, we have integrated the specificity of NUMA allocation schemes in our placement algorithm.

### *7.7.2. Security and logic language*

Besides, the main issue with existing security languages is the ability to formally guarantee the required property. On the one hand, security policies described in a natural language have quite ambiguous semantics. On the other hand, a formal language or logic provides clear syntax and semantics. Moreover, existing mechanisms are dedicated to secure specific type of entities (e.g., VM, Service, Data, VNet). Therefore, the problem is to have a formal definition of security properties and proven procedures to transform the end-user's global security properties into multiple local properties enforceable by several local mechanisms. For these reasons, we proposed a logic language called IF-PLTL (Information Flow Past Linear Time Logic). Our logic is dedicated to controlling the propagation of information i.e., direct and indirect information flows. As these information flows cannot be obtained directly, we have explained their construction from low-level observable events. Security decisions are naturally expressed according to past actions. Accordingly, IF-PLTL is based on the past fragment of LTL. In addition to using IF-PLTL to transform properties, we have proposed a dynamic monitor that can enforce the full expressivity of IF-PLTL even if its complexity (in time and space) would incur a high overhead in practice.

## 7.8. Autonomic Middleware Deployment using Self-Stabilization

**Participants:**  Eddy Caron, Maurice Faye.

Dynamic nature of distributed architecture is a major challenge to avail the benefits of distributed computing. An effective solution to deal with this dynamic nature is to implement a self-adaptive mechanism to sustain the distributed architecture. Self-adaptive systems can autonomously modify their behavior at run-time in response to changes in their environment. This capability may be included in the software systems at design time or later by external mechanisms. We have created a self-adaptive algorithm for the DIET middleware. Once the middleware is deployed, it can detect a set of events which indicate an unstable deployment state. When an event is detected, some instructions are executed to handle the event. We have designed a simulator to have a deeper insights of our proposed self-adaptive algorithm.

# CTRL-A Team

# 7. New Results

## 7.1. Discrete control and reactive language support

**Participants:** Gwenaël Delaval, Eric Rutten, Stéphane Mocanu, Alia Hajjar, Abdoul-Razak Hassimi Harouna.

Concerning language support, we have designed and implemented BZR, a mixed imperative/declarative programming language: declarative contracts are enforced upon imperatively described behaviors (see 6.1 ). The semantics of the language uses the notion of Discrete Controller Synthesis (DCS) [5]. This work is done in close cooperation with the Inria team Sumo at Inria Rennes (H. Marchand).

New results concern the master internship of Alia Hajjar, co-directed by Gwenaël Delaval an Stéphane Mocanu, on the subject of Application of control of reactive environments and probabilistic models on Transactional Memory. Multiprocessor environments which use concurrent programs and data structures showed the need of techniques to organize the usage of the shared structures, to reduce the unpredicted delay and reduce the contention between concurrent processors. Transactional Memory (TM) is a programming model that eases development of concurrent applications. Concurrent programming causes conflicts and TM is a way to resolve these conflicts with the transaction paradigm. To control conflict, techniques are provided to optimize (identify the best) degree of parallelism. In this framework, the aim is to control the TM system by adapting the degree of parallelism in order to maximize the throughput, i.e., number of committed transactions per time. The main objective is to minimize the execution time of a parallel application, thus maximize the throughput. During this master's thesis, the behavior of a multithreaded TM environment has been modeled as a stochastic discrete event system. The Heptagon/BZR language has then been used to implement this model for simulation, and evaluation of control strategies.

Ongoing work concerns aspects of compilation and debugging and exploring the notion of adaptive discrete control, which is yet an open question in discrete control in contrast to the well-known adaptive continuous control.

Another activity related to discrete control is or work with Leiden University and CWI (N. Khakpour, now at Linnaeus U., and F. Arbab) on enforcing correctness of the behavior of an adaptive software system during dynamic adaptation is an important challenge along the way to realize correct adaptive systems [11].

## 7.2. Design and programming

### 7.2.1. *Component-based approaches*

**Participants:** Frederico Alvares de Oliveira Junior, Eric Rutten.

Architecting in the context of variability has become a real need in today's software development. Modern software systems and their architecture must adapt dynamically to events coming from the environment (e.g., workload requested by users, changes in functionality) and the execution platform (e.g., resource availability). Component-based architectures have shown to be very suited for self-adaptation especially with their dynamical reconfiguration capabilities. However, existing solutions for reconfiguration often rely on low level, imperative, and non formal languages. We have defined Ctrl-F, a domain-specific language whose objective is to provide high-level support for describing adaptation behaviors and policies in component-based architectures. It relies on reactive programming for formal verification and control of reconfigurations. We integrate Ctrl-F with the FraSCAti Service Component Architecture middleware platform, and apply it to the Znn.com self-adaptive case study [20], [15], [14], [18].

We work on the topic in cooperation with the Spirals Inria team at Inria Lille (L. Seinturier). It constitutes a follow-up on previous work in the ANR Minalogic project MIND, industrializing the Fractal component-based framework, with a continuation of contacts with ST Microelectronics (V. Bertin). Our integration of BZR and Fractal [4], [2] is at the basis of our current work.

### *7.2.2. Rule-based systems*

**Participants:**  Adja Sylla, Eric Rutten.

We are starting a cooperation with CEA LETI/DACLE on the topic of a high-level language for safe rule-based programming in the LINC platform. The general context os that of the runtime redeployment of distributed applications, for example managing smart buildings. Motivations for redeployment can be diverse: load balancing, energy saving, upgrading, or fault tolerance. Redeployment involves changing the set of components in presence, or migrating them. The basic functionalities enabling to start, stop, migrate, or clone components, and the control managing their safe coordination, will have to be designed in the LINC middleware developed at CEA.

The transactional nature of the LINC platform insures the correct execution of each of the rules constituting the program, but there still is a need to insure the safety of their coordination, and of the behavior resulting from their sequential execution. For example, in the smart environments application domain, we must insure safety of control decisions, so that all the configurations that can be reached are safe, as well as the sequences of actions in switching between them. For this we will rely on automata-based models and control, using the BZR language, and integrating it in a domains specific language. Our work builds upon preliminary results involving colored Petri nets models [17].

The PhD of Adja Sylla at CEA on this topic is co-advised with F. Pacull and M. Louvel.

## 7.3. Infrastructure-level support

We apply the results of the previous axes of the team's activity 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 [13], towards a general view of Feedback Control as MAPE-K loop in Autonomic Computing [21].

### *7.3.1. Autonomic Cloud and Big-Data systems*

#### *7.3.1.1. Coordination in multiple-loop autonomic Cloud systems*
**Participants:**  Soguy Gueye, Gwenaël Delaval, Eric Rutten.

Complex computing systems are increasingly self-adaptive, with an autonomic computing approach for their administration. Real systems require the co-existence of multiple autonomic management loops, each complex to design. However their uncoordinated co-existence leads to performance degradation and possibly to inconsistency. There is a need for methodological supports facilitating the coordination of multiple autonomic managers. To tackle this problem, we take a global view and underscore that Autonomic Management Systems (AMS) are intrinsically reactive, as they react to flows of monitoring data by emitting flows of reconfiguration actions. Therefore we propose a new approach for the design of AMSs, based on synchronous programming and discrete controller synthesis techniques. They provide us with high-level languages for modeling the system to manage, as well as means for statically guaranteeing the absence of logical coordination problems. Hence, they suit our main contribution, which is to obtain guarantees at design time about the absence of logical inconsistencies in the taken decisions. We detail our approach, illustrate it by designing an AMS for a realistic multi-tier application, and evaluate its practicality with an implementation [10].

In order to coordinate managers without breaking their natural modularity. we address the problem with a method stressing modularity, and focusing on the discrete control of the interactions of managers. We make proposals for the distributed execution of modular controllers, first in synchronized way, and then relaxing this synchronization. We apply and validate our method on a multi-loop multi-tier system in a data-center [16].

We addressed these problems in the context of the ANR project Ctrl-Green, in cooperation with LIG (N. de Palma) in the framework of the PhD of S. Gueye and the post-doc of N. Berthier.

#### *7.3.1.2. Control for Big data*
**Participants:**  Bogdan Robu [Gipsa-lab], Mihaly Berekmeri [Gipsa-lab], Nicolas Marchand [Gipsa-lab].

Companies have a fast growing amounts of data to process and store, a data explosion is happening next to us. Currently one of the most common approaches to treat these vast data quantities is the MapReduce parallel programming paradigm. While it?s use is widespread in the industry, ensuring performance constraints, while also minimizing costs, provides considerable challenges. To deal with these issues we propose a control theoretical approach, based on techniques that have already proved their usefulness in the control community. We developed an algorithm to create the first linear dynamic model for a Big Data MapReduce Cloud system, running a concurrent workload. Furthermore we identify two important control use cases: relaxed performance - minimal resource and strict performance. We developed the first feedback control mechanism for such systems. Then to minimize the number of control actuations, an event-based feedback controller was also introduced. Furthermore to address the strict performance challenges a feedforward controller that efficiently suppresses the effects of large workload size variations is developed. On top of this issues an optimal predictive control which deals with concurrent objectives (dependability and performance) is implemented. The approach is validated online in a benchmark running in a real 60 node MapReduce cluster, using a data intensive Business Intelligence [22], [23].

This work is performed in cooperation with LIG (S. Bouchenak) in the framework of the PhD of M. Berekmeri.

### 7.3.2. *Reconfiguration control in DPR FPGA*
**Participant:** Eric Rutten.

Dynamically reconfigurable hardware has been identified as a promising solution for the design of energy efficient embedded systems. However, its adoption is limited by the costly design effort including verification and validation, which is even more complex than for non dynamically reconfigurable systems. We worked on this topic in the context of a ensign environment, developed in the framework of the ANR project Famous, in cooperation with LabSticc in Lorient and Inria Lille (DaRT team) [12]. We proposed a tool-supported formal method to automatically design a correct-by-construction control of the reconfiguration. By representing system behaviors with automata, we exploit automated algorithms to synthesize controllers that safely enforce reconfiguration strategies formulated as properties to be satisfied by control. We design generic modeling patterns for a class of reconfigurable architectures, taking into account both hardware architecture and applications, as well as relevant control objectives. We validate our approach on two case studies implemented on FPGAs [1].

We are currently valorizing results in more publications [12], [9], and extending the use of control techniques by evaluating the new tool ReaX developed at Inria Rennes (Sumo).

We are starting a new ANR project called HPeC, within which some of these topics will be extended, especially regarding hierarchical and modular control, and logico-numeric aspects.

### 7.3.3. *Autonomic memory management in HPC*
**Participants:** Naweiluo Zhou, Gwenaël Delaval, Bogdan Robu, Eric Rutten.

Parallel programs need to manage the time trade-off between synchronization and computation. A high parallelism may decrease computing time but meanwhile increase synchronization cost among threads. Software Transactional Memory (STM) has emerged as a promising technique, which bypasses locks, to address synchronization issues through transactions. A way to reduce conflicts is by adjusting the parallelism, as a suitable parallelism can maximize program performance. However, there is no universal rule to decide the best parallelism for a program from an offline view. Furthermore, an offline tuning is costly and error-prone. Hence, it becomes necessary to adopt a dynamical tuning-configuration strategy to better manage a STM system. Autonomic control techniques begin to receive attention in computing systems recently. Control technologies offer designers a framework of methods and techniques to build autonomic systems with well-mastered behaviors. The key idea of autonomic control is to implement feedback control loops to design safe, efficient and predictable controllers, which enable monitoring and adjusting controlled systems dynamically while keeping overhead low. We propose to design feedback control loops to automate the choice of parallelism at runtime and diminish program execution time.

In the context of the action-team HPES of the Labex Persyval-lab[0] (see 9.1 ), this work is performed in cooperation with LIG (J.F. Méhaut) in the framework of the PhD of N. Zhou.

### 7.3.4. *Control of smart environments*

**Participants:** Adja Sylla, Mengxuan Zhao, Eric Rutten, Hassane Alla [Gipsa-lab].

#### 7.3.4.1. *Generic supervision architecture*

New application domains of control, such as in the Internet of Things (IoT) and Smart Environments, require generic control rules enabling the systematization and the automation of the controller synthesis. We worked on an approach for the generation of Discrete Supervisory Controllers for these applications. A general modeling framework is proposed for the application domain of smart home. We formalize the design of the environment manager as a Discrete Controller Synthesis (DCS) problem, w.r.t. multiple constraints and objectives, for example logical issues of mutual exclusion, bounding of power peaks. We validate our models and manager computations with the BZR language and an experimental simulator This work was performed in cooperation with Orange labs (G. Privat) in the framework of the Cifre PhD of M. Zhao [8].

#### 7.3.4.2. *Rule-based specification*

In the context of IoT applications like mart home environments, the rules for programming in the LINC framework are used as a flexible tool to govern the relations between sensors and actuators. Runtime coordination and formal analysis becomes a necessity to avoid side effects mainly when applications are critical. In cooperation with CEA LETI/DACLE, we are working on a case study for safe applications development in IoT and smart home environments [17].

---

[0]https://persyval-lab.org/en/sites/hpes

<p style="text-align:center;color:red;font-weight:bold;">DANTE Project-Team</p>

# 7. New Results

## 7.1. Graph & Signal Processing

**Participants:** Paulo Gonçalves Andrade, Éric Fleury, Benjamin Girault, Sarra Ben Alaya.

**Isometric Graph shift operator.**   In [14], [40], we proposed a new shift operator for graph signals, enforcing that our operator is isometric. Doing so, we ensure that as many properties of the time shift as possible get carried over. Finally, we show that our operator behaves reasonably for graph signals.

**Stationary graph signals.**   We extended the concept of stationary temporal signals to stationary graph signals [24]. We introduced the concept of strict sense stationary and wide sense stationary graph signals as a statistical invariance through an isometric graph translation. Using these definitions, we proposed a spectral characterisation of WSS graph signals allowing to study stationarity using only the spectral components of a graph signal. Finally, we applied this characterisation to a synthetic graph in order to study a few important stochastic graph signals. Also, using geographic data, we analysed data from a graph set of weather stations and showed evidence of stationarity in the temperature signal [36].

**Community mining with graph filters for correlation matrices.**   Communities are an important type of structure in networks. Graph filters, such as wavelet filter-banks, have been used to detect such communities as groups of nodes more densely connected together than with the outsiders. When dealing with times series, it is possible to build a relational network based on the correlation matrix. However, in such a network, weights assigned to each edge have different properties than those of usual adjacency matrices. As a result, classical community detection methods based on modularity optimisation are not consistent and the modularity needs to be redefined to take into account the structure of the correlation from random matrix theory. In our contribution [34], we addressed how to detect communities from correlation matrices, by filtering global modes and random parts using properties that are specific to the distribution of correlation eigenvalues. Based on a Louvain approach, an algorithm to detect multiscale communities was also developed, which yields a weighted hierarchy of communities. The implementation of the method using graph filters was also discussed.

**A strong Tauberian theorem for characteristic functions.**   In [20], we showed that a characteristic function which can be approximated at 0 by any polynomial of order n is actually n-times differentiable at 0. This fact is exploited to strengthen a tauberian-type result by Lukacs and provides the theoretical basis for a wavelet based non-parametric estimator of the tail index of a distribution. This work is a technical improvement of our previous contribution [53].

**Fractal Analysis of Fetal Heart Rate Variability.**   The fetal heart rate (FHR) is commonly monitored during labor to detect early fetal acidosis. FHR variability is traditionally investigated using Fourier transform, often with adult predefined frequency band powers and the corresponding LF/HF ratio. However, fetal conditions differ from adults and modify spectrum repartition along frequencies. The study we reported in [12] questioned the arbitrariness definition and relevance of the frequency band splitting procedure, and thus of the calculation of the underlying LF/HF ratio, as efficient tools for characterising intrapartum FHR variability. Then, we showed that the intrapartum FHR is characterised by fractal temporal dynamics and promotes the Hurst parameter as a potential marker of fetal acidosis. This parameter preserves the intuition of a power frequency balance, while avoiding the frequency band splitting procedure and thus the arbitrary choice of a frequency separating bands. The study also shows that extending the frequency range covered by the adult-based bands to higher and lower frequencies permits the Hurst parameter to achieve better performance for identifying fetal acidosis.

# 7.2. Performance analysis and networks protocols

**Participants:** Paulo Gonçalves Andrade, Thomas Begin, Anthony Busson, Isabelle Guérin Lassous, Laurent Reynaud, Thiago Wanderley Matos de Abreu.

**Global computing-network-visualisation.**    The PetaFlow application aims to contribute to the use of high performance computational resources for the benefit of society. To this goal the emergence of adequate information and communication technologies with respect to high performance computing-networking-visualisation and their mutual awareness is required. In the work published in [5], we present the developed technology and the algorithms that we applied to a real global peta-scale data intensive scientific problem with social and medical importance, i.e. human upper airflow modeling.

**Performance analysis of multi-hop flows in IEEE 802.11 networks**    Multi-hop wireless networks are often regarded as a promising means to extend the limited coverage area offered by WLANs. However, they are usually associated with poor and uncertain performance in terms of available bandwidth and packet losses, which clearly stands as a limitation to their use. In [7], we consider the performance evaluation of a multi-hop path (also called chain), based on the IEEE 802.11 DCF. The proposed modeling framework is constructive and versatile, so that it can handle various types of multi-hop wireless paths, including scenarios with two flows in opposite directions, and topologies where nodes are exposed to the well-known hidden node problem. The models derived from our framework are conceptually simple, easy to implement and produce generally accurate results for the attained goodput of flows, as well as the datagram loss probability. Typical relative errors for these two quantities are below a few percent. Also, fundamental phenomena occurring in multi-hop wireless networks such as performance collapse and starvation, are well captured by the models.

**Passive Measurement-based Estimator for the Standard Deviation of the End-to-End Delay.**
Emerging architectures for computer networks such as SDN aim at offering a better handling of flows with stringent requirements of QoS. On the one hand, operators would benefit from a detailed description of common network performance (e.g., end-to-end delay and end-to end loss ratio) including their first two moments, namely mean and standard deviation. Indeed, for many applications, the variability in the end-to-end delay (e.g., jitter) deeply affects the actual QoS experienced by a flow. On the other hand, the cost and nuisance associated with the instrumentation, the measurements, and the computations must be kept as low as possible. This typically prevents the availability of end-to-end measurements. In [30], we propose an algorithm to estimate the second moment of the end-to-end delay experienced by the packets of a flow based only on delay measurements locally collected by the network nodes. Our solution estimates the standard deviation of the end-to-end delay in an easy and computationally efficient way. Based on thousands of simulations using a real-life trace, our solution is found to be accurate, typically differing by only a few percent from the actual value of the standard deviation of the end-to-end delay.

**Design of a force-based controlled mobility on aerial vehicles for pest management.**    Vespa velutina, also known as the Asian hornet, is considered as an invasive species out of its native zone. In particular, since it preys on honey bees, its recent progression in Europe could soon pose a significant risk to the local apiculture activity. European beekeepers are therefore investigating adapted control strategies, including V. velutina nest destruction. Unfortunately, nest location pinpointing generally follows a manual process which can prove tedious, time-consuming and inaccurate. In [31], we propose the use of a network of micro aerial vehicles featuring autonomous and cooperative flight capabilities. We describe an adapted controlled mobility strategy and detail the design of our Virtual Force Protocol (VFP) which allows a swarm of vehicles to track and follow hornets to their nests, while maintaining connectivity through a wireless multi-hop communication route with a remote ground station used to store applicative data such as hornet trajectory and vehicle telemetry. In order to achieve the mission objectives with a minimum of vehicles, we identify through simulations appropriate value for the key parameters of VFP and discuss the obtained network performance.

**Channel assignment in IEEE 802.11-based substitution networks.**    A substitution network is a rapidly deployable wireless network that provides a backup solution to quickly react to failures

on an existing network. We assume that the substitution network uses Wi-Fi technology and that wireless routers are equipped with several Wi-Fi cards. The problem, addressed in this work, deals with the channel assignment to these wireless interfaces. In this particular context, there is only one source-destination pair for which paths are known in advance. It is then possible to derive an objective function, function of the channel assignment, that very precisely reflects the overall throughput that can be achieved in this network. This problemis formulated through a linear optimization problem for which we propose different heuristics. Simulation results, performed with ns-3, consider several scenarios, and compare our heuristics to the optimum. Simulations show that, with only a few wireless cards, the throughput is significantly increased. Also, we show that the objective function fits to the throughput measured with ns-3.

**Performance evaluation and message dissemination in vehicular networks.**     Vehicular     Ad-Hoc Network (VANET) is becoming a promising technology for improving the efficiency and the safety of Intelligent Transportation Systems (ITS). Smart vehicles are expected to continuously exchange a huge amount of data either through safety or non-safety messages dedicated for road safety or infotainment and passenger comfort applications, respectively. In this context we proposed two contributions: the estimation of the capacity offered by the wireless network [13] in order to dimension the applications, and the proposal of an efficient message dissemination protocol [25].

**Performance Evaluation of Cloud Computing Centers with General Arrivals and Service.**     Cloud providers need to size their systems to determine the right amount of resources to allocate as a function of customer's needs so as to meet their SLAs (Service Level Agreement), while at the same time minimizing their costs and energy use. Queueing theory based tools are a natural choice when dealing with performance aspects of the QoS (Quality of Service) part of the SLA and forecasting resource utilization. The characteristics of a cloud center lead to a queueing system with multiple servers (nodes) in which there is potentially a very large number of servers and both the arrival and service process can exhibit high variability. We propose to use a G/G/c-like model to represent a cloud system and assess expected performance indices. Given the potentially high number of servers in a cloud system, we present an efficient, fast and easy-to-implement approximate solution. We have extensively validated our approximation against discrete-event simulation for several QoS performance metrics such as task response time and blocking probability with excellent results. We apply our approach to examples of system sizing and our examples clearly demonstrate the importance of taking into account the variability of the tasks arrivals and thus expose the risk of under- or over-provisioning if one relies on a model with Poisson assumptions [8].

**Prediction of the System Performance from components models.**     In this paper we consider the problem of combining calibrated performance models of system components in order to predict overall system performance. We focus on open workload system models, in which, under certain conditions, obtaining and validating the overall system performance measures can be a simple application of Little's law. We discuss the conditions of applicability of such a simple validation methodology, including examples of successful application, as well as examples where this approach fails. Additionally, we propose to analyze the deviations between the model predictions and system measurements, so as to decide if they correspond to "measurement noise" or if an important system component has not been correctly represented. This approach can be used as an aid in the design of validated system performance models [26].

## 7.3. Modeling of Dynamics of Complex Networks

**Participants:**  Christophe Crespelle, Éric Fleury, Márton Karsai, Yannick Leo, Matteo Morini.

**Non-Altering Time Scales for Aggregation of Dynamic Networks into Series of Graphs** [29]    Many dynamic networks coming from real-world contexts are *link streams*, i.e. a finite collection of triplets $(u, v, t)$ where $u$ and $v$ are two nodes having a link between them at time $t$. A great number of studies on these objects start by aggregating the data on disjoint time windows of length $\Delta$ in order to obtain a series of graphs on which are made all subsequent analyses. Here we are concerned

with the impact of the chosen $\Delta$ on the obtained graph series. We address the fundamental question of knowing whether a series of graphs formed using a given $\Delta$ faithfully describes the original link stream. We answer the question by showing that such dynamic networks exhibit a threshold for $\Delta$, which we call the *saturation scale*, beyond which the properties of propagation of the link stream are altered, while they are mostly preserved before. We design an automatic method to determine the saturation scale of any link stream, which we apply and validate on several real-world datasets.

**Termination of the Iterated Strong-Factor Operator on Multipartite Graphs** [10]    The clean-factor operator is a multipartite graph operator that has been introduced in the context of complex network modelling. Here, we consider a less constrained variation of the clean-factor operator, named strong-factor operator, and we prove that, as for the clean-factor operator, the iteration of the strong-factor operator always terminates, independently of the graph given as input. Obtaining termination for all graphs using minimal constraints on the definition of the operator is crucial for the modelling purposes for which the clean-factor operator has been introduced. Moreover we show that the relaxation of constraints we operate not only preserves termination but also preserves the termination time, in the sense that the strong-factor series always terminates before the clean-factor series.

**On the Termination of Some Biclique Operators on Multipartite Graphs** [9]    We define a new graph operator, called the *weak-factor graph*, which comes from the context of complex network modelling. The weak-factor operator is close to the well-known clique-graph operator but it rather operates in terms of bicliques in a multipartite graph. We address the problem of the termination of the series of graphs obtained by iteratively applying the weak-factor operator starting from a given input graph. As for the clique-graph operator, it turns out that some graphs give rise to series that do not terminate. Therefore, we design a slight variation of the weak-factor operator, called *clean-factor*, and prove that its associated series terminates for all input graphs. In addition, we show that the multipartite graph on which the series terminates has a very nice combinatorial structure: we exhibit a bijection between its vertices and the chains of the inclusion order on the intersections of the maximal cliques of the input graph.

**Directed Cartesian-Product Decomposition** [11].    In this paper, we design an algorithm that, given a directed graph $G$ and the Cartesian-product decomposition of its underlying undirected graph $\widetilde{G}$, produces the directed Cartesian-product decomposition of $G$ in linear time. This is the first time that the linear complexity is achieved for this problem, which has two major consequences. Firstly, it shows that the directed and undirected versions of the Cartesian-product decomposition of graphs are linear-time equivalent problems. And secondly, as there already exists a linear-time algorithm for solving the undirected version of the problem, combined together, it provides the first linear-time algorithm for computing the directed Cartesian-product decomposition of a directed graph.

**An $O(n^2)$ time Algorithm for the Minimal Permutation Completion Problem** [28]    We provide an $O(n^2)$ time algorithm computing a minimal permutation completion of an arbitrary graph $G = (V, E)$, i.e., a permutation graph $H = (V, F)$ on the same vertex set, such that $E \subseteq F$ and $F$ is inclusion-minimal among all possibilities.

**Linearity is Strictly More Powerful than Contiguity for Encoding Graphs** [27]    Linearity and contiguity are two parameters devoted to graph encoding. Linearity is a generalisation of contiguity in the sense that every encoding achieving contiguity $k$ induces an encoding achieving linearity $k$, both encoding having size $\Theta(k.n)$, where $n$ is the number of vertices of $G$. In this paper, we prove that linearity is a strictly more powerful encoding than contiguity, i.e. there exists some graph family such that the linearity is asymptotically negligible in front of the contiguity. We prove this by answering an open question asking for the worst case linearity of a cograph on $n$ vertices: we provide an $O(\log n / \log \log n)$ upper bound which matches the previously known lower bound.

**Socioeconomic correlations in communication networks** [37], [38]    In this work we study the socioeconomic structure of a communication network by combining mobile communication records and bank credit informations of a large number of individuals living in Mexico. We provide empirical evidences about present economic unbalances suggesting not only the distribution of wealth but also

the distribution of debts to follow the Pareto principle. Further we study the internal and interconnected structure of socioeconomic groups. Through a weighted core analysis we signal assortative correlations between people regarding their economic capacities, and show the existence of "rich-clubs" indicating present social stratification in the social structure. This project is ongoing with final results expected in 2016.

**Detecting global bridges in networks** [15]   The identification of nodes occupying important positions in a network structure is crucial for the understanding of the associated real-world system. Usually, betweenness centrality is used to evaluate a node capacity to connect different graph regions. However, we argue here that this measure is not adapted for that task, as it gives equal weight to "local" centers (i.e. nodes of high degree central to a single region) and to "global" bridges, which connect different communities. This distinction is important as the roles of such nodes are different in terms of the local and global organisation of the network structure. In this paper we propose a decomposition of betweenness centrality into two terms, one highlighting the local contributions and the other the global ones. We call the latter bridgeness centrality and show that it is capable to specifically spot out global bridges. In addition, we introduce an effective algorithmic implementation of this measure and demonstrate its capability to identify global bridges in air transportation and scientific collaboration networks.

**Collective attention in the age of (mis)information** [17]   We study, on a sample of 2.3 million individuals, how Facebook users consumed different information at the edge of political discussion and news during the last Italian electoral competition. Pages are categorized, according to their topics and the communities of interests they pertain to, in a) alternative information sources (diffusing topics that are neglected by science and main stream media); b) online political activism; and c) main stream media. We show that attention patterns are similar despite the different qualitative nature of the information, meaning that unsubstantiated claims (mainly conspiracy theories) reverberate for as long as other information. Finally, we categorize users according to their interaction patterns among the different topics and measure how a sample of this social ecosystem (1279 users) responded to the injection of 2788 false information posts. Our analysis reveals that users which are prominently interacting with alternative information sources (i.e. more exposed to unsubstantiated claims) are more prone to interact with false claims.

**The Scaling of Human Contacts in Reaction-Diffusion Processes**  [22]   We present new empirical evidence, based on millions of interactions on Twitter, confirming that human contacts scale with population sizes. We integrate such observations into a reaction-diffusion metapopulation framework providing an analytical expression for the global invasion threshold of a contagion process. Remarkably, the scaling of human contacts is found to facilitate the spreading dynamics. Our results show that the scaling properties of human interactions can significantly affect dynamical processes mediated by human contacts such as the spread of diseases, and ideas.

**From calls to communities: a model for time varying social networks** [16]   Social interactions vary in time and appear to be driven by intrinsic mechanisms, which in turn shape the emerging structure of the social network. Large-scale empirical observations of social interaction structure have become possible only recently, and modelling their dynamics is an actual challenge. Here we propose a temporal network model which builds on the framework of activity-driven time-varying networks with memory. The model also integrates key mechanisms that drive the formation of social ties - social reinforcement, focal closure and cyclic closure, which have been shown to give rise to community structure and the global connectedness of the network. We compare the proposed model with a real-world time-varying network of mobile phone communication and show that they share several characteristics from heterogeneous degrees and weights to rich community structure. Further, the strong and weak ties that emerge from the model follow similar weight-topology correlations as real-world social networks, including the role of weak ties.

**Kinetics of Social Contagion** [21]   Diffusion of information, behavioural patterns or innovations follows diverse pathways depending on a number of conditions, including the structure of the underlying social network, the sensitivity to peer pressure and the influence of media. Here we study

analytically and by simulations a general model that incorporates threshold mechanism capturing sensitivity to peer pressure, the effect of 'immune' nodes who never adopt, and a perpetual flow of external information. While any constant, non-zero rate of dynamically-introduced innovators leads to global spreading, the kinetics by which the asymptotic state is approached show rich behaviour. In particular we find that, as a function of the density of immune nodes, there is a transition from fast to slow spreading governed by entirely different mechanisms. This transition happens below the percolation threshold of fragmentation of the network, and has its origin in the competition between cascading behaviour induced by innovators and blocking of adoption due to immune nodes. This change is accompanied by a percolation transition of the induced clusters.

# MESCAL Project-Team

# 6. New Results

## 6.1. Reproducible Research

In the field of large-scale distributed systems, experimentation is particularly difficult. The studied systems are complex, often non-deterministic and unreliable, software is plagued with bugs, whereas the experiment workflows are unclear and hard to reproduce. In [11], we provide an extensive list of features offered by general-purpose experiment management tools dedicated to distributed systems research on real platforms. We then use it to assess existing solutions and compare them, outlining possible future paths for improvements.

In [20], we address the question of developing a lightweight and effective workflow for conducting experimental research on modern parallel computer systems in a reproducible way. Our workflow simply builds on two well-known tools (Org-mode and Git) and enables us to address issues such as provenance tracking, experimental setup reconstruction, replicable analysis. Although this workflow is perfectible and cannot be seen as a final solution, we have been using git for two years now and we have recently published a fully reproducible article, which demonstrates the effectiveness of our proposal.

## 6.2. Performance Characterization and Optimization of IOs

In high-performance computing environments, parallel file systems provide a shared storage infrastructure to applications. In the situation where multiple applications access this shared infrastructure concurrently, their performance can be impaired because of interference. In [22], we improve improve performance by alleviating interference effects through a smart I/O scheduler scheduler that organizes and optimizes the applications' requests and adjusts the access pattern to the device characteristics. We apply machine learning techniques to automatically select the best scheduling algorithm for each situation. Our approach improves performance by up to 75

In [33], we present a new storage device profiling tool that characterizes the sequential to random throughput ratio for reads and writes of different sizes. As we explained previously, several optimizations aim at adapting applications' access patterns in order to generate contiguous accesses for improved performance when accessing storage devices like hard disks. However, when considering other storage options like RAID arrays and SSDs, the access time ratio between contiguous and non-contiguous accesses may not compensate for these optimizations' cost. In this scenario, the information provided by our tool could be used to dynamically decide if optimizations are beneficial for performance, which is why we took a particular attention to obtain accurate information in a minimal benchmarking time.

## 6.3. Application of Game Theory and Distributed Optimization to Wireless Networks

In wireless networks, channel conditions and user quality of service (QoS) requirements vary, often quite arbitrarily, with time (e.g. due to user mobility, fading, etc.) and users only have a very limited information about their environment. In such context optimizing transmission while taking power consumption into account is extremely challenging. We apply game theory technique to MIMO wireless network using OFDM or OFDMA where multi-path channels can be handled efficiently

In [25], [9], we show that distributed power allocation in heterogeneous OFDMA cognitive radio networks can be modeled as a game where each user equipment in the network engages in a non-cooperative game and allocates its available transmit power over subcarriers to maximize its individual utility. The corresponding equilibrium (Debreu, an extension of Nash Equilibrium) can be characterized with fractional programming and we provide sufficient conditions for computing such equilibria as fixed points of a water-filling best response operator. Using such approach can however be quite slow and is very sensitive to delay and information uncertainty (it may not converge). Therefore, we explain in [17] how signal covariance matrices in Gaussian MIMO multiple access channel can be learnt in presence of imperfect (and possibly delayed) feedback. The algorithm we propose is based on on the method of matrix exponential learning (MXL) and it has the same information and computation requirements as distributed water-filling. However our algorithm converge much faster even for large numbers of users and/or antennas per user and in the presence of user update asynchrony, random delays and/or ergodically changing channel conditions. Yet, since the system may evolve over time in an unpredictable fashion (e.g. due to changes in the wireless medium or the users' QoS requirements), static solution concepts (such as Nash equilibrium) may be no longer relevant and users must adapt to changes in the environment "on the fly", without being able to predict the system's evolution ahead of time. Hence, we focus on the concept of no-regret : policies that perform at least as well as the best fixed transmit profile in hindsight. In [31] and [41], we provide a formulation of power control as an online optimization problem and we show that the FM dynamics lead to no regret in this dynamic context. In [40] we apply this approach energy efficient transmission in MIMO-OFDM systems and we show through numerical simulations that, in realistic network environments even under rapidly changing channel conditions, users can track their individually optimum transmit profile, achieving gains of up to 600in energy efficiency over uniform power allocation policies.

We also apply this technique to multi-carrier cognitive radio systems. Such systems allow opportunistic secondary users (SUs) to access portions of the spectrum that are unused by the network's licensed primary users (PUs), provided that the induced interference does not compromise the PUs' performance guarantees. In [14], we introduce a flexible spectrum access pricing schemes such that the corresponding Nash equilibrium is unique under very mild assumptions and satisfies the performance constraints. In addition, we derive a dynamic power allocation policy that converges to equilibrium within a few iterations (even for large numbers of users) and that relies only on local—and possibly imperfect—signal-to-interference-and-noise ratio measurements. In [24], we draw on exponential learning techniques to design an algorithm that is able to adapt to system changes "on the fly", i.e. such that the proposed transmit policy leads to no regret even under rapidly changing network conditions.

## 6.4. General Results in Game Theory

Our work on game theory is often motivated by applications to wireless networks but can often have a more general application.

In [38], motivated by applications to multi-antenna wireless networks, we propose a distributed and asynchronous algorithm for stochastic semidefinite programming This algorithm is a stochastic approximation of a continous-time matrix exponential scheme regularized by the addition of an entropy-like term to the problem's objective function. We show that the resulting algorithm converges almost surely to an $(\epsilon)$-approximation of the optimal solution requiring only an unbiased estimate of the gradient of the problem's stochastic objective.

As explained in the previous section, classical Nash equilibrium concepts become irrelevant in situations where the environment evolves over time. In [15], we study one of the main concept of online learning and sequential decision problem known as regret minimization. Our objective is to provide a quick overview and a comprehensive introduction to online learning and game theory.

In practice, it is rarely reasonable to assume that players have access to the strategy of the others and implementing a best response can thus become cumbersome. Replicator dynamics is a fundamental approach in evolutionary game theory in which players adjust their strategies based on their actions' cumulative payoffs over time – specifically, by playing mixed strategies that maximize their expected cumulative payoff.

- In [19], we investigate the impact of payoff shocks on the evolution of large populations of myopic players that employ simple strategy revision protocols such as the "imitation of success". In the

noiseless case, this process is governed by the standard (deterministic) replicator dynamics; in the presence of noise however, the induced stochastic dynamics are different from previous versions of the stochastic replicator dynamics (such as the aggregate-shocks model of Fudenberg and Harris, 1992). In this context, we show that strict equilibria are always stochastically asymptotically stable, irrespective of the magnitude of the shocks; on the other hand, in the high-noise regime, non-equilibrium states may also become stochastically asymptotically stable and dominated strategies may survive in perpetuity (they become extinct if the noise is low). Such behavior is eliminated if players are less myopic and revise their strategies based on their cumulative payoffs. In this case, we obtain a second order stochastic dynamical system whose attracting states coincide with the game's strict equilibria and where dominated strategies become extinct (a.s.), no matter the noise level.

- In [13], we study a new class of continuous-time learning dynamics consisting of a replicator-like drift adjusted by a penalty term that renders the boundary of the game's strategy space repelling. These penalty-regulated dynamics are equivalent to players keeping an exponentially discounted aggregate of their ongoing payoffs and then using a smooth best response to pick an action based on these performance scores. Building on the duality with evolutionary game theory, we design a discrete-time, payoff-based learning algorithm that converges to (arbitrarily precise) approximations of Nash equilibria in potential games. Moreover, the algorithm remains robust in the presence of stochastic perturbations and observation errors, and it does not require any synchronization between players, which is a very important property when applying such technique to traffic engineering.

- In [18], we investigate an other class of reinforcement learning dynamics in which the players strategy adjustment is regularized with a strongly convex penalty term. In contrast to the class of penalty functions used to define smooth best responses in models of stochastic fictitious play, the regularizers used in this paper need not be infinitely steep at the boundary of the simplex. Dropping this requirement gives rise to an important dichotomy between steep and non-steep cases. In this general setting, our main results extend several properties of the replicator dynamics such as the elimination of dominated strategies, the asymptotic stability of strict Nash equilibria and the convergence of time-averaged trajectories to interior Nash equilibria in zero-sum games.

- In [37], we study a general class of game-theoretic learning dynamics in the presence of random payoff disturbances and observation noise, and we provide a unified framework that extends several rationality properties of the (stochastic) replicator dynamics and other game dynamics. In the unilateral case, we show that the stochastic dynamics under study lead to no regret, irrespective of the noise level. In the multi-player case, we find that dominated strategies become extinct (a.s.) and strict Nash equilibria remain stochastically asymptotically stable – again, independently of the perturbations' magnitude. Finally, we establish an averaging principle for 2-player games and we show that the empirical distribution of play converges to Nash equilibrium in zero-sum games under any noise level.

## 6.5. Simulation

Simgrid is a toolkit providing core functionalities for the simulation of distributed applications in heterogeneous distributed environments. Although it was initially designed to study large distributed computing environments such as grids, we have recently applied it to performance prediction of HPC configurations.

- Indeed, multi-core architectures comprising several GPUs have become mainstream but obtaining the maximum performance of such heterogeneous machines is challenging as it requires to carefully offload computations and manage data movements between the different processing units. The most promising and successful approaches so far build on task-based runtimes that abstract the machine and rely on opportunistic scheduling algorithms. As a consequence, the problem gets shifted to choosing the task granularity, task graph structure, and optimizing the scheduling strategies. Trying different combinations of these different alternatives is also itself a challenge. Indeed, getting accurate measurements requires reserving the target system for the whole duration of experiments. Furthermore, observations are limited to the few available systems at hand and may be difficult

to generalize. In [21], we show how we crafted a coarse-grain hybrid simulation/emulation of StarPU, a dynamic runtime for hybrid architectures, over SimGrid. This approach allows to obtain performance predictions of classical dense linear algebra kernels accurate within a few percents and in a matter of seconds, which allows both runtime and application designers to quickly decide which optimization to enable or whether it is worth investing in higher-end GPUs or not. Additionally, it allows to conduct robust and extensive scheduling studies in a controlled environment whose characteristics are very close to real platforms while having reproducible behavior. In [30], we have extended this approach to the simulation of a multithreaded multifrontal QR solver of sparse matrices: QR-MUMPS. In our approach, the target high-end machines are calibrated only once to derive sound performance models. These models can then be used at will to quickly predict and study in a reproducible way the performance of such irregular and resource-demanding applications using solely a commodity laptop. Our approach also allows to study the memory consumption along time, which is a critical factor for such applications.

- Beside the inherent heterogeneity of distributed computing infrastructures, storage is also a essential component to cope with the tremendous increase in scientific data production and the ever-growing need for data analysis and preservation. Understanding the performance of a storage subsystem or dimensioning it properly is an important concern for which simulation can help. In [29], we detail how we have extended SimGrid with storage simulation capacities and we list several concrete use cases of storage simulations in clusters, grids, clouds, and data centers for which the proposed extension would be beneficial.

$\Psi^2$ is a simulation software of markovian models that is able to provide a perfect sampling of the stationary distribution. In [12], we consider open Jackson networks with losses with mixed finite and infinite queues and analyze the efficiency of sampling from their exact stationary distribution. We show that perfect sampling is possible, although the underlying Markov chain may have an infinite state space. The main idea is to use a Jackson network with infinite buffers (that has a product form stationary distribution) to bound the number of initial conditions to be considered in the coupling from the past scheme. We also provide bounds on the sampling time of this new perfect sampling algorithm for acyclic or hyper-stable networks. These bounds show that the new algorithm is considerably more efficient than existing perfect samplers even in the case where all queues are finite. We illustrate this efficiency through numerical experiments. We also extend our approach to variable service times and non-monotone networks such as queueing networks with negative customers.

## 6.6. Asymptotic Models

Analyzing a set of $n$ stochastic entities interacting with each others can be particularly difficult but the *mean field approximation* is a very effective technique to characterize the probability distribution of such systems when the number of entities $n$ grows very large. The limit system is generally deterministic and characterized by a differential equation that is more amenable to analysis and optimization. Such approximation however typically requires that the dynamics of the entities depend only on their state (the state space of each object does not scale with $n$ the number of objects) but neither on their identity nor on their spatial location.

- In [28], we analyze a family of list-based cache replacement algorithms. We present explicit expressions for the cache content distribution and miss probability under some assumptions and we develop an algorithm with a time complexity that is polynomial in the cache size and linear in the number of items to compute the exact miss probability. We further introduce a mean field model to approximate the transient behavior of the miss probability and prove that this model becomes exact as the cache size and number of items tends to infinity. We show that the set of ODEs associated to the mean field model has a unique fixed point that can be used to approximate the miss probability in case the exact computation becomes too time consuming. Using this approximation, we provide guidelines on how to select a replacement algorithm within the family considered such that a good trade-off is achieved between the cache reactivity and its steady-state hit probability

- For distributed systems where /locality/ is essential in the dynamics the mean-field approach requires to resort to discretization of space into a finite number of cells to fit in the classical framework.

Such approach not only scales badly but also requires that spatial interactions are weak. One of the tool to tackle this difficult problem comes from statistical physics and is popular in biology: pair approximation. In [26], we successfully apply this approach to the "Power of Two Choice" load balancing paradigm: each incoming task is allocated to the least loaded of two servers picked at random among a collection of $n$ servers. We study the power of two-choice in a setting where the two servers are not picked independently at random but are connected by an edge in an underlying graph. Our problem is motivated by systems in which choices are geometrically constrained (e.g., a bike-sharing system). We study a dynamic setting in which jobs leave the system after being served by a server to which is was allocated. Our focus is when each server has few neighbors (typically 2 to 4) for which an mean-field approximation is not accurate. We build the pair-approximation equations and show that they describe accurately the steady-state of the system. Our results show that, even in a graph of degree 2, choosing between two neighboring improve dramatically the performance compared to a random allocation.

- In [8], we consider a queueing system composed of a dispatcher that routes deterministically jobs to a set of non-observable queues working in parallel. In this setting, the fundamental problem is which policy should the dispatcher implement to minimize the stationary mean waiting time of the incoming jobs. We present a structural property that holds in the classic scaling of the system where the network demand (arrival rate of jobs) grows proportionally with the number of queues. Assuming that each queue of type $r$ is replicated $k$ times, we consider a set of policies that are periodic with period $k \sum_r p_r$ and such that exactly $p_r$ jobs are sent in a period to each queue of type $r$. When $k \to \infty$, our main result shows that all the policies in this set are equivalent, in the sense that they yield the same mean stationary waiting time, and optimal, in the sense that no other policy having the same aggregate arrival rate to all queues of a given type can do better in minimizing the stationary mean waiting time. Furthermore, the limiting mean waiting time achieved by our policies is a convex function of the arrival rate in each queue, which facilitates the development of a further optimization aimed at solving the fundamental problem above for large systems.

## 6.7. Trace and Statistical Analysis

Although we often use Markovian approaches to model large scale distributed system, these probabilistic tools can also be used to lay the foundation of statistical analysis of traces of real systems.

- In [36], we explain how we apply statistical statistical modelling and statistical inference of the ANR GEOMEDIA corpus, that is a collection of international RSS news feeds. Central to this project, RSS news feeds are viewed as a representation of the information in geopolitical space. As such they allow us to study media events of global extent and how they affect international relations. Here we propose hidden Markov models (HMM) as an adequate modelling framework to study the evolution of media events in time. This set of models respect the characteristic properties of the data, such as temporal dependencies and correlations between feeds. Its specific structure corresponds well to our conceptualisation of media attention and media events. We specify the general model structure that we use for modelling an ensemble of RSS news feeds. Finally, we apply the proposed models to a case study dedicated to the analysis of the media attention for the Ebola epidemic which spread through West Africa in 2014.

- The use of stochastic formalisms, such as Stochastic Automata Networks (SAN), can be very useful for statistical prediction and behavior analysis. Once well fitted, such formalisms can generate probabilities about a target reality. These probabilities can be seen as a statistical approach of knowledge discovery. However, the building process of models for real world problems is time consuming even for experienced modelers. Furthermore, it is often necessary to be a domain specialist to create a model. In [34], we present a new method to automatically learn simple SAN models directly from a data source. This method is encapsulated in a tool called SAN GEnerator (SANGE). Through examples we show how this new model fitting method is powerful and relatively easy to use, which can grant access to a much broader community to such powerful modeling formalisms.

- In [32], we have presented our recent results on macroscopic analysis of huge traces of parallel/distributed applications. To identify a *macroscopic phenomenon* over large traces, one needs to change the representation scale and to aggregate data both in time, space and application structure through meaningful operators to propose *multi-scale visualizations*. The question is then to know the quantity of information lost by such scaling to be able to correctly interpret them. The principles underlying this approach are based on information theory since the conditional entropy of an aggregation indicates the quantity of information loss when data are aggregated. This approach has been integrated in the Framesoc framework [35].

- In [27], We study the problem of making forecasts about the future availability of bicycles in stations of a bike-sharing system (BSS). This is relevant in order to make recommendations guaranteeing that the probability that a user will be able to make a journey is sufficiently high. To this end, we use probabilistic predictions obtained from a queuing theoretical time-inhomogeneous model of a BSS. The model is parametrized and successfully validated using historical data from the Vélib ' BSS of the City of Paris. We develop a critique of the standard root-mean-square-error (RMSE), commonly adopted in the bike-sharing research as an index of the prediction accuracy, because it does not account for the stochasticity inherent in the real system. Instead we introduce a new metric based on scoring rules. We evaluate the average score of our model against classical predictors used in the literature. We show that these are outperformed by our model for prediction horizons of up to a few hours. We also discuss that, in general, measuring the current number of available bikes is only relevant for prediction horizons of up to few hours.

# MOAIS Project-Team

# 5. New Results

## 5.1. TABARNAC: Tools for Analyzing Behavior of Applications Running on NUMA Architecture.

In modern parallel architectures, memory accesses represent a common bottleneck. We develop TABARNAC, a tool for analyzing the memory behavior of parallel applications with a focus on NUMA architectures. TABARNAC provides a new visualization of the memory access behavior, focusing on the distribution of accesses by thread and by structure. Such visualization allows the developer to easily understand why performance issues occur. Using TABARNAC, we explain why some applications do not benefit from data and thread mapping. Moreover, we propose several code modifications to improve the memory access behavior of several parallel applications [29].

## 5.2. Computing the Rank Profile Matrix

We propose the definition of a new matrix invariant, the rank profile matrix, summarizing all information on the row and column rank profiles of all the leading sub-matrices. We also explore the conditions for a Gaussian elimination algorithm to compute all or part of this invariant, through the corresponding PLUQ decomposition [12].

## 5.3. Parallel Algebraic Linear Algebra Dedicated Interface

We propose a domain specific language based on C/C++ macros, PALADIn (Parallel Algebraic Linear Algebra Dedicated Interface) [15]. This domain specific language allows the user to write C++ code and benefits from sequential and parallel executions on shared memory architectures. With a unique syntax, the user can switch between different parallel runtime systems such as OpenMP, TBB and xKaapi. This interface provides data and task parallelism and has been used for recursion-based parallelization of exact dense linear algebra routines[7].

## 5.4. Communication models insights meet simulations

It is well-known that taking into account communications while scheduling jobs in large scale parallel computing platforms is a crucial issue. In modern hierarchical platforms, communication times are highly different when occurring inside a cluster or between clusters. Thus, allocating the jobs taking into account locality constraints is a key factor for reaching good performances. However, several theoretical results prove that imposing such constraints reduces the solution space and thus, possibly degrades the performances. In practice, such constraints simplify implementations and most often lead to better results. Our aim in this work is to bridge theoretical and practical intuitions, and check the differences between constrained and unconstrained schedules (namely with respect to locality and node contiguity) through simulations. We have developed a generic tool, using SimGrid as the base simulator, enabling interactions with external batch schedulers to evaluate their scheduling policies. The results confirm that insights gained through theoretical models are ill-suited to current architectures and should be reevaluated [13].

## 5.5. Adaptive Resource and Job Management for Limited Power Consumption

The last decades have been characterized by an evergrowing requirement in terms of computing and storage resources. This tendency has recently put the pressure on the ability to efficiently manage the power required to operate the huge amount of electrical components associated with state-of-the-art high performance computing systems. The power consumption of a supercomputer needs to be adjusted based on varying power budget or electricity availabilities. As a consequence, Resource and Job Management Systems have to be adequately adapted in order to efficiently schedule jobs with optimized performance while limiting power usage whenever needed. We introduce in this paper a new scheduling strategy that can adapt the executed workload to a limited power budget. The originality of this approach relies upon a combination of speed scaling and node shutdown techniques for power reductions. It is implemented into the widely used resource and job management system SLURM. Finally, it is validated through large scale emulations using real production workload traces of the supercomputer Curie [17].

## 5.6. Lessons Learned from Building In Situ Coupling Frameworks

Over the past few years, the increasing amounts of data produced by large-scale simulations have motivated a shift from traditional offline data analysis to in situ analysis and visualization. In situ processing began as the coupling of a parallel simulation with an analysis or visualization library, motivated primarily by avoiding the high cost of accessing storage. Going beyond this simple pairwise tight coupling, complex analysis workflows today are graphs with one or more data sources and several interconnected analysis components. In this paper, we review four tools that we have developed to address the challenges of coupling simulations with visualization packages or analysis workflows: Damaris, Decaf, FlowVR and Swift. This self-critical inquiry aims to shed light not only on their potential, but most importantly on the forthcoming software challenges that these and other in situ analysis and visualization frameworks will face in order to move toward exascale [11]. Besides, focusing on asynchronous In Situ Processing with Gromacs, we have exhibited how to take Advantage of GPUs [25].

## 5.7. Design and analysis of scheduling strategies for multi-CPU and multi-GPU architectures

In [8], we present a comparison of scheduling strategies for heterogeneous multi-CPU and multi-GPU architectures. We designed and evaluated four scheduling strategies on top of XKaapi runtime: work stealing, data-aware work stealing, locality-aware work stealing, and Heterogeneous Earliest-Finish-Time (HEFT). On a heterogeneous architecture with 12 CPUs and 8 GPUs, we analysed our scheduling strategies with four benchmarks: a BLAS-1 AXPY vector operation, a Jacobi 2D iterative computation, and two linear algebra algorithms Cholesky and LU. We conclude that the use of work stealing may be efficient if task annotations are given along with a data locality strategy. Furthermore, our experimental results suggests that HEFT scheduling performs better on applications with very regular computations and low data locality.

<p style="text-align:center; color:red;">**ROMA Project-Team**</p>

# 7. New Results

## 7.1. Scheduling computational workflows on failure-prone platforms

**Participants:** Guillaume Aupy, Anne Benoit, Henri Casanova [University of Hawaii], Yves Robert.

We study the scheduling of computational workflows on compute resources that experience exponentially distributed failures. When a failure occurs, rollback and recovery is used to resume the execution from the last checkpointed state. The scheduling problem is to minimize the expected execution time by deciding in which order to execute the tasks in the workflow and whether to checkpoint or not checkpoint a task after it completes. We give a polynomial-time algorithm for fork graphs and show that the problem is NP-complete with join graphs. Our main result is a polynomial-time algorithm to compute the execution time of a workflow with specified to-be-checkpointed tasks. Using this algorithm as a basis, we propose efficient heuristics for solving the scheduling problem. We evaluate these heuristics for representative workflow configurations.

This work has been published in the 17th Workshop on Advances in Parallel and Distributed Computational Models [20].

## 7.2. Efficient checkpoint/verification patterns

**Participants:** Anne Benoit, Saurabh K. Raina [Jaypee Institute of Information Technology], Yves Robert.

Errors have become a critical problem for high performance computing. Checkpointing protocols are often used for error recovery after fail-stop failures. However, silent errors cannot be ignored, and their peculiarity is that such errors are identified only when the corrupted data is activated. To cope with silent errors, we need a verification mechanism to check whether the application state is correct. Checkpoints should be supplemented with verifications to detect silent errors. When a verification is successful, only the last checkpoint needs to be kept in memory because it is known to be correct.

In this work, we analytically determine the best balance of verifications and checkpoints so as to optimize platform throughput. We introduce a balanced algorithm using a pattern with $p$ checkpoints and $q$ verifications, which regularly interleaves both checkpoints and verifications across same-size computational chunks. We show how to compute the waste of an arbitrary pattern, and we prove that the balanced algorithm is optimal when the platform MTBF (Mean Time Between Failures) is large in front of the other parameters (checkpointing, verification and recovery costs). We conduct several simulations to show the gain achieved by this balanced algorithm for well-chosen values of $p$ and $q$, compared to the base algorithm that always perform a verification just before taking a checkpoint ($p = q = 1$), and we exhibit gains of up to $19\%$.

This work has been published in the International Journal of High Performance Computing Applications [8].

## 7.3. Assessing the impact of partial verifications against silent data corruptions

**Participants:** Aurélien Cavelan, Saurabh K. Raina [Jaypee Institute of Information Technology], Yves Robert, Hongyang Sun.

Silent errors, or silent data corruptions, constitute a major threat on very large scale platforms. When a silent error strikes, it is not detected immediately but only after some delay, which prevents the use of pure periodic checkpointing approaches devised for fail-stop errors. Instead, checkpointing must be coupled with some verification mechanism to guarantee that corrupted data will never be written into the checkpoint file. Such a guaranteed verification mechanism typically incurs a high cost. In this work, we assess the impact of using partial verification mechanisms in addition to a guaranteed verification. The main objective is to investigate to which extent it is worthwhile to use some light cost but less accurate verifications in the middle of a periodic computing pattern, which ends with a guaranteed verification right before each checkpoint. Introducing partial verifications dramatically complicates the analysis, but we are able to analytically determine the optimal computing pattern (up to the first-order approximation), including the optimal length of the pattern, the optimal number of partial verifications, as well as their optimal positions inside the pattern. Performance evaluations based on a wide range of parameters confirm the benefit of using partial verifications under certain scenarios, when compared to the baseline algorithm that uses only guaranteed verifications.

This work has been published in the proceedings of ICPP'15 [22].

## 7.4. Which Verification for Soft Error Detection?

**Participants:** Leonardo Bautista-Gomez [Argonne National Laboratory], Anne Benoit, Aurélien Cavelan, Saurabh K. Raina [Jaypee Institute of Information Technology], Yves Robert, Hongyang Sun.

This work is an extension of the work described in Section 7.4 to cope with imperfect verifications. Many methods are available to detect silent errors in high-performance computing (HPC) applications. Each comes with a given cost and recall (fraction of all errors that are actually detected). The main contribution of this work is to characterize the optimal computational pattern for an application: which detector(s) to use, how many detectors of each type to use, together with the length of the work segment that precedes each of them. We conduct a comprehensive complexity analysis of this optimization problem, showing NP-completeness and designing an FPTAS (Fully Polynomial-Time Approximation Scheme). On the practical side, we provide a greedy algorithm whose performance is shown to be close to the optimal for a realistic set of evaluation scenarios.

This work has been published in the proceedings of HiPC'15 [21].

## 7.5. Composing resilience techniques: ABFT, periodic and incremental checkpointing

**Participants:** George Bosilca [University of Tennessee, Knoxville], Aurélien Bouteiller [University of Tennessee, Knoxville], Thomas Hérault [University of Tennessee, Knoxville], Yves Robert, Jack Dongarra [University of Tennessee, Knoxville].

Algorithm Based Fault Tolerant (ABFT) approaches promise unparalleled scalability and performance in failure-prone environments. Thanks to recent advances in the understanding of the involved mechanisms, a growing number of important algorithms (including all widely used factorizations) have been proven ABFT-capable. In the context of larger applications, these algorithms provide a temporal section of the execution, where the data is protected by its own intrinsic properties, and can therefore be algorithmically recomputed without the need of checkpoints. However, while typical scientific applications spend a significant fraction of their execution time in library calls that can be ABFT-protected, they interleave sections that are difficult or even impossible to protect with ABFT. As a consequence, the only practical fault-tolerance approach for these applications is checkpoint/restart. In this work, we propose a model to investigate the efficiency of a composite protocol, that alternates between ABFT and checkpoint/restart for the effective protection of an iterative application composed of ABFT-aware and ABFT-unaware sections. We also consider an incremental checkpointing composite approach in which the algorithmic knowledge is leveraged by a novel optimal dynamic programming to compute checkpoint dates. We validate these models using a simulator. The model and simulator show that the composite approach drastically increases the performance delivered by an execution platform, especially at scale, by providing the means to increase the interval between checkpoints while simultaneously decreasing the volume of each checkpoint.

This work has been published in the International Journal of Networking and Computing [9].

## 7.6. Voltage Overscaling Algorithms for Energy-Efficient Workflow Computations With Timing Errors

**Participants:** Aurélien Cavelan, Yves Robert, Hongyang Sun, Frédéric Vivien.

We proposed a software-based approach using dynamic voltage overscaling to reduce the energy consumption of HPC applications. This technique aggressively lowers the supply voltage below nominal voltage, which introduces timing errors, and we used Algorithm-Based Fault-Tolerance (ABFT) to provide fault tolerance for matrix operations. We introduced a formal model, and we designed optimal polynomial-time solutions, to execute a linear chain of tasks. Evaluation results obtained for matrix multiplication demonstrated that our approach indeed leads to significant energy savings, compared to the standard algorithm that always operates at nominal voltage.

This work has been published in the proceedings of the 5th Workshop on Fault Tolerance for HPC at eXtreme Scale [23].

## 7.7. Approximation algorithms for energy, reliability and makespan optimization problems

**Participants:** Guillaume Aupy, Anne Benoit.

We consider the problem of scheduling an application on a parallel computational platform. The application is a particular task graph, either a linear chain of tasks, or a set of independent tasks. The platform is made of identical processors, whose speed can be dynamically modified. It is also subject to failures: if a processor is slowed down to decrease the energy consumption, it has a higher chance to fail. Therefore, the scheduling problem requires us to re-execute or replicate tasks (i.e., execute twice the same task, either on the same processor, or on two distinct processors), in order to increase the reliability. It is a tri-criteria problem: the goal is to minimize the energy consumption, while enforcing a bound on the total execution time (the makespan), and a constraint on the reliability of each task.

Our main contribution is to propose approximation algorithms for linear chains of tasks and independent tasks. For linear chains, we design a fully polynomial-time approximation scheme. However, we show that there exists no constant factor approximation algorithm for independent tasks, unless P=NP, and we propose in this case an approximation algorithm with a relaxation on the makespan constraint.

This work has been published in the Parallel Processing Letters [4].

## 7.8. Co-scheduling algorithms for high-throughput workload execution

**Participants:** Guillaume Aupy, Manu Shantharam [San Diego Supercomputer Center], Anne Benoit, Yves Robert, Padma Raghavan [Penn State University].

This work investigates co-scheduling algorithms for processing a set of parallel applications. Instead of executing each application one by one, using a maximum degree of parallelism for each of them, we aim at scheduling several applications concurrently. We partition the original application set into a series of packs, which are executed one by one. A pack comprises several applications, each of them with an assigned number of processors, with the constraint that the total number of processors assigned within a pack does not exceed the maximum number of available processors. The objective is to determine a partition into packs, and an assignment of processors to applications, that minimize the sum of the execution times of the packs.

We thoroughly study the complexity of this optimization problem, and propose several heuristics that exhibit very good performance on a variety of workloads, whose application execution times model profiles of parallel scientific codes. We show that co-scheduling leads to to faster workload completion time and to faster response times on average (hence increasing system throughput and saving energy), for significant benefits over traditional scheduling from both the user and system perspectives.

This work has been published in the Journal of Scheduling [6].

## 7.9. Scheduling the I/O of HPC Applications Under Congestion

**Participants:** Ana Gainaru [University of Illinois at Urbana Champaign], Guillaume Aupy, Anne Benoit, Franck Cappello, Yves Robert.

A significant percentage of the computing capacity of large-scale platforms is wasted due to interferences incurred by multiple applications that access a shared parallel file system concurrently. One solution to handling I/O bursts in large-scale HPC systems is to absorb them at an intermediate storage layer consisting of burst buffers. However, our analysis of the Argonne's Mira system shows that burst buffers cannot prevent congestion at all times. As a consequence, I/O performance is dramatically degraded, showing in some cases a decrease in I/O throughput of 67%.

In this work, we analyze the effects of interference on application I/O bandwidth, and propose several scheduling techniques to mitigate congestion. We focus on typical HPC applications, which have a periodic pattern consisting of some amount of computation followed by some volume of I/O to be transferred. We show through extensive experiments that our global I/O scheduler is able to reduce the effects of congestion, even on systems where burst buffers are used, and can increase the overall system throughput up to 56%. We also show that it outperforms current Mira I/O schedulers, even for non-periodic applications.

This work has been published in IPDPS'15 [26].

## 7.10. Scheduling trees of malleable tasks for sparse linear algebra

**Participants:** Abdou Guermouche [Univ. Bordeaux/Inria Bordeaux Sud-Ouest], Loris Marchal, Bertrand Simon, Oliver Sinnen [Univ. Auckland/New Zealand], Frédéric Vivien.

Scientific workloads are often described by directed acyclic task graphs. This is in particular the case for multifrontal factorization of sparse matrices —the focus of this work— whose task graph is structured as a tree of parallel tasks. Prasanna and Musicus [84], [85] advocated using the concept of *malleable* tasks to model parallel tasks involved in matrix computations. In this powerful model each task is processed on a time-varying number of processors. Following Prasanna and Musicus, we consider malleable tasks whose speedup is $p^\alpha$, where $p$ is the fractional share of processors on which a task executes, and $\alpha$ ($0 < \alpha \leq 1$) is a task-independent parameter. Firstly, we use actual experiments on multicore platforms to motivate the relevance of this model for our application. Then, we study the optimal time-minimizing allocation proposed by Prasanna and Musicus using optimal control theory. We greatly simplify their proofs by resorting only to pure scheduling arguments. Building on the insight gained thanks to these new proofs, we extend the study to distributed (homogeneous or heterogeneous) multicore platforms. We prove the NP-completeness of the corresponding scheduling problem, and we then propose some approximation algorithms [28].

In a second step, we studied a simplified speed-up function for malleable tasks, corresponding to perfect parallelism for a number of processors below a given threshold. The threshold depends on the task. We proved that scheduling independent chains of malleable tasks under this model is NP-complete. We study the performance of a classical allocation policy which is agnostic of the threshold and a simple greedy heuristic, and proved that both are 2-approximation algorithms, even if in practice, the latter often outperforms the former.

## 7.11. Parallel scheduling of task trees with limited memory

**Participants:** Clément Brasseur [ENS Lyon], Guillaume Aupy, Loris Marchal.

Scientific workloads are often described by directed acyclic task graphs. This is in particular the case for multifrontal factorization of sparse matrices —the focus of this work— whose task graph is structured as a tree of parallel tasks. When processing this tree on a multicore machine, we have to find a tradeoff between task parallelism and memory usage. In this context, Agullo et al. [62] proposed an activation scheme which follows a postorder traversal and books the memory needed for the task. This strategy has a low complexity and thus has been implemented in the lightweight runtime system StarPU [65], but may lead to excessive memory booking, which limits the task parallelism. In this work, we proposed a new booking strategy that books exactly what is necessary for a task, given what is already booked by its predecessors in the tree. We have shown by extensive simulations on realistic trees that this leads to better task parallelism and reduces the overall processing time.

## 7.12. Locality of Map tasks in MapReduce computations

**Participants:**  Olivier Beaumont [Inria Bordeaux Sud-Ouest], Loris Marchal.

In data parallel system such as MapReduce, large data files are distributed among the storage attached to computing nodes, and the computation is afterwards allocated close to the data whenever it is possible. Several parameters may affect the locality of the data, and thus the amount of data that needs to be communicated during the computation: the possible replication of the data when it is distributed on the platform, and the load-balancing mechanism that transmits new data to node which have exhausted their own data. In this work, we have proposed a simple analytical model to estimate the amount of data transfer of various scenarios for the Map phase of MapReduce computations and we have validated this model using simulations.

## 7.13. Improving multifrontal methods by means of block low-rank representations

**Participants:**  Patrick Amestoy [INPT-IRIT, Université of Toulouse], Cleve Ashcraft [LSTC], Olivier Boiteau [EDF], Alfredo Buttari [CNRS-IRIT, Université of Toulouse], Jean-Yves L'Excellent, Clément Weisbecker [INPT-IRIT, now at LSTC].

Matrices coming from elliptic Partial Differential Equations (PDEs) have been shown to have a low-rank property: well defined off-diagonal blocks of their Schur complements can be approximated by low-rank products. Given a suitable ordering of the matrix which gives the blocks a geometrical meaning, such approximations can be computed using an SVD or a rank-revealing QR factorization. The resulting representation offers a substantial reduction of the memory requirement and gives efficient ways to perform many of the basic dense linear algebra operations.

Several strategies, mostly based on hierarchical formats, have been proposed to exploit this property. We study a simple, non-hierarchical, low-rank format called Block Low-Rank (BLR), and explain how it can be used to reduce the memory footprint and the complexity of sparse direct solvers based on the multifrontal method. We present experimental results on matrices coming from elliptic PDEs and from various other applications. We show that even if BLR based factorizations are asymptotically less efficient than hierarchical approaches, they still deliver considerable gains. The BLR format is compatible with numerical pivoting, and its simplicity and flexibility make it easy to use in the context of a general purpose, algebraic solver. This work has been published in the SIAM Journal on Scientific Computing [2].

## 7.14. Parallel Computation of a subset of entries of the inverse

**Participants:**  Patrick Amestoy [INPT-IRIT, Université of Toulouse], Iain Duff [RAL and CERFACS], Jean-Yves L'Excellent, François-Henry Rouet.

We consider the computation in parallel of several entries of the inverse of a large sparse matrix. We assume that the matrix has already been factorized by a direct method and that the factors are distributed. Entries are efficiently computed by exploiting sparsity of the right-hand sides and the solution vectors in the triangular solution phase. We demonstrate that in this setting, parallelism and computational efficiency are two contrasting objectives. We develop an efficient approach and show its efficiency on a general purpose parallel multifrontal solver. This work has been published in the SIAM Journal on Scientific Computing [3].

## 7.15. Efficient 3D frequency-domain seismic modeling with a parallel block low-rank (BLR) direct solver

**Participants:** Patrick Amestoy [INPT-IRIT, University of Toulouse], Romain Brossier [ISTerre, University of Grenoble-Alpes], Alfredo Buttari [CNRS-IRIT, University of Toulouse], Jean-Yves L'Excellent, Théo Mary [UPS-IRIT, University of Toulouse], Ludovic Métivier [ISTerre-JK-CNRS], Alain Miniussi [Geoazur-CNRS-UNSA], Stéphane Operto [Geoazur-CNRS-UNSA], Alessandra Ribodetti [Geoazur-CNRS-UNSA], Jean Virieux [ISTerre-UJF, University of Grenoble-Alpes], Clément Weisbecker [INPT-IRIT, now at LSTC].

Three-dimensional frequency-domain full waveform inversion (FWI) of fixed-spread data can be efficiently performed in the visco-acoustic approximation when seismic modeling is based on a sparse direct solver. Based on the work in [3] and its extension to a parallel environment, we studied the application of a parallel algebraic Block Low-Rank (BLR) multifrontal solver providing an approximate solution of the time-harmonic wave equation with a reduced operation count, memory demand, and volume of communication relative to the full-rank solver. We analyzed the parallel efficiency and the accuracy of the solver with a realistic FWI case [19]. The application of this parallel BLR solver to a real data case from the North Sea for full waveform inversion of ocean-bottom cable data was also presented in [18], where a multiscale frequency-domain FWI is applied by successive inversions of 11 discrete frequencies in the 3.5Hz-10Hz frequency band. The velocity model built by FWI reveals short-scale features such as channels, scrapes left by drifting icebergs, fractures and deep reflectors below the reservoir level, alhough the presence of gas in the overburden. The quality of the FWI results is controlled by time-domain modeling and source wavelet estimation. This work was done in the context of an on-going collaboration with the Seiscope consortium (https://seiscope2.obs.ujf-grenoble.fr/?lang=en?).

## 7.16. Approximation algorithms for bipartite matching on multicore architectures

**Participants:** Fanny Dufossé [DOLPHIN/Inria Lille - Nord Europe], Kamer Kaya [BMI, The Ohio State Univ., USA], Bora Uçar.

We proposed [13] two heuristics for the bipartite matching problem that are amenable to shared-memory parallelization. The first heuristic is very intriguing from a parallelization perspective. It has no significant algorithmic synchronization overhead and no conflict resolution is needed across threads. We showed that this heuristic has an approximation ratio of around 0.632 under some common conditions. The second heuristic was designed to obtain a larger matching by employing the well-known Karp-Sipser heuristic on a judiciously chosen subgraph of the original graph. We showed that the Karp-Sipser heuristic always finds a maximum cardinality matching in the chosen subgraph. Although the Karp-Sipser heuristic is hard to parallelize for general graphs, we exploited the structure of the selected subgraphs to propose a specialized implementation which demonstrates very good scalability. We proved that this second heuristic has an approximation guarantee of around 0.866 under the same conditions as in the first algorithm. We discussed parallel implementations of the proposed heuristics on a multicore architecture. Experimental results, for demonstrating speed-ups and verifying the theoretical results in practice, were also provided.

## 7.17. Hypergraph partitioning for multiple communication cost metrics

**Participants:** Mehmet Deveci [BMI, The Ohio State Univ., USA], Kamer Kaya [BMI, The Ohio State Univ., USA], Umit V. Çatalyürek [BMI, The Ohio State Univ., USA], Bora Uçar.

We investigated [12] hypergraph partitioning-based methods for efficient parallelization of communicating tasks. A good partitioning method should divide the load among the processors as evenly as possible and minimize the inter-processor communication overhead. The total communication volume is the most popular communication overhead metric which is reduced by the existing state-of-the-art hypergraph partitioners. However, other metrics such as the total number of messages, the maximum amount of data transferred by a processor, or a combination of them are equally, if not more, important. Existing hypergraph-based

solutions use a two phase approach to minimize such metrics where in each phase, they minimize a different metric, sometimes at the expense of others. We proposed a one-phase approach where all the communication cost metrics can be effectively minimized in a multi-objective setting and reductions can be achieved for all metrics together. For an accurate modeling of the maximum volume and the number of messages sent and received by a processor, we proposed the use of directed hypergraphs. The directions on hyperedges necessitate revisiting the standard partitioning heuristics. We did so and proposed a multi-objective, multi-level hypergraph partitioner. The partitioner takes various prioritized communication metrics into account, and optimizes all of them together in the same phase. Compared to the state-of-the-art methods which only minimize the total communication volume, we showed on a large number of problem instances that the new method produced better partitions in terms of several communication metrics.

## 7.18. Comments on the hierarchically structured bin packing problem

**Participants:** Thomas Lambert [Inria Bordeaux Sud-Ouest], Loris Marchal, Bora Uçar.

We studied [16] the hierarchically structured bin packing problem. In this problem, the items to be packed into bins are at the leaves of a tree. The objective of the packing is to minimize the total number of bins into which the descendants of an internal node are packed, summed over all internal nodes. We investigated an existing algorithm and made a correction to the analysis of its approximation ratio. Further results regarding the structure of an optimal solution and a strengthened inapproximability result were given.

## 7.19. Semi-two-dimensional partitioning for parallel sparse matrix-vector multiplication

**Participants:** Enver Kayaaslan, Cevdet Aykanat [Bilkent Univ., Turkey], Bora Uçar.

We proposed [31] a novel sparse matrix partitioning scheme, called semi-two-dimensional (s2D), for efficient parallelization of sparse matrix-vector multiply (SpMV) operations on distributed memory systems. In s2D, matrix nonzeros are more flexibly distributed among processors than one dimensional (rowwise or columnwise) partitioning schemes. Yet, there is a constraint which renders s2D less flexible than two-dimensional (nonzero based) partitioning schemes. The constraint is enforced to confine all communication operations in a single phase, as in 1D partition, in a parallel SpMV operation. In a positive view, s2D thus can be seen as being close to 2D partitions in terms of flexibility, and being close to 1D partitions in terms of computation/communication organization. We described two methods that take partitions on the input and output vectors of SpMV and produce s2D partitions while reducing the total communication volume. The first method obtains optimal total communication volume, while the second one heuristically reduces this quantity and takes computational load balance into account. We demonstrated that the proposed partitioning method improves the performance of parallel SpMV operations both in theory and practice with respect to 1D and 2D partitionings.

## 7.20. Combining backward and forward recovery to cope with silent errors in iterative solvers

**Participants:** Massimiliano Fasi [Univ Manchester, UK], Julien Langou [Univ. Colorado Denver, USA], Yves Robert, Bora Uçar.

We proposed combining checkpointing and verification for coping with silent errors in iterative solvers. We used algorithm based fault tolerance for error detection and error correction, allowing a forward recovery (and no rollback nor re-execution) when a single error is detected. We introduced an abstract performance model to compute the performance of all schemes, and we instantiated it using the Conjugate Gradient (CG) algorithm. Finally, we validate our new approach through a set of simulations both in normal and preconditioned CG [48], [25], [47].

## 7.21. Load-balanced local time stepping for large-scale wave propagation

**Participants:** Max Rietmann [Univ. Lugano, CH], Daniel Peter [Univ. Lugano, CH], Olaf Schenk [Univ. Lugano, CH], Bora Uçar, Marcus J. Grote [Univ. Basel, CH].

In complex acoustic or elastic media, finite element meshes often require regions of refinement to honor external or internal topography, or small-scale features. These localized smaller elements create a bottleneck for explicit time-stepping schemes due to the Courant-Friedrichs-Lewy stability condition. Recently developed local time stepping (LTS) algorithms reduce the impact of these small elements by locally adapting the time-step size to the size of the element. The recursive, multi-level nature of our LTS scheme introduces an additional challenge, as standard partitioning schemes create a strong load imbalance across processors. We examined [33] the use of multi-constraint graph and hypergraph partitioning tools to achieve effective, load-balanced parallelization. We implemented LTS-Newmark in the seismology code SPECFEM3D and compared performance and scalability between different partitioning tools on CPU and GPU clusters using examples from computational seismology.

## 7.22. Fast and high quality topology-aware task mapping

**Participants:** Mehmet Deveci [BMI, The Ohio State Univ., USA], Kamer Kaya [BMI, The Ohio State Univ., USA], Umit V. Çatalyürek [BMI, The Ohio State Univ., USA], Bora Uçar.

Considering the large number of processors and the size of the interconnection networks on exascale-capable supercomputers, mapping concurrently executable and communicating tasks of an application is a complex problem that needs to be dealt with care. For parallel applications, the communication overhead can be a significant bottleneck on scalability. Topology-aware task-mapping methods that map the tasks to the processors (i.e., cores) by exploiting the underlying network information are very effective to avoid, or at worst bend, this limitation. We proposed [24] novel, efficient, and effective task mapping algorithms employing a graph model. The experiments showed that the methods are faster than the existing approaches proposed for the same task, and on 4096 processors, the algorithms improved the communication hops and link contentions by 16% and 32%, respectively, on the average. In addition, they improved the average execution time of a parallel SpMV kernel and a communication-only application by 9% and 14%, respectively.

## 7.23. Distributed memory tensor computations

**Participants:** Oguz Kaya, Bora Uçar.

There are two prominent tensor decomposition formulations. CANDECOMP/PARAFAC (CP) formulation approximates a tensor as a sum of rank-one tensors. *Tucker* formulation approximates a tensor with a core tensor multiplied by a matrix along each mode. Both of these formulations have uses in applications. The most common algorithms for both decompositions are based on the alternating least squares method. The algorithms of this type are iterative, where the computational core of an iteration is a special operation operation between an $N$-mode tensor and $N$ matrices. These key operations are called the matricized tensor times Khatri-Rao product (MTTKRP) in the CP-ALS case, and the $n$-mode product in the Tucker decomposition case. We have investigated efficient parallelizations of full fledged algorithms for obtaining these two decompositions in distributed memory systems [30], [51] with a special focus on the mentioned key operations. In both studies, hypergraphs are used for computational load balancing and communication cost reduction. We are currently finalizing our last touches on the Tucker decomposition algorithms [51] to submit it to a conference. We are also working towards a unified view of the parallelization of the two algorithms. This work with its whole extend is carried out in the context of the thesis of Oguz Kaya.

## 7.24. Bridging the gap between performance and bounds of Cholesky factorization on heterogeneous platforms

**Participants:** Emmanuel Agullo [Inria Bordeaux Sud-Ouest], Olivier Beaumont [Inria Bordeaux Sud-Ouest], Lionel Eyraud-Dubois, Julien Herrmann, Suraj Kumar [Inria Bordeaux Sud-Ouest], Loris Marchal, Samuel Thibault [Inria Bordeaux Sud-Ouest].

In this work, we consider the problem of allocating and scheduling dense linear application on fully heterogeneous platforms made of CPUs and GPUs. More specifically, we focus on the Cholesky factorization since it exhibits the main features of such problems. Indeed, the relative performance of CPU and GPU highly depends on the sub-routine: GPUs are for instance much more efficient to process regular kernels such as matrix-matrix multiplications rather than more irregular kernels such as matrix factorization. In this context, one solution consists in relying on dynamic scheduling and resource allocation mechanisms such as the ones provided by PaRSEC or StarPU. We analyze the performance of dynamic schedulers based on both actual executions and simulations, and we investigate how adding static rules based on an offline analysis of the problem to their decision process can indeed improve their performance, up to reaching some improved theoretical performance bounds which we introduce [17].

## 7.25. Assessing the cost of redistribution followed by a computational kernel: Complexity and performance results

**Participants:** Julien Herrmann, George Bosilca [University of Tennessee, Knoxville], Thomas Hérault [University of Tennessee, Knoxville], Loris Marchal, Yves Robert, Jack Dongarra [University of Tennessee, Knoxville].

The classical redistribution problem aims at optimally scheduling communications when reshuffling from an initial data distribution to a target data distribution. This target data distribution is usually chosen to optimize some objective for the algorithmic kernel under study (good computational balance or low communication volume or cost), and therefore to provide high efficiency for that kernel. However, the choice of a distribution minimizing the target objective is not unique. This leads to generalizing the redistribution problem as follows: find a re-mapping of data items onto processors such that the data redistribution cost is minimal, and the operation remains as efficient. This work studies the complexity of this generalized problem. We compute optimal solutions and evaluate, through simulations, their gain over classical redistribution. We also show the NP-hardness of the problem to find the optimal data partition and processor permutation (defined by new subsets) that minimize the cost of redistribution followed by a simple computational kernel. Finally, experimental validation of the new redistribution algorithms are conducted on a multicore cluster, for both a 1D-stencil kernel and a more compute-intensive dense linear algebra routine.

This work has been published in the Parallel Computing journal [15].

## 7.26. STS-k: A Multi-level Sparse Triangular Solution Scheme for NUMA Multicores

**Participants:** Humayun Kabir [Penn State University], Joshua Booth [Sandia National Laboratories], Guillaume Aupy, Anne Benoit, Yves Robert, Padma Raghavan [Penn State University].

We consider techniques to improve the performance of parallel sparse triangular solution on non-uniform memory architecture multicores by extending earlier coloring and level set schemes for single-core multiprocessors. We develop STS-k, where k represents a small number of transformations for latency reduction from increased spatial and temporal locality of data accesses. We propose a graph model of data reuse to inform the development of STS-k and to prove that computing an optimal cost schedule is NP-complete. We observe significant speed-ups with STS-3 on 32-core Intel Westmere-EX and 24-core AMD 'MagnyCours' processors. Incremental gains solely from the 3-level transformations in STS-3 for a fixed ordering, correspond to reductions in execution times by factors of 1.4 (Intel) and 1.5 (AMD) for level sets and 2 (Intel) and 2.2 (AMD) for coloring. On average, execution times are reduced by a factor of 6 (Intel) and 4 (AMD) for STS-3 with coloring compared to a reference implementation using level sets.

This work has been published in SC'15 [29].

## 7.27. Mono-parametric Tiling

**Participants:** Guillaume Iooss [Inria/ENS-Lyon/UCBL/CNRS], Sanjay Rajopadhye [Colorado State University], Christophe Alias, Yun Zou [Colorado State University].

Tiling is a crucial program transformation with many benefits: it improves locality, exposes parallelism, allows for adjusting the ops-to-bytes balance of codes, and can be applied at multiple levels. Allowing tile sizes to be symbolic parameters at compile time has many benefits, including efficient autotuning, and run-time adaptability to system variations. For polyhedral programs, parametric tiling in its full generality is known to be non-linear, breaking the mathematical closure properties of the polyhedral model. Most compilation tools therefore either avoid it by only performing fixed size tiling, or apply it in only the final, code generation step. Both strategies have limitations.

We first introduce mono-parametric partitioning, a restricted parametric, tiling-like transformation which can be used to express a tiling. We show that, despite being parametric, it is a polyhedral transformation. We first prove that applying mono-parametric partitioning (i) to a polyhedron yields a union of polyhedra, and (ii) to an affine function produces a piecewise-affine function. We then use these properties to show how to partition an entire polyhedral program, including one with reductions. Next, we generalize this transformation to tiles with arbitrary tile shapes that can tesselate the iteration space (e.g., hexagonal, trapezoidal, etc). We show how mono-parametric tiling can be applied at multiple levels, and enables a wide range of polyhedral analyses and transformations to be applied.

This work has been published as an Inria research report [49] and will be submitted to a journal.

## 7.28. Data-aware Process Networks

**Participants:**  Christophe Alias, Alexandru Plesco [XtremLogic SAS].

High-level circuit synthesis (HLS, high-level synthesis) consists in compiling a C-like high-level program to a circuit. The circuit must be as efficient as possible while using properly the resources (energy, memory, FPGA building blocks, etc). Thought many progresses were achieved on the low aspects of circuit generation (pipeline, place/route), the front-end aspects (parallelism, communications) are still rudimentary compared to the state-of-the-art techniques in the HPC community.

We introduce the Data-aware Process Networks (DPN), a new parallel execution model adapted to the hardware constraints of high-level synthesis, where the data transfers are made explicit. We show that the DPN model is consistant in the meaning where any translation of a sequential program produces an equivalent DPN without deadlocks. Finally, we show how to compile a sequential program to a DPN and how to optimize the input/output and the parallelism.

This work was published as an Inria research report [63] and will be submitted to a journal.

## 7.29. Termination of C programs

**Participants:**  Laure Gonnord, David Monniaux [CNRS/VERIMAG], Gabriel Radanne [Univ Paris 7/ PPS].

We designed a complete method for synthesizing lexicographic linear ranking functions (and thus proving termination), supported by inductive invariants, in the case where the transition relation of the program includes disjunctions and existentials (large block encoding of control flow).

Previous work would either synthesize a ranking function at every basic block head, not just loop headers, which reduces the scope of programs that may be proved to be terminating, or expand large block transitions including tests into (exponentially many) elementary transitions, prior to computing the ranking function, resulting in a very large global constraint system. In contrast, our algorithm incrementally refines a global linear constraint system according to extremal counterexamples: only constraints that exclude spurious solutions are included.

Experiments with our tool Termite 6.5  show marked performance and scalability improvements compared to other systems.

This work has been published in the proceedings of PLDI'15 [27].

## 7.30. Analysing C programs with arrays

**Participants:**  Laure Gonnord, David Monniaux [CNRS/VERIMAG].

Automatically verifying safety properties of programs is hard, and it is even harder if the program acts upon arrays or other forms of maps. Many approaches exist for verifying programs operating upon 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 contrast to most preceding approaches, we do not introduce a new abstract domain or a new interpolation procedure for arrays. Instead, we generate an abstraction as a scalar problem and feed it to a preexisting solver. The intuition is that if there is a proof of safety of the program, it is likely that it can be expressed by elementary steps between properties involving only a small (tunable) number $N$ of cells from the array.

Our transformed problem is expressed using Horn clauses over scalar variables, a common format with clear and unambiguous logical semantics, for which there exist several solvers. In contrast, solvers directly operating over Horn clauses with arrays are still very immature.

An important characteristic of our encoding is that it creates a nonlinear Horn problem, with tree unfoldings, contrary to the linear problems obtained by flatly encoding the control-graph structure. Our encoding thus cannot be expressed by encoding into another control-flow graph problem, and truly leverages the Horn clause format.

Experiments with our prototype VAPHOR show that this approach can prove automatically the functional correctness of several classical examples of the literature, including *selection sort*, *bubble sort*, *insertion sort*, as well as examples from previous articles on array analysis.

This work has been published as a research report [53] and is currently under submission.

## 7.31. Symbolic Range Analysis of Pointers in C programs

**Participants:** Maroua Maalej, Vitor Paisante [Univ. Mineas Gerais, Brasil], Laure Gonnord, Fernando Pereira [Univ. Mineas Gerais, Brasil], Vitor Paisante [Univ. Mineas Gerais, Brasil].

Alias analysis is one of the most fundamental techniques that compilers use to optimize languages with pointers. However, in spite of all the attention that this topic has received, the current state-of-the-art approaches inside compilers still face challenges regarding precision and speed. In particular, pointer arithmetic, a key feature in C and C++, is yet to be handled satisfactorily. We designed a new alias analysis algorithm to solve this problem. The key insight of our approach is to combine alias analysis with symbolic range analysis. This combination lets us disambiguate fields within arrays and structs, effectively achieving more precision than traditional algorithms. To validate our technique, we have implemented it on top of the LLVM compiler. Tests on a vast suite of benchmarks show that we can disambiguate several kinds of C idioms that current state-of-the-art analyses cannot deal with. In particular, we can disambiguate 1.35x more queries than the alias analysis currently available in LLVM. Furthermore, our analysis is very fast: we can go over one million assembly instructions in 10 seconds.

This work has been published at CGO'16 [32].

An extended version of the related work has also been published as an Inria research report [52] and will be the basis of a journal submission.

<p align="center" style="color:red"><b>SOCRATE Project-Team</b></p>

# 6. New Results

## 6.1. Flexible Radio Front-End

The contributions of members of this axis are mainly on four topics: Wake-Up Radio,Full-Duplex transceivers, SDR Gateways for Urban Networks, and Channel Estimation. In the global concept of enhancing wireless communications, those four topics are complimentary, addressing the reduction of energy consumption, the increase of throughput and/or flexibility of the transmission and the performance evaluation.

### 6.1.1. *Wake-Up Radio*

The last decades have been really hungry in new ways to reduce energy consumption. That is especially true when talking about wireless sensor networks in general and home multimedia networks in particular, since electrical energy consumption is the bottleneck of the network. One of the most energy-consuming functional block of an equipment is the radio front end, and methods to switch it off during the time intervals where it is not active must be implemented. This study [10] has proposed a wake-up radio circuit which is capable of both addressing and waking up not only a more efficient but also more energy-consuming radio front end. By using a frequency footprint to differentiate each sensor, awaking all the sensors except for the one of interest is avoided. The particularity of the proposed wake-up receiver is that the decision is taken in the radio-frequency part and no baseband treatment is needed. The global evaluation in theory and in simulation was performed, and a first testbed of this technology was fabricated.

#### 6.1.1.1. Full-Duplex

This work studies [8] a Full-Duplex Dual-Band (FDDB) OFDM radio architecture that enables the radio transceiver to be more flexible and provides a viable radio link capacity gain. A simple but practical I/Q imbalance estimation and compensation method, based on the frequency-flat-fading behavior of the self-interference channel, is proposed. The performance of the proposed I/Q imbalance compensation method is evaluated by system level simulations conducted with ADS and Matlab. The co-simulation results show that the proposed radio transceiver could potentially increase the physical layer transmission rate by four times compared to the conventional radio link at the cost of tolerable loss of BER performance. The I/Q imbalance compensation method can effectively compensate both high and low I/Q imbalance without the problem of algorithm convergence. Application of this technique for physical layer security has already been proposed.

#### 6.1.1.2. SDR for SRDs

The technologies employed in urban sensor networks are permanently evolving, and thus the gateways of these networks have to be regularly upgraded. The existing method to do so is to stack-up receivers dedicated to one communication protocol. However, this implies to have to replace the gateway every time a new protocol is added to the network. A more practical way to do this is to perform a digitization of the full band and to perform digitally the signal processing, as done in Software-Defined Radio (SDR). The main hard point in doing this is the dynamic range of the signals: indeed the signals are emitted with very different features because of the various propagation conditions. It has been proved that the difference of power between two signals can be so important that no existing Analog-to-Digital Converter (ADC) is able to properly digitize the signals. We propose a solution to reduce the dynamic range of signals before digital conversion. In this study [28], the assumption is made that there is one strong signal, and several weak signals. This assumption is made from the existing urban sensor networks topology. A receiver architecture with two branches is proposed with a "Coarse Digitization Path" (CDP) and a "Fine Digitization Path" (FDP). The CDP allows to digitize the strong signal and to get data on it that is used to reconfigure the FDP. The FDP then uses a notch filter to attenuate the strong signal (and then to reduce the dynamic range of the signals) and digitizes the rest of the band. Another way to relax these specifications on ADCs is an analog processing, such as companding, that should be performed before digitization. The companding technique is usually employed on one signal (and

not on multiple signals that are only separated on the frequency domain). This work [36], [29] studies three companding laws to test their efficiency in relaxing the digitization constraints with multiple signals. A $\mu$-law, a Piecewise-Linear (PL) law and a Piecewise-Linear, Constant Gain with Offsets (PLCGO) law are tested. We have described how to use a PLCGO approach to reduce ADC's complexity, and two implementations of the compressing law are proposed.

*6.1.1.3. Channel Estimation*

In modern mobile telecommunications, shadow fading has to be modeled by a two-dimensional (2D) correlated random variable since shadow fading may present both cross-correlation and spatial correlation due to the presence of similar obstacles during the propagation. In our study, 2D correlated random shadowing is generated based on the multi-resolution frequency domain ParFlow (MR-FDPF) model. The MR-FDPF model is a 2D deterministic radio propagation model, so a 2D deterministic shadowing can be firstly extracted from it. Then, a 2D correlated random shadowing can be generated by considering the extracted 2D deterministic shadowing to be a realization of it. Moreover, based on the generated 2D correlated random shadowing, a complete 2D semi-deterministic path loss model can be proposed. The proposed methodology [5] can be implemented into system-level simulators where it will be very useful due to its ability to generate realistic shadow fading.

[23] presents the first implementation on software defined radio nodes in the large scale testbed CorteXlab of a radio link estimation technique based on OFDM transmissions. The purpose of this large scale testbed is to offer to the whole scientific community an open tool to test new techniques for multiuser , cooperative and cognitive radio networks in a controlled environment. As the experimentation room was defined in order to offer reproducible measurements, it is important to be able to characterize each radio link between all transceivers. Therefore, we present here the development of a channel sounder directly implemented on the software radio nodes. This paper presents the first implementation on software defined radio nodes in the large scale testbed called CorteXlab of a radio link estimation technique based on OFDM transmissions. The purpose of this large scale testbed is to offer to the whole scientific community an open tool to test new techniques for multiuser , cooperative and cognitive radio networks in a controlled environment. As the experimentation room was defined in order to offer reproducible measurements, it is important to be able to characterize each radio link between all transceivers. Therefore, we proposed the development of a channel sounder directly implemented on the software radio nodes.

# 6.2. Agile Radio Resource Sharing

This axis addresses the challenges relative to the network perspective of software radio. While the two other axes have their focus on the design of the software radio nodes, axis 2 deals with coexistence and cooperation in a multi-user communications perspective.

A first research direction concerns theoretical limits of different reference scenarios where trade-offs between spectral efficiency, energy efficiency, stability and/or fairness are analyzed. This work exploits multi-users information theory, game theory and stochastic geometry. This year, a particular focus has been put on the interference channel with feedback and on dense wireless networks. New problems have been also investigated with the simultaneous energy and information transmission problem for energy harvesting and some specific attacks in smart grids.

In parallel our research activities are also driven by applicative frameworks. Concerning 4G RAN, a new interference alignment scheme has been proposed, simulated and implemented on CortexLab. This work has been presented as one of the promising technologies proposed by Greentouch. IoT has been identified as a new challenge for 5G with the objective of serving a very large number of nodes per cell, in a connectionless manner and with very small packets. The original transmission technology using ultra narrow band modulation and proposed by Sigfox for large area of IoT nodes has been investigated. A multiband CSMA strategy has been also evaluated in collaboration with CEA-Leti for dense Wifi like IoT access networks. Body area networks (BANs) represent also a very challenging applicative framework, with strong dynamics, interference environments, and low energy requirements. In partnership with Euromedia and Hikob, our studies focused on

dynamic algorithms for information gathering in a sport event broadcast system. Additionnally, localization capabilities at the body scale may offer interesting perspetives but require specific MAC protocols.

## 6.2.1. Fundamental Limits

*6.2.1.1. Energy efficiency - Spectral Efficiency (EE-SE) Tradeoffs in Wireless RANs*

The spectral and energy efficiency (SE-EE) trade-off in cellular networks has attracted significant recent interest in the wireless community [1]. The work in [7] studies this fundamental limit with a simple and effective method. The proposed theoretical framework is based on an optimal radio resource allocation of transmit power and bandwidth for the downlink direction, applicable for an orthogonal cellular network. The analysis is initially focused on a single cell scenario, for which in addition to the solution of the main SE-EE optimization problem, it is proved that a traffic repartition scheme can also be adopted as a way to simplify this approach. By exploiting this interesting result along with properties of stochastic geometry, this work is extended to a more challenging multi-cell environment, where interference is shown to play an essential role and for this reason several interference reduction techniques are investigated. Special attention is also given to the case of low signal to noise ratio (SNR) and a way to evaluate the upper bound of EE in this regime is provided. This methodology leads to tractable analytical results under certain common channel properties, and thus allows the study of various models without the need for demanding system level simulations.

*6.2.1.2. Interference Channels with Feedback*

The capacity region of the two-user linear deterministic (LD) interference channel with noisy output feedback (IC-NOF) is fully characterized in [35], [26]. This result allows the identification of several asymmetric scenarios in which implementing channel-output feedback in only one of the transmitter-receiver pairs is as beneficial as implementing it in both links, in terms of achievable individual rate and sum-rate improvements w.r.t. the case without feedback. In other scenarios, the use of channel-output feedback in any of the transmitter-receiver pairs benefits only one of the two pairs in terms of achievable individual rate improvements or simply, it turns out to be useless, i.e., the capacity regions with and without feedback turn out to be identical even in the full absence of noise in the feedback links. As a byproduct, the exact conditions on the signal to noise ratios on the feedback links to observe an improvement on either a single rate, both single rates, or the sum-rate capacity, for any IC-NOF are also fully described in [41].

*6.2.1.3. Simultaneous Energy and Information Transmission*

The fundamental limits of simultaneous information and energy transmission in the two-user Gaussian multiple access channel (G-MAC) with and without feedback are fully characterized in [33], [9]. All the achievable information and energy transmission rates (in bits per channel use and energy-units per channel use respectively) are identified. Thus, the information-energy capacity region is defined in both cases. In the case without feedback, an achievability scheme based on power-splitting and successive interference cancelation is shown to be optimal. Alternatively, in the case with feedback (G-MAC-F), a simple yet optimal achievability scheme based on power-splitting and Ozarow's capacity achieving scheme is presented. Three of the most important observations in this work are: (a) The capacity-energy region of the G-MAC without feedback is a proper subset of the capacity-energy region of the G-MAC-F; (b) Feedback can at most double the energy rate for a fixed information rate; and (c) Time-sharing with power control is strictly suboptimal in terms of sum-rate in the G-MAC without feedback.

*6.2.1.4. Multiple Access Channel and Broadcast Channel with Linear Feedback Schemes*

In [11], it is shown that for the two-user Gaussian broadcast channel with correlated noises and perfect feedback the largest region that can be achieved by linear-feedback schemes equals the largest region that can be achieved over a dual multi-access channel when in this latter the channel inputs are subject to a "non-standard" sum-power constraint that depends on the BC-noise correlation. Combining this new duality result with Ozarow's MAC-scheme gives an elegant achievable region for the Gaussian BC with correlated noises. A constructive iterative coding scheme is then presented for the non-symmetric Gaussian BC with uncorrelated noises that is sum-rate optimal among all linear-feedback schemes. This coding scheme shows that the connection between the MAC and the BC optimal schemes is tighter than what is suggested by our duality result on achievable rates. In fact, it is linear-feedback sum-rate optimal to use Ozarow MAC-encoders and MAC-decoders— rearranged—to code over the BC.

### 6.2.2. Low Complexity Receivers for Massive MIMO Systems

In wireless communications, Multi-user massive MIMO network is a scenario that has been recently proposed, where many mobile terminals are served by a Base Station (BS) equipped with a very high number of antennas. In such a scenario, the detection in the uplink remains a challenge, since the BS is required to detect signals transmitted from all users while trying to exploit full received diversity. The optimal detection criterion that fulfills the diversity requirement is the Maximum-Likelihood (ML) joint detection which has been proposed to detect jointly the transmitted signals. However, such a criterion is not applicable to the addressed multi-user massive MIMO scenario due to its computational complexity that increases exponentially with the number of signals to be detected. In our work paper, we have proposed a relaxed ML detector based on an iterative decoding strategy that reduces the computational cost. We exploit the fact that the transmit constellation is discrete, and remodel the channel as a MIMO channel with sparse input belonging to the binary $\{0, 1\}$. The sparsity property allows us to relax the ML problem as a quadratic minimization under linear and $\ell$1-norm constraint. We then prove the equivalence of the relaxed problem to a convex optimization problem solvable in polynomial time. Simulation results illustrate the efficiency of the low-complexity proposed detector compared to other existing ones in very large and massive MIMO context.

### 6.2.3. Distributed Radio Resource Management

#### 6.2.3.1. Interference Alignment in Cellular Networks with no-Explicit Coordination

Current networks aim to support high data rates for end users by increasing the spectral efficiency in bits-per-Hertz, at the expense of the energy efficiency of the network. Indeed, an important part of the energy consumption of mobile networks is proportional to the radiated energy, which relies on the frequency bandwidth and the transmission power. Any energy efficient transmission scheme should exploits the whole system bandwidth by allocating the entire available spectrum to each base station. Such an approach, however, leads to significant interference increase and performance degradation for mobiles located at the cell edges. The key challenge is to balance interference avoidance and spectrum use to reach an optimal spectral efficiency – energy efficiency (EE-SE) trade-off. The work achieved in the framework of Greentouch collaboration is based on the non classical interference alignement scheme proposed by Suh and Tse in dowlink mode. The key contribution relies on users scheduling with a unique criteria based as well on QoS priorities and orthogonality of precoding directions. The spectral efficiency is improved by a factor 2 for edge users and a energy saving of about $30\%$ is made possible. This scheme has been evaluated on simulation scenarios as defined by Greentouch partners and a simplified version has been implemented on FIT/CorteXlab and demontrastrated during the final event of Greentouch (New-York, June 2015).

### 6.2.4. RANs for IoT : Dense and Connectionless Solutions

Internet of Things (IoT) is going to take a major place in the telecommunications market as announced in technical and public medias. The paradigm of IoT relies on the deployment of billions of objects having the capability of transmitting information about their context and environment and to create a real-time, secured and efficient interaction between the real and the virtual worlds, pushing them to evolve from the state of cousins to the state of Siamese twins. IoT revealed to be a key technology for solving societal issues such as digital cities, intelligent transportation, green environment monitoring or medical care and elderly person monitoring.

IoT has strong connections with machine-to-machine (M2M), and sometimes in literature, both terms refer to the same idea. From our point of view, IoT covers a broader scope including as well the technology and the applications. On the opposite, M2M refers to the technologies that allow machines or objects to communicate.

In any case, from the technical point of view, the main challenge of this new paradigm is to let a huge number of machine type devices (MTDs) be connected to the Internet at a low cost, with a limited infrastructure and featuring a very long life time with very small battery or energy needs [4].

In this global picture, we may consider different technical issues. M2M has first been defined to connect MTDs in their vicinity. The proposed solutions extensively rely on the research results produced over the last 20 years for ad-hoc and wireless sensor networks. Initiated 20 years ago from theoretical concepts, this very active research area has gone up to the definition of full standards (802.15.4, 802.15.6, Zigbee, Bluetooth) which have already found a market.

More recently, the IoT paradigm has been extended to the problem of connecting all these MTDs to the Internet, and through Internet to anyone or anything. The massive connection of objects spread over the world is a challenge that has some similarities with the paradigm of cellular networks which aimed at connecting people. This similarity attracted the interest of mobile network providers, to exploit such attractive potential market and IoT has been identified as a target for the future 5G.

### 6.2.4.1. Performance of Ultra-NarrowBand Techniques

The Ultra-narrow-band technology is an appealing solution for the low throughput wireless sensor networks (10b/s - 1 kb/s). It is complementary to the classical cellular networks thanks to its low energy consumption and very long range communication (up to 50 km in free-space) [4]. This technology has already been deployed and is proved to be ultra-efficient for point-to-point communications in Sigfoxs' network. Nodes are transmitting at a random time and random frequency carrier (random frequency division multiple access schemes : R-FDMA), so the uplink is exposed to interference. In our approach, we have proposed to model this interference for the UNB network when taking into account the path-loss and Rayleigh effects, with stochastic geometry tools. The obtained model allows us to estimate the system performance, and its capacity in terms of maximum number average of simultaneous nodes in a unique cell [37]. We have also considered the replication mechanism, and identified the optimum number of replications.

### 6.2.4.2. Multiband CSMA for Dense Wireless Networks in Uplink

In this approach, the objective is to mitigate the degradation of the throughput and delay performance in wireless local area networks (WLAN) that employ carrier sense multiple access collision avoidance (CSMA/CA) protocol with request to send and clear to send (RTS/CTS) mechanism, when a large number of IoT like nodes are deployed. In our approach, the overhead is reduced with a modified handshake mechanism. The medium access control (MAC) overhead caused by the RTS and CTS messages is high comparing to the total duration of successful transmission. In order to reduce the MAC overhead we propose in this work a new strategy to serve many users successively. This strategy consists on sending many RTS in parallel by different stations on different frequency sub-bands. Once the RTS messages do not collide with each other, there will be no need to resend the RTS and wait for a CTS to gain the channel access [21].

## 6.2.5. Algorithms and Protocols for BANs

### 6.2.5.1. Information Gathering in a Group of Mobile Users

Distributed decisions within any group of agents, is a very active research area and theoretical results as well as efficient algorithms have already been proposed but in the context of wireless networks, the task is made harder due to possible transmission errors, channel asymmetry, dynamic behaviour of the channel and node mobility. In this work, we consider a group of mobile agents moving roughly in a common direction. We study different algoritimic solutions allowing each agent to periodically discover its neighbours: one-hop neighbours as well as multi-hop neighbors. The reference scenario is a bike race, during which groups are susceptible to split or merge. The objective is a live gathering of information about who is present in a group for live TV broadcasting. For that, we need a fully distributed approach allowing every agent to discover with a consensus algorithm the list of neighbours participating to the same pack. This study may be of interest for various other applications such as group navigation support in crowded environments, autonomous navigation of a fleet of robots. . . This problem exhibits some similarities with a clustering problem. However, a clustering problem aims at exploiting the structure of a graph and to form some subgroups to ensure a good structure of the network for further communications while our objective is rather to estimate the groups naturally formed in the real world. Hence, we have focused on distributed decision algorithms, which are widely present in the literature. Max-consensus problem has been much less studied than average consensus. The proposed algorithms are based on the N-dimension generalization of the Random Broadcast Max-Consensus algorithm,

allowing each agent to build and share the list of its muli-hop neighbors. We extend this approach to a dynamic context where the group information needs to be updated according to possible group merge or split. Experimental validation has been done in the context of a cycling race with 10 agents, equipping each bicycle with a wireless sensor node to assess the interactions between the racers and to provide a live monitoring of the dynamic evolution of the cyclists groups that form during the race.

*6.2.5.2. MAC Protocols and Algorithms for Localization at the Body Scale*

The purpose of this work is to evaluate the impact of the node speed on the ranging estimation for location applications with Wireless Body Area Networks (WBAN). While estimated with the 3-Way ranging protocol (3-WR), this distance between two nodes placed on the body can be affected by the human movements [30], [17]. Thus, we study theoretically the ranging error with the 3-WR, based on a perfect channel, a MAC layer based on TDMA using two scheduling strategies (Single node localization (P2P-B) and Aggregated & Broadcast (A&B)) and a PHY layer based on Ultra Wideband (IR-UWB) [31]. We demonstrate the accuracy of the model, and show that the distance error is highly correlated with the speed of nodes [16], while the associated mobility model has an impact on the design of MAC strategies by simulation [18].

### 6.2.6. Other Topics

*6.2.6.1. Data Injection Attacks in Smart Grids*

Multiple attacker data injection attack construction in electricity grids with minimum-mean-square-error (MMSE) state estimation is studied for centralized and decentralized scenarios in [34]. A performance analysis of the trade-off between the maximum distortion that an attack can in- troduce and the probability of the attack being detected by the network operator is considered. Within this setting, optimal centralized attack construction strategies are studied. The decentralized case is examined in a game-theoretic setting. A novel utility function is proposed to model this trade-off and it is shown that the resulting game is a potential game. The existence and cardinality of the corresponding set of Nash Equilibria (NE) in the game is analyzed. For the particular case of two attackers, numerical results based on IEEE test systems are presented. These results suggest that attackers perform better when they seize control of power flow measurements instead of power injection measurements.

# 6.3. Software Radio Programming Model

## 6.3.1. Data Flow Programming

Streaming languages have been proven to be a natural and efficient approach for taking advantage of the intrinsic parallelism of modern CPU architectures. The focus of many previous work has been to improve the throughput of streaming programs. In [27], we rather focus on satisfying quality-of-service requirements of streaming applications executed alongside non-streaming processes. We monitor synchronous dataflow (SDF) programs at runtime both at the application and system levels, in order to identify violations of quality-of-service requirements. Our monitoring requires the programmer to provide the expected throughput of its application (e.g 25 frames per second for a video decoder), then takes full benefit from the compilation of the SDF graph to detect bottlenecks in this graph and identify causes among processor or memory overloading. It can then be used to perform dynamic adaptations of the applications in order to optimize the use of computing and memory resources.

## 6.3.2. Smart Sensors

The article [19] presents the development of a wireless wearable sensor for the continuous, long-term monitoring of cardiac activity. Heart rate assessment, as well as heart rate variability parameters are computed in real time directly on the sensor, thus only a few parameters are sent via wireless communication for power saving. Hardware and software methods for heart beat detection and variability calculation are described and preliminary tests for the evaluation of the sensor are presented. With an autonomy of 48 hours of active measurement and a Bluetooth Low Energy radio technology, this sensor will form a part of a wireless body network for the remote mobile monitoring of vital signals in clinical applications requiring automated collection of health data from multiple patients.

### 6.3.3. Cryptography

For security applications in wireless sensor networks (WSNs), choosing best algorithms in terms of energy-efficiency and of small memory requirements is a real challenge because the sensor networks are composed of low-power entities. Previous works benchmarked 12 block-ciphers on an ATMEL AVR ATtiny45 8-bit microcontroller. In [2], most of the recent lightweight block cipher proposals, as well as some conventional block ciphers, are studied on the Texas Instruments MSP430 16-bit microcontroller . The chosen block ciphers are described with a security and an implementation summary. Implementations are then evaluated on a dedicated platform.

### 6.3.4. Hardware Arithmetic

#### 6.3.4.1. Hardware Implementations of Fixed-Point Atan2

The atan2 function computes the polar angle arctan(x/y) of a point given by its cartesian coordinates. It is widely used in digital signal processing to recover the phase of a signal. The article [14] studies for this context the implementation of atan2 with fixed-point inputs and outputs. It compares the prevalent CORDIC shift-and-add algorithm to two multiplier-based techniques. The first one reduces the bivariate atan2 function to two functions of one variable: the reciprocal, and the arctangent. These two functions may be tabulated, or evaluated using bipartite or polynomial approximation methods. The second technique directly uses piecewise bivariate polynomial approximations, in degree 1 and degree 2. It requires larger tables but has the shortest latency. Each of these approaches requires a relevant argument reduction, which is also discussed. All the algorithms are described with the same accuracy target (faithful rounding) and implemented with similar care in FloPoCo. Based on synthesis results on FPGAs, their relevance domains are discussed.

#### 6.3.4.2. Fixed-Point Implementations of the Reciprocal, Square Root and Reciprocal Square Root Functions

Implementations of the reciprocal, square root and reciprocal square root often share a common structure. The article [39] is a survey and comparison of methods for computing these functions. It compares classical methods (direct tabulation, multipartite tables, piecewise polynomials, Taylor-based polynomials, Newton-Raphson iterations). It also studies methods that are novel in this context: the Halley method and, more generally, the Householder method. The comparisons are made in the context of the same accuracy target (faithful rounding) and of an arbitrary fixed-point format for the inputs and outputs (precisions of up to 32 bits). Some of the methods discussed might require some form of range reduction, depending on the input range. The objective of the article is to optimize the use of fixed-size FPGA resources (block multipliers and block RAMs). The discussions and conclusions are based on synthesis results for FPGAs.

#### 6.3.4.3. Fixed-Point Hardware Polynomials

Polynomial approximation is a general technique for the evaluation of numerical functions of one variable such as atan, reciprocal and square roots studied above. The article [38] addresses the automatic construction of fixed-point hardware polynomial evaluators. By systematically trying to balance the accuracy of all the steps that lead to an architecture, it simplifies and improves the previous body of work covering polynomial approximation, polynomial evaluation, and range reduction. This work is supported by an open-source implementation in FloPoCo.

### 6.3.5. Software Elementary Functions

#### 6.3.5.1. Code Generators for Mathematical Functions

A typical floating-point environment includes support for a small set of about 30 mathematical functions such as exponential, logarithms and trigonometric functions. These functions are provided by mathematical software libraries (libm), typically in IEEE754 single, double and quad precision. The article [13] suggests to replace this libm paradigm by a more general approach: the on-demand generation of numerical function code, on arbitrary domains and with arbitrary accuracies. First, such code generation opens up the libm function space available to programmers. It may capture a much wider set of functions, and may capture even standard functions on non-standard domains and accuracy/performance points. Second, writing libm code requires fine-tuned instruction selection and scheduling for performance, and sophisticated floating-point

techniques for accuracy. Automating this task through code generation improves confidence in the code while enabling better design space exploration, and therefore better time to market, even for the libm functions. This article discusses, with examples, the new challenges of this paradigm shift, and presents the current state of open-source function code generators.

*6.3.5.2. Computing Floating-Point Logarithms with Fixed-Point Operations*

Elementary functions from the mathematical library input and output floating-point numbers. However it is possible to implement them purely using integer/fixed-point arithmetic. This option was not attractive between 1985 and 2005, because mainstream processor hardware supported 64-bit floating-point, but only 32-bit integers. Besides, conversions between floating-point and integer were costly. This has changed in recent years, in particular with the generalization of native 64-bit integer support. The purpose of the article [40] is therefore to reevaluate the relevance of computing floating-point functions in fixed-point. For this, several variants of the double-precision logarithm function are implemented and evaluated. Formulating the problem as a fixed-point one is easy after the range has been (classically) reduced. Then, 64-bit integers provide slightly more accuracy than 53-bit mantissa, which helps speed up the evaluation. Finally, multi-word arithmetic, critical for accurate implementations, is much faster in fixed-point, and natively supported by recent compilers. Novel techniques of argument reduction and rounding test are introduced in this context. Thanks to all this, a purely integer implementation of the correctly rounded double-precision logarithm outperforms the previous state of the art, with the worst-case execution time reduced by a factor 5. This work also introduces variants of the logarithm that input a floating-point number and output the result in fixed-point. These are shown to be both more accurate and more efficient than the traditional floating-point functions for some applications.

<span style="color:red">**URBANET Team**</span>

# 7. New Results

## 7.1. Characterizing and deploying urban networks

*Participants: Ahmed Boubrima, Angelo Furno, Diala Naboulsi, Patrice Raveneau, Walid Bechkit, Marco Fiore, Hervé Rivano, Razvan Stanica.*

### 7.1.1. Collection and Analysis of Mobile Phone Data

Cellular communications are undergoing significant evolutions in order to accommodate the load generated by increasingly pervasive smart mobile devices. At the same time, recent generations of mobile phones, embedding a wide variety of sensors, have fostered the development of open sensing applications, while cellular operators are looking for new services they can provide using the data collected on their side, in the access or the core network.

The analysis of operator-side data is a recently emerged research field, and, apart a few outliers, relevant works cover the period from 2005 to date, with a sensible densification over the last three years. In [9], we provided a thorough review of the multidisciplinary activities that rely on mobile traffic datasets, identifying major categories and sub-categories in the literature, so as to outline a hierarchical classification of research lines and proposing a complete introductory guide to the research based on mobile traffic analysis. The usage of these datasets in the design of new networking solutions, in order to achieve the so-called *cognitive networking* paradigm, is discussed in detail in the PhD thesis of Diala Naboulsi [2], where the examples of green networking and virtualized radio access networks are given.

When constructing a social network from interactions among people (e.g., phone calls, encounters), a crucial task is to define the threshold that separates social from random (or casual) relationships. The ability to accurately identify social relationships becomes essential to applications that rely on a precise description of human routines, such as recommendation systems, forwarding strategies and opportunistic dissemination protocols. We thus proposed a strategy to analyze users' interactions in dynamic networks where entities act according to their interests and activity dynamics [10]. Our strategy allows classifying users interactions, separating random ties from social ones, and unveils significant differences among the dynamics of users' wireless interactions in the datasets.

Furthermore, mobile traffic data has been recently used to characterize the urban environment in terms of urban fabric profiles. While showing promising results, the existing urban fabric detection solutions are built without a clear understanding of the detection process chain. In [16], we distinguished and analyzed the different steps common to all urban profiling techniques. By evaluating the impact of each step of the process, we were able to propose a new solution that outperforms the state of the art techniques. Our approach uses the weekly periodicity of human activities, as well as a median-based filtering technique, resulting in a better clustering in terms of both coverage and entropy, as shown by results obtained on two large scale mobile traffic datasets covering the urban areas of Milan and Turin, in Italy. The solution proposed in this work was selected among the 10 finalists of the Telecom Italia Big Data challenge.

A second source of mobile data is the smartphone itself. In the context of the PrivaMov project, funded by the Labex IMU, we developed and deployed a data collection platform on more than 100 Android devices. A first step in the study of this enormous dataset (more than 50 Gb have been collected to date) was presented in [21], with a focus on the extraction of user mobility information and Wi-Fi mapping. This led us to the study of Wi-Fi tracking, a method relying on signals emitted by portable devices to track individuals for commercial, security or surveillance purposes. Wi-Fi tracking has the potential to passively track a large fraction of the population and is therefore an ideal population surveillance technology and a serious privacy threat. In [19], we argue that Wi-Fi routers make an ideal building block to create a large scale Wi-Fi tracking system, showing how they can be easily turned into Wi-Fi tracking devices through software modification. We

provided a first evaluation of the tracking capabilities of an hypothetical Wi-Fi tracking system through a set of simulations based on real-world datasets. Results showed that the spatial distribution of Wi-Fi routers is such that compromising even a small fraction of Wi-Fi routers is sufficient to track people for a large fraction of the time.

Preservation of user privacy is therefore paramount in the publication of datasets that contain fine-grained information about individuals. The problem is especially critical in the case of mobile traffic datasets collected by cellular operators, as discussed above, as they feature high subscriber trajectory uniqueness and they are resistant to anonymization through spatiotemporal generalization. In [17], we first unveiled the reasons behind such undesirable features of mobile traffic datasets, by leveraging an original measure of the anonymizability of users' mobile fingerprints. Building on such findings, we proposed GLOVE, an algorithm that grants k-anonymity of trajectories through specialized generalization. We evaluated our methodology on two nationwide mobile traffic datasets, and show that it achieves k-anonymity while preserving a substantial level of accuracy in the data.

### 7.1.2. Deployment of Wireless Sensor Networks for Pollution Monitoring

Recently, air pollution monitoring emerged as one of the main services of smart cities because of the increasing industrialization and the massive urbanization. Wireless Sensor Networks are a suitable technology for this purpose, thanks to their substantial benefits including low cost and autonomy. Minimizing the deployment cost is one of the major challenges in the design of such networks, therefore sensors positions have to be carefully determined. In [13], we proposed two integer linear programming formulations based on real pollutants dispersion modeling to deal with the minimum cost sensor network deployment for air pollution monitoring. We illustrated the concept by applying our models on real world data, namely the Nottingham City street lights. We compared the two models in terms of execution time and showed that the second flow-based formulation is much better. We finally conducted extensive simulations to study the impact of some parameters and derive some guidelines for efficient urban sensor deployment for air pollution monitoring.

## 7.2. Technology specific solutions

*Participants: Jin Cui, Walid Bechkit, Khaled Boussetta, Hervé Rivano, Fabrice Valois.*

### 7.2.1. Temperature-Aware Algorithms for Wireless Sensor Networks

Temperature variations have a significant effect on low power wireless sensor networks as wireless communication links drastically deteriorate when temperature increases. A reliable deployment should take temperature into account to avoid network connectivity problems resulting from poor wireless links when temperature increases. A good deployment needs also to adapt its operation and save resources when temperature decreases and wireless links improve. Taking into account the probabilistic nature of the wireless communication channel, in [12] we investigated the effect of temperature on percolation-based connectivity in large scale wireless sensor networks and showed that more energy can be saved by allowing some nodes to go to deep sleep mode when temperature decreases and links improve. Based on this result, we proposed a simple, yet efficient, Temperature-Aware MAC plugin (TA-MAC), which can be potentially used with any MAC protocol, enabling it to dynamically adapt the network effective density in order to allow further energy savings, while maintaining network connectivity. We carried out simulations and demonstrated that sate of the art protocols augmented with the TA-MAC plugin allow a significant energy efficiency improvement.

Going one step further, we developed a mathematical model that provides the most energy efficient deployment in function of temperature without compromising the correct operation of the network by preserving both connectivity and coverage [3]. We used our model to design three temperature-aware algorithms that seek to save energy (i) by putting some nodes in hibernate mode as in the SO (Stop-Operate) algorithm in TA-MAC, or (ii) by using transmission power control as in PC (Power-Control), or (iii) by doing both techniques as in SOPC (Stop-Operate Power-Control). All proposed algorithms are fully distributed and solely rely on temperature readings without any information exchange between neighbors, which makes them low overhead and robust. Our results identified the optimal operation of each algorithm and showed that a significant amount of energy can be saved by taking temperature into account.

### 7.2.2. Resilience in Wireless Sensor Networks

The concept of resilience for routing protocols in wireless sensor networks has been proposed and developed in the team in the last few years. In our previous works, a general overview of the resilience, including definition, metric and resilient techniques based on random behavior and data replication have been proposed. Following these previous methods, in [6] we proposed a new resilient solution based on network coding techniques, to improve resilience in wireless sensor networks for smart metering applications. More precisely, using our resilience metric based on a performance surface, we compared several variants of a well-known gradient based routing protocol with the previous methods (random routing and packet replications) and the new proposed methods (two network coding techniques). The proposed methods outperformed the previous methods in terms of data delivery success even in the presence of high attack intensity.

We also continued to study the resilience of routing protocols against malicious insiders willing to disrupt network communications. Previously, the simulation results showed that introducing randomness in routing protocols increases uncertainty for an adversary, making the protocols unpredictable. When combined with data replication, it permits route diversification between a source and a destination, thus enhancing the resilience. In [15], we proposed a theoretical framework to quantify analytically the performance of random protocols against attacks based on biased random walks on a torus lattice. The objective is to evaluate analytically the influence of bias and data replication introduced to random walks. The bias allows to decrease the route length by directing random walks toward the destination, thus reducing the probability of a data packet to meet a malicious insider along the route; however, it decreases also the degree of randomness (entropy). When random protocols are combined with data replication, the reliability is improved thanks to route diversity despite an additional overhead in terms of energy consumption.

### 7.2.3. Data aggregation in Wireless Sensor Networks

Aggregation functions are intended to save energy and capacity in Wireless Sensor Networks, by avoiding unnecessary transmissions. Aggregation functions take benefit from spatial and/or temporal correlations to forecast or to compress the real data which are collected. Although several works have focused on data aggregation in Wireless Sensor Networks, there is a lack of a formal unified framework that can compare several aggregation functions suitable for a given network topology, a given application and a target accuracy. In [14], we address this question by proposing a Markov Decision Process that can help to evaluate the performances of aggregation functions. The performances are expressed using two new proposed metrics, which can assess the energy and capacity savings of aggregation functions. As illustrative examples, we use our Markov Decision Process to evaluate and analyze the performances of basic aggregation functions (e.g. average) and more complex ones (time series, polynomial functions).

### 7.2.4. Data Gathering in Mesh Networks

In the gathering problem in mesh networks, a particular node in a graph, the base station, aims at receiving messages from some nodes in the graph. At each step, a node can send one message to one of its neighbors (such an action is called a call). However, a node cannot send and receive a message during the same step. Moreover, the communication is subject to interference constraints, more precisely, two calls interfere in a step, if one sender is at distance at most $d_I$ from the other receiver. Given a graph with a base station and a set of nodes having some messages, the goal of the gathering problem is to compute a schedule of calls for the base station to receive all messages as fast as possible, i.e., minimizing the number of steps (called makespan). The gathering problem is equivalent to the personalized broadcasting problem where the base station has to send messages to some nodes in the graph, with same transmission constraints.

In [5], we focused on the gathering and personalized broadcasting problem in grids. Moreover, we considered the non-buffering model: when a node receives a message at some step, it must transmit it during the next step. In this setting, though the problem of determining the complexity of computing the optimal makespan in a grid is still open, we presented linear (in the number of messages) algorithms that compute schedules for gathering with $d_I \in 0, 1, 2$. In particular, we presented an algorithm that achieves the optimal makespan up to an additive constant 2 when $d_I = 0$. If no messages are "close" to the axes (the base station being the origin),

our algorithms achieve the optimal makespan up to an additive constant 1 when $d_I = 0$, 4 when $d_I = 2$, and 3 when both $d_I = 1$ and the base station is in a corner.

## 7.3. Capillary Network Solutions

*Participants: Patrice Raveneau, Trista Lin, Marco Fiore, Hervé Rivano, Razvan Stanica.*

### 7.3.1. Connected Vehicles

Managing user mobility is historically one of the most critical issues in cellular radio access networks (RANs). That task will become an even greater challenge due to cellular users on-board vehicles and networked cars that autonomously access Internet-based services, whose number is expected to grow dramatically in the next few years. There is thus a need to characterize RAN access from/by vehicles in a similar way to what has been done for traditional pedestrian access. In [11], we proposed a first study of the macroscopic and microscopic features of pervasive vehicular access in a case-study large-scale urban environment, in presence of realistic datasets of the road traffic and RAN deployment. We found that pervasive vehicular access is characterized by unique temporal and spatial variability in the urban region, such that it may require a dedicated RAN capacity planning: the presence of stable vehicular access load patterns and mobility flows can help to that end. Also, we identified the theoretical distributions that best fit key metrics for RAN planning, i.e., the vehicular users' inter-arrival and residence times at cells, and discuss how their parameters vary over time and space.

Smart parking, allowing drivers to access parking information through their smart-phone, is another important service for vehicular users, which can be provided not only through cellular networks, but also by using metropolitan wireless networks, whose deployment strategy needs to be guided by efficiency and functionality. In [8], we introduced and studied a deployment strategy for wireless on-street parking sensor networks. We defined a multiple-objective problem in our analysis, and solved it with two real-world street parking maps. We presented the results on the tradeoff among minimum energy consumption, sensing information delay and the amount of deployed mesh routers and Internet gateways, i.e., the cost of city infrastructure. We also analyzed these tradeoffs to see how different urban layouts affect the optimal solutions. The overall smart parking architecture and services made the object of the PhD thesis of Trista Lin [1], where the analysis of the entire system can be found, including results on the wireless sensor networks used to collect data from parking places and the Publish-Subscribe service used to disseminate this information to users.

### 7.3.2. Offloading Cellular Networks

Offloading is a promising technique for alleviating the ever-growing traffic load from infrastructure-based networks such as the Internet. Offloading consists in using alternative methods of transmission as a cost-effective solution for network operators to extend their transport capacity. Wi-Fi offloading is one of the most effective approaches to relieve the cellular radio access from part of the burgeoning mobile demand. To date, Wi-Fi offloading has been mainly leveraged in limited contexts, such as home, office or campus environments. In [18], we investigated the scaling properties of Wi-Fi offloading, by studying how it would perform on a much larger scope than those considered today. To that end, we considered a real-world citywide scenario, built on data about actual infrastructure deployments and mobile traffic demand, and observed which amount of traffic could be accommodated by the existing pervasive Wi-Fi access infrastructure, were it opened to mobile users. We found that more than 80% of the mobile traffic demand in a large urban area may be easily served by Wi-Fi access points, under a wide range of system settings.

A new offloading technique was introduced in [20] and further detailed in [4], where we advocate the use of conventional vehicles equipped with storage devices as data carriers whilst being driven for daily routine journeys. The road network can be turned into a large-capacity transmission system to offload bulk transfers of delay-tolerant data from the Internet. The challenges we addressed include how to assign data to flows of vehicles and while coping with the complexity of the road network. We proposed an embedding algorithm that computes an offloading overlay where each logical link spans over multiple stretches of road from the underlying road infrastructure. We then formulated the data transfer assignment problem as a novel linear programming model we solve to determine the optimal logical paths matching the performance requirements

of a data transfer. We evaluated our road traffic allocation scheme using actual road traffic counts in France. The numerical results show that 20% of vehicles in circulation in France equipped with only one Terabyte of storage can offload Petabyte transfers in a week.

<span style="color:red">**Chroma Team**</span>

# 7. New Results

## 7.1. Sensor Fusion

### *7.1.1. Observability properties of the visual-inertial structure*
**Participant:** Agostino Martinelli.

We continued to investigate the visual-inertial structure from motion problem by further addressing the following issues:

1. analytically deriving its observability properties in challenging scenarios (i.e., when some of the system inputs are unknown and act as disturbances);
2. obtaining simple and efficient methods for data matching and localization.

Regarding the first issue, we extended our previous results (published last year on the journal Foundations and Trends in Robotics [43]) by also including the extreme case of a single point feature and when the camera is not extrinsically calibrated. Even if this extension seems to be simple, the analytic computation must be totally changed. Indeed, by including in the state the camera extrinsic parameters, the computation, as carried out in [43] in the case when the camera is calibrated, becomes prohibitive.

The problem of deriving the observability properties of the visual-inertial structure from motion problem, when the number of inertial sensors is reduced, corresponds to solve a problem that in control theory is known as the Unknown Input Observability (UIO). This problem is still unsolved in the nonlinear case. In [43] we introduced a new method able to provide sufficient conditions for the state observability. On the other hand, this method is based on a state augmentation. Specifically, the new extended state includes the original state together with the unknown inputs and their time-derivatives up to a given order. Then, the method introduced in [43] is based on the computation of a codistribution defined in the augmented space. This makes the computation necessary to derive the observability properties dependent on the dimension of the augmented state and consequently prohibitive in our case. Our effort to deal with this fundamental issue, was devoted to separate the information on the original state from the information on its extension. We fully solved this problem in the case of a single unknown input. For the general case, we partially solved this problem and we suggested a technique able to partially perform this separation. Since these results are very general (their validity is not limited to the visual-inertial structure from motion problem) we presented them at two international conferences on automatic control (SIAM on Control and Applications, [18] and MED, [16]). By applying these new methods to the the visual-inertial structure from motion problem, we obtained the following result. Even in the case of a single point feature, the information provided by a sensor suit composed by a monocular camera and two inertial sensors (along two independent axes and where at least one is an accelerometer) is the same as in the case of a complete inertial measurement unit (i.e., when the inertial sensors consist of three orthogonal accelerometers and three orthogonal gyroscopes). This result has been presented at ICRA, [17].

Regarding the second issue, our focus was in the framework of Micro Aerial Vehicle navigation. State of the art approaches for visual-inertial sensor fusion use filter-based or optimization-based algorithms. Due to the nonlinearity of the system, a poor initialization can have a dramatic impact on the performance of these estimation methods. Last year, we published, on the journal of computer vision, a closed-form solution providing such an initialization [42]. This solution determines the velocity (angular and linear) of a monocular camera in metric units by only using inertial measurements and image features acquired during a short time interval. This year, we study the impact of noisy sensors on the performance of this closed-form solution. Additionally, starting from this solution, we proposed new methods for both localization and data matching in the context of micro aerial navigation. These methods have been tested in collaboration with the vision and perception team in Zurich (in the framework of the ANR-VIMAD) and published on the journal of Robotics and Autonomous Systems [4].

### 7.1.2. *Sensing floor for Human & objects localisation and tracking*

**Participants:** Mihai Andries (inria Nancy, Larsen), Olivier Simonin, François Charpillet (inria Nancy, Larsen).

In the context of the PhD of Mihai Andries, co-advised by François Charpillet (Inria Nancy, Larsen) and Olivier Simonin, we investigated a large distributed sensor — a grid of connected sensing tiles on the floor — that was developed by the Maia team, at Nancy, in 2012.

Localization, tracking, and recognition of objects, robots and humans are basic tasks that are of high value in the applications of ambient intelligence. Sensing floors were introduced to address these tasks in a non-intrusive way. To recognize the humans moving on the floor, they are usually first localized, and then a set of gait features are extracted (stride length, cadence, and pressure profile over a footstep). However, recognition generally fails when several people stand or walk together, preventing successful tracking. In the Phd, defended on December 15 [27], we proposed a detection, tracking, and recognition technique which uses objects' weight. It continues working even when tracking individual persons becomes impossible. Inspired by computer vision, this technique processes the floor pressure-image by segmenting the blobs containing objects, tracking them, and recognizing their contents through a mix of inference and combinatorial search. The result lists the probabilities of assignments of known objects to observed blobs. The concept was successfully evaluated in daily life activity scenarii, involving multi-object tracking and recognition on low-resolution sensors, crossing of user trajectories, and weight ambiguity. This model can be used to provide a probabilistic input for multi-modal object tracking and recognition systems. The model and the experimental results have been published in Journal IEEE Sensors [1] and international conference ICRA 2015 [7].

## 7.2. Bayesian Perception

**Participants:** Christian Laugier, Lukas Rummelhard, Amaury Nègre, Jean-Alix David, Procópio Silveira-Stein, Jerome Lussereau, Tiana Rakotovao, Nicolas Turro (sed), Jean-François Cuniberto (sed), Diego Puschini (cea Dacle), Julien Mottin (cea Dacle).

### 7.2.1. *Conditional Monte Carlo Dense Occupancy Tracker (CMCDOT)*

**Participants:** Lukas Rummelhard, Amaury Nègre, Christian Laugier.

In 2015, the research work on *Bayesian Perception* has been done as a continuation and an extension of some previous research results obtained in the scope of the former Inria team-project e-Motion. This work exploits the *Bayesian Occupancy Filter (BOF)* paradigm [28], developed and patented by the team several years ago [0]. It also extends the more recent concept of *Hybrid Sampling BOF(HSBOF)* [46], whose purpose was to adapt the concept to highly dynamic scenes and to analyse the scene through a static-dynamic duality. In this new approach, the static part is represented using an occupancy grid structure, and the dynamic part (motion field) is modeled using moving particles. The *HSBOF* software has been implemented and tested on our experimental platforms (equipped Toyota Lexus and Renault Zoe) in 2014 and 2015; it has also been implemented in 2015 on the experimental autonomous car of Toyota Motor Europe in Brussels.

The objective of the research work performed in 2015 was to overcome some of the shortcomings of the *HSBOF* approach [0], and to obtain a better understanding of the observed dynamic scenes through the introduction an additional object level into the model. The new framework, whose development will be continued in 2016, is called *Conditional Monte Carlo Dense Occupancy Tracker (CMCDOT)* [10]. This work has mainly been performed in the scope of the project *Perfect* of IRT Nanoelec [0] (financially supported by the French ANR agency [0]), and also used in the scope of our long-term collaboration with Toyota.

---

[0]The *Bayesian programming formalism* developed in e-Motion, pioneered (together with the contemporary work of Thrun, Burgards and Fox [53]) a systematic effort to formalize robotics problems under Probability theory –an approach that is now pervasive in Robotics.

[0]In the current implementation of the HSBOF algorithm, many particles are still allocated to irrelevant areas, since no specific representation models are associated to dataless areas. Moreover, if the filtered low level representation can directly be used for various applications (for example mapping process, short-term collision risk assessment [31], [48], etc), the retrospective object level analysis by dynamic grid segmentation can be computationally expensive and subjected to some data association errors.

[0]Nanoelec Technological Research Institute (Institut de Recherche Technologique Nanoelec)

[0]National Research Agency (Agence Nationale de la recherche)

The *CMCDOT* approach introduces an drastic change in the underlying formal expressions: instead of directly filtering the occupancy data, we have added *hidden states* for representing what is currently present in a cell. Then, the occupancy distribution can then be inferred from those hidden states. Besides presenting a clear distinction between static and dynamic parts, the main interest of this new approach is to introduce a specific processing of dataless areas, excluding them from the velocity estimation (and consequently optimizing the processing of the dynamic parts) and disabling their temporal persistence (which is used to generate estimation bias in newly discovered areas). This updated formalism also enables the introduction of an appropriate formal model for the particle initialization and management (which was previously more isolated).

Another important added feature is the automatic segmentation of the dynamic parts of the occupancy grid, according to its shapes and dynamics. While the *CMCDOT* tracks spatial occupancy in the scene without object segmentation, Detection and Tracking of Moving Objects (DATMO) is often required for high level processing. A standard approach would be to analyse the *CMCDOT* outputs, to apply a clustering algorithm on the occupancy grid (enhanced by velocities), and to use those clusters as potential object level targets. This clustering can turn out to be computationally expensive, considering the grid dimensions and the size and complexity of the dynamic particle model. The basic idea of our new approach is to exploit the particle propagation process within the *CMCDOT*: the way particles are resampled can leads to the wanted segmentation after a number of time steps. After initialization, at each step, the particles that correctly fit the motion of a dynamic object are multiplied, those which do not are forgotten. In a few steps, the best particles propagate in the object, and the object motion is fully described by a set of particles deriving from a common particle root. By marking each particle at the initialization step with a unique identification number, all the dynamic areas which are coherent in term of space and motion are marked after few iterations. The convergence of those markers is fastened by additional rules.

| Sequence | Grid size (m) | HSBOF | CMCDOT |
|---|---|---|---|
| Highway | 20x70 | 76.9% | 23.5% |
| Semi-Urban | 30x60 | 89.3% | 46.7% |
| City Center | 30x60 | 93.2% | 40.1% |

*Figure 2. Estimation of irrelevant particle allocation ratio.*



|     (a)     |     (b)     |     (c)     |     (d)     |     (e)     |
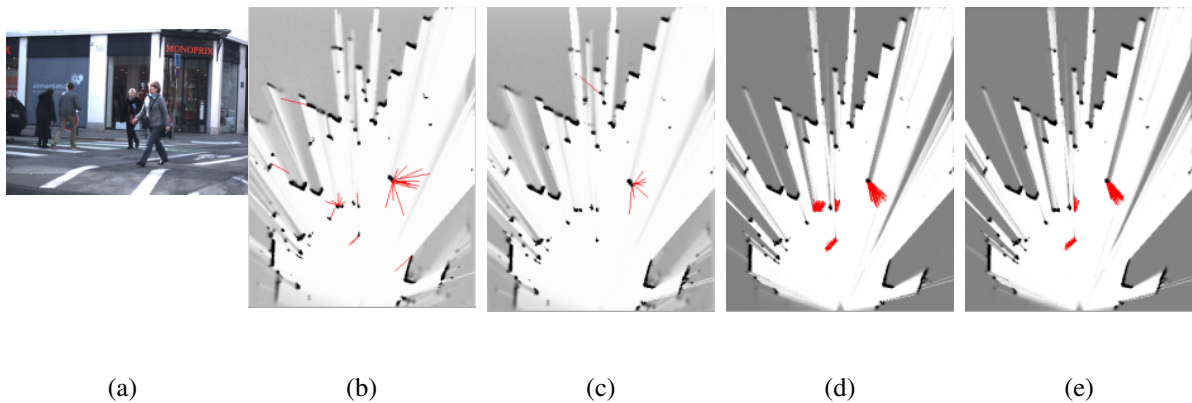
*Figure 3. Results of the HSBOF with 262144 and 32768 particles (b) and (c), and of the CMCDOT with the same number of particles (d) and (e). Red segments represent the average estimated per-cell velocity. They show that the CMCDOT is more accurate and still manages to track most of the moving pedestrians (even with a severely reduced number of samples), whereas the HSBOF loses track of almost all objects.*

Experimental results showed that the insertion of an "unknown" state in the model leads to a better distribution of dynamic samples on observed areas (see figure 2 ) and also allows us to be more reactive and accurate on the velocity distributions, while requiring less computing power (see figure 3 ).

The intrinsic clustering approach has also been tested on real road data, showing promising results in real-time tracking of moving objects, regardless of their type. The method could be improved by managing split-and-merge events that can occur in complex urban environment (see figure 4 ).



(a)(b)             (c)             (d)
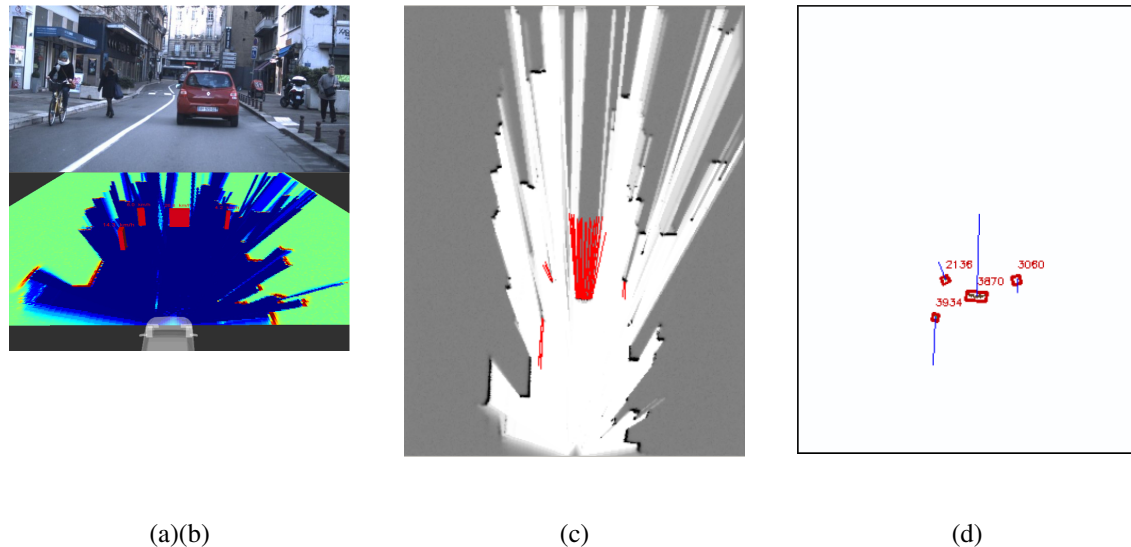
*Figure 4. Result of the dynamic objects clustering. (a) Camera image; (b) 3D view of the grid with detected objects; (c) resulting occupancy grid with velocity; (d) extracted dynamic objects (red boxes) with velocity (blue segments) and id.*

### 7.2.2. *Multimodal dynamic objects classification*

**Participants:** Amaury Nègre, Jean-Alix David.

The method described in section 7.2.1  allows to obtain a list of dynamic objects and to track each object over time. In order to increase the level of representation of the environment, we have developed a method to classify detected objects using both the camera images and the occupancy grid representation estimated by the *CMCDOT*. For each detected object, the bounding box of the object is projected in the camera image and a local image is extracted from the camera. Jointly, we can extract a patch from the occupancy grid around the dynamic object position. The extracted camera image and the occupancy grid patch can then be used as the input of a Deep Neural Network (DNN) to identify the class of the object. The DNN we designed is a combination on two classic neural networks, the "ImageNet" Convolutional Neural Networks [35] for the camera image input and the "LeNet" [37] for the occupancy grid input (see fig 5 ).

To train and evaluate the model, a dataset has been created from the data recorded with the Lexus platform. We extracted the camera images and the occupancy grid for each object detected by the *CMCDOT* module, then we manually annotated the object class among "pedestrian", "crowd", "car", "truck", "two-wheelers" and "misc" categories. The resulting dataset contains more than 100000 camera images & occupancy grid pairs. The training process and the classification module has been done by using the open source library caffe [33]. An example of the obtained results is shown on fig 6 . The percentage of good classification is greater than 90% on our evaluation dataset.

### 7.2.3. *Visual Map-Based Localisation with OSM*

**Participants:** Jean-Alix David, Amaury Nègre.
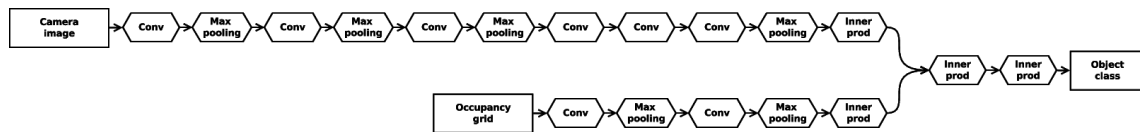
*Figure 5. Deep neural network used for obstacle detection.*
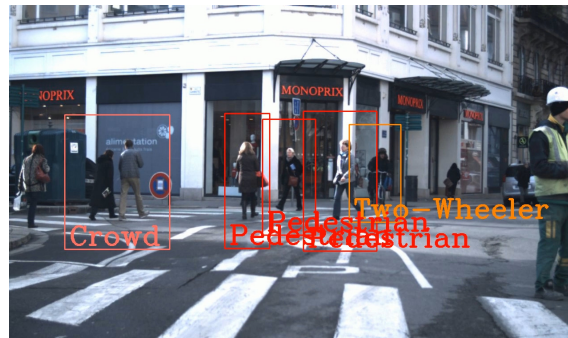


*Figure 6. Result of the dynamic obstacle classification.*

This module aims to improve both the global localization provided by the GPS [0] and the lane-relative localization information estimated by a lane tracker by combining their mutual strengths. The idea is to detect lane markings on the road using a camera, and then to compare the extracted lines with those stored in the map. This is done using the ICP [0] algorithm. This work is described in a confidential Toyota project report entitled *Real Traffic Data Acquisition and Risk Assessment Experiments*.

*7.2.3.1. The map*

Our solution is based on a post-processed OSM [0] map shown on figure 7 . Typically, these maps contain information on the roads and lanes, but contain no information about lane markers on the ground. Thus, we ran a semi-manual process to complete the existing maps with information about the number and type of markers.

New data are stocked in a local server. Requests can be sent to this server to fetch map data using HTTP protocol.

*7.2.3.2. Line detection*

The line extraction is done using ridge detection on a top-down view of the camera image. Only one monocular camera is used, as it is an inexpensive sensor, and needs only to be calibrated once. The line detection is based on an algorithm using Laplacian to extract ridges of the monochrome image. The algorithm is implemented for parallelized calculation using CUDA on a GPU, for an improved performance. Figure 8  shows the results of the ridge detector.

---

[0] Global Positioning System
[0] Iterative Closest Point
[0] OpenStreetMap
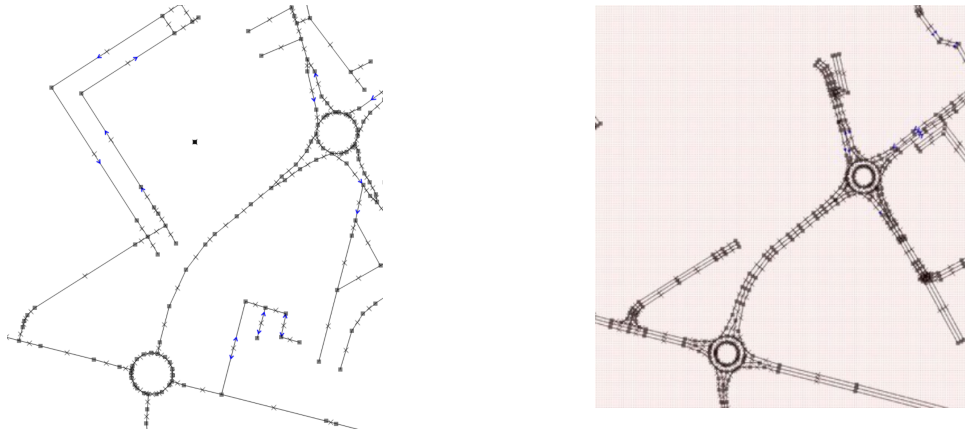
*Figure 7. Data conversion. (a) Raw OSM data, a line represents a road and it is not possible to see the lanes. (b) Modified OSM data, with lane markings.*
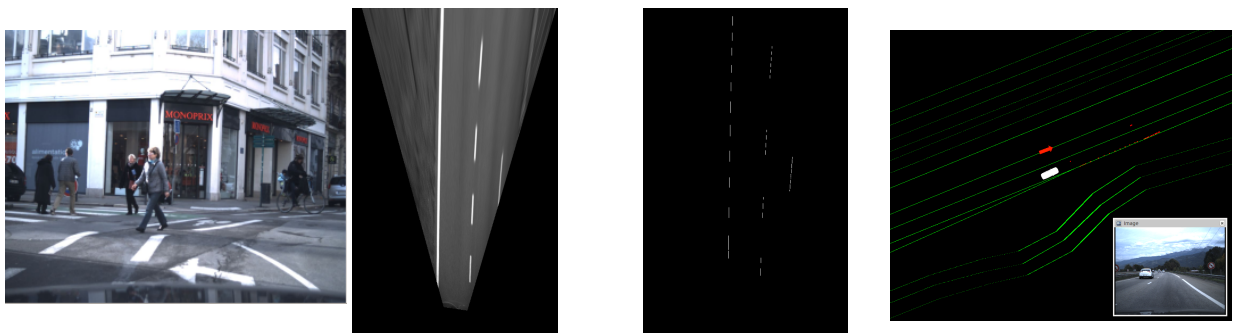


*Figure 8. Ridges detection: (a) Input image (b) Projected image (c) Detected ridges. ICP correction on highway (d)*

*7.2.3.3. ICP-based line matching*

The extracted lines are matched and aligned with the map using the ICP algorithm to improve the localization of the vehicle. The ICP algorithm iteratively minimizes the total alignment error between the points detected as ridges and the segments of line extracted from the map. Finally, figure 8 .d shows how the algorithm can correct the vehicle localization. The algorithm is able to accurately track the orientation and position. However, the lateral displacement may be off by a multiple of the lane width, depending on how the algorithm has been initialized. In practice, this effect is often mitigated due to the existence of single-lane roads such as highway entrances.

The results are very promising on highways, but the algorithm has a lower performance on other types of roads, mostly due to irregularities.

### 7.2.4. Integration of Bayesian Perception System on Embedded Platforms

**Participants:**  Tiana Rakotovao, Christian Laugier, Diego Puschini(cea Dacle), Julien Mottin(cea Dacle).

Safe autonomous vehicles will emerge when comprehensive perception systems will be successfully integrated into vehicles. However, our Bayesian Perception approach requires high computational loads that are not supported by the embedded architectures currently used in standard automotive ECUs.

To address this issue, we first explored new embedded hardware architecture credible for the integration of OGs [0] into autonomous vehicles [19]. We studied in particular recent emerging many-core architectures, which offer higher computing performance while drastically reducing the required power consumption (typically less than 1W). In such architectures, the computation of OGs can be divided into several independent tasks, executed simultaneously on separated processing core of a many-core.

Experiments were conducted on data collected from urban traffic scenario, produced by 8 LIDAR layers mounted on the Inria-Toyota experimental Lexus vehicle. These experiments demonstrate that the many-core produces OGs largely in real-time: 6 time faster than the sensor reading rate.

Besides, we also proposed a mathematical improvement of the OG model, for performing multi-sensor fusion more efficiently than the standard approach presented in [29]. In our approach, the fusion of occupancy probabilities requires fewer operations. This model improvement makes it possible the implementation of OG-based multi-sensor fusion on simple hardware architectures. This perspective applies to microcontroller, ASICs or FPGAs which are more and more present in computing platforms recently present on the automotive market.

### 7.2.5. Experimental Vehicle Renault ZOE

**Participants:**  Nicolas Turro (sed), Jean-François Cuniberto (sed), Procópio Silveira-Stein, Amaury Nègre, Lukas Rummelhard, Jean-Alix David, Christian Laugier.

*7.2.5.1. Experimental Vehicle Renault ZOE*

In the scope of the *Perfect* projet of the IRT nanoelec, we have started to develop in 2014, an experimental platform based on an equipped Renault Zoe. The development of this platform has been pursed in 2015.

The vehicle has been enhanced with a tablet to display the new HMI [0], figure 9 (a) illustrates. The HMI displays the detected dynamic objects over the camera image and the graph of collision risk at different time horizon.

New experiments have also been designed to test the perception algorithms and the recent implementation of the collision risk alert. These experiments simulate collisions with people using a fabric mannequin, as shown on figure 9 (b), and an inflatable ball.

---

[0]Occupancy Grids
[0]Human Machine Interface
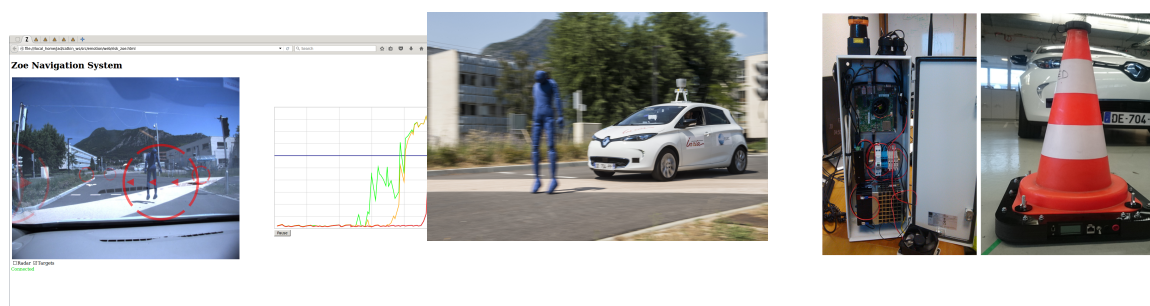
*Figure 9. (a) Display of the HMI (b) Collision simulation with a mannequin (c) On left: picture of the smartbox, on right: picture of the cone.*

Finally, we have also developed two movable devices in order to enhance V2X [0] communication experiments (see figure 9 (c)):

1. A movable communicating cone equipped with a GPS and a V2X communication box, which broadcast its position to near V2X listeners.

2. A movable smartbox equipped with a GPS, a V2X communication box, a LIDAR sensor and a Nvidia Tegra K1 board. The *CMCDOT* algorithm is implemented on it, and the detected objects are broadcasted to other communicating devices. The smartbox can be mounted on another vehicle or be placed as part of a static infrastructure. Both are alimented by batteries and aim at minimizing their energy consumption.

## 7.3. Situation Awareness

**Participants:** Christian Laugier, Alejandro Dizan Vasquez Govea, Procópio Silveira-Stein, David Sierra-Gonzalez, Mathieu Barbier, Stephanie Lefevre(uc Berkeley).

### 7.3.1. *Framework for Motion Prediction and Collision Risk Assessment*

**Participants:** Christian Laugier, Alejandro Dizan Vasquez Govea, Procópio Silveira-Stein, Stephanie Lefevre(uc Berkeley).

For several years, the challenging scientific problem of Motion Prediction, Risk Assessment and Decision-Making in open and dynamic environments has been one of our main research topics (see activity reports of the former e-Motion Inria team-project). Throughout 2016, we have continued this line of work by developing several new frameworks for Motion Prediction and Collision Risk Assessment in complex dynamic scenes involving multiple moving agents having various behaviors.

A first contribution has been the extensive experimental validation in real conditions –together with the University of Berkeley– of our *Intention-Expectation* approach: a high-level approach to risk assessment which avoids the complexity of trajectory-level reasoning while being able to take multi-vehicle interactions into account [9]. These results have also been integrated into a Mooc course at the graduate and undergraduate levels [25]. They have also been presented in several invited talks [24] [21] [22] [23].

Another contribution relies in the implementation of some the proposed models on two experimental vehicles (Lexus and Zoé experimental platforms). As mentioned in section 7.2.5 , several experiments on short-term collision risk assessement have been successfuly conducted with these platforms (c.f. [10], [15]). This work will be continued in 2016, in the scope of our ongoing collaborative projects with Toyota, Renault and IRT nanoelec.

---

[0]Vehicle-to-Vehicle and Vehicle-to-Infrastructure

### 7.3.2. *Planning-based motion prediction for collision risk estimation in autonomous driving scenarios*

**Participants:** David Sierra-Gonzalez, Alejandro Dizan Vasquez Govea, Christian Laugier.

The objective is to develop a collision risk estimation system capable of reliably finding the risk of collision associated to the different feasible trajectories of the ego-vehicle. This research work is done in the scope of the Inria-Toyota long-term cooperation and of the PhD thesis work of David Sierra- Gonzales.
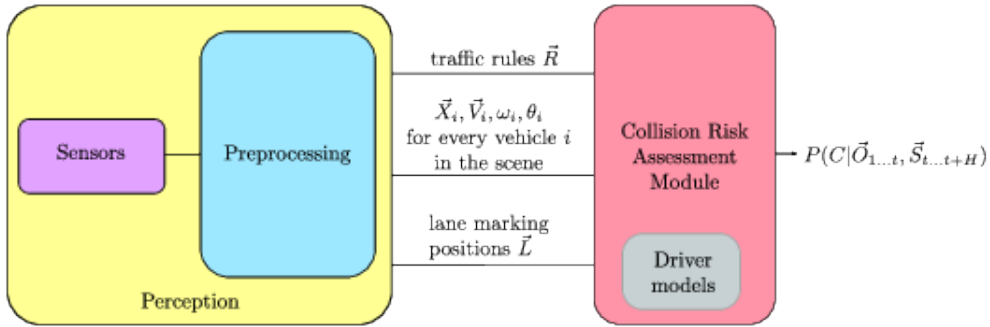
*Figure 10. Black box model for the Collision Risk Assessment Module. The number of inputs expected from the perception module is not comprehensive at this point.*

Figure 10 shows the black box model of the system. At a given timestep, the system takes the following inputs: the traffic rules in effect; the position, velocity, angular velocity and heading of each vehicle $i$ in the scene; and the position of the lane markings. Thus, at each timestep $t$ we construct an observation vector $\vec{O}_t = \left( \vec{R}_t, \{\vec{X}_i, \vec{V}_i, \omega_i, \theta_i\}_t, \vec{L}_t \right)$ with all the high-level perception inputs, and a state vector $\vec{S}_t$ with only the minimum variables necessary to describe the scene. The proposed system aims to calculate the probability of collision $C$ of the ego-vehicle for a sequence of future states up until a fixed time horizon $H$. That can be expressed as $P(C|\vec{O}_{1\cdots t}, \vec{S}_{t\cdots t+H})$. This information can then be used by a path-planner to decide upon the safest trajectory.

One key factor for the correct estimation of collision risk is the ability to predict the motion of the dynamic obstacles in the scene, that is, the other drivers. We opt here for a planning-based approach, which assumes that drivers instinctively act as to maximize a reward (or equivalently, minimize a cost). This reward function encodes the preferences of the driver to, for instance, keep a minimum distance with the vehicle in front, drive in the right lane in the highway, or respect the speed limits. Given such a reward function, Markov Decision Processes (MDP) constitute an adequate framework for the motion prediction problem. Moreover, by using Inverse Reinforcement Learning (IRL) algorithms, we can obtain such reward function directly from expert demonstrations (i.e. simply observing how people drive).

At this point, two well-known IRL algorithms ( [26], [58]) have been implemented and used to obtain a generic driver model from human demonstrations performed on a highway simulator. This driver model can now be used to predict the future behavior of the dynamic obstacles in the scene.

## 7.4. Motion-planning in human-populated environment

### 7.4.1. *Planning-based motion prediction for pedestrians in crowded environments*

**Participant:** Alejandro Dizan Vasquez Govea.
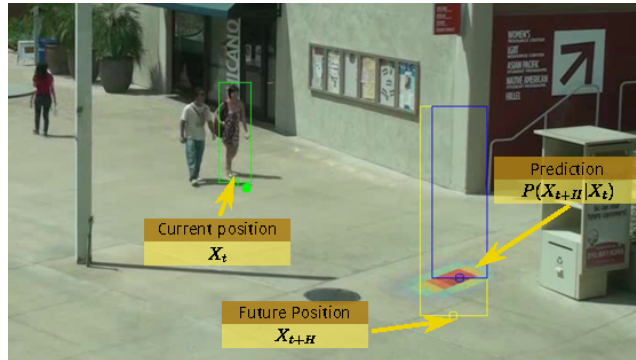
*Figure 11. Example of spatiotemporal state prediction for a single agent, indicating its current and future position after 8 sec (green and yellow boxes) as well as the predicted position probability and its maximum value (blue box)*

We have also explored the application of motion planning algorithms to the prediction of human motion (Fig. 11 ). We have proposed a novel planning-based motion prediction approach [12] which addresses the weaknesses of the previous state-of-the-art motion prediction technique [34], namely:

1. *High computational complexity*. This is dealt with by using the Fast Marching Method (FMM) [49] an efficient deterministic planning algorithm which computes the cost-to-go to a given location for every cell of a grid representing the agent's workspace. This grid is then used in a novel goal prediction algorithm and to produce a path-like prediction equivalent to the output of the Markov Decision Processes (MDPs) used by Kitani.

2. *Limited ability to model the temporal evolution along the predicted path*: this is addressed through the use of a velocity-dependent probabilistic motion model which is used to estimate a probability distribution of the future agent's position. This is then fused with a novel cost-based model to produce a full spatiotemporal prediction.

3. *Constant-goal assumption*. We propose a gradient-based goal prediction approach which does not rely on filtering, making it capable of quickly recognizing intended destination changes as they happen.

In our preliminary experiments, the proposed method significantly outperforms the accuracy of Kitani's approach while reducing the computation time by a factor of 30 using a parallel version of our algorithm.
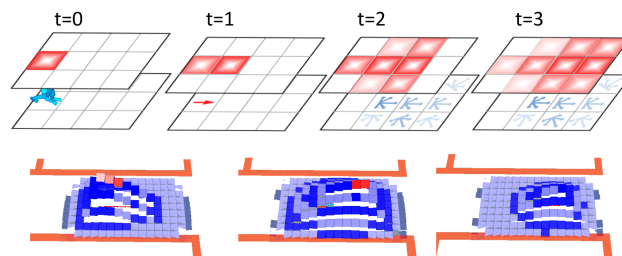


*Figure 12. Illustration of (a) the flow-grid model (b) the pheromone model of human motion prediction.*

### 7.4.2. *Modeling human-flows from robot(s) perception*

**Participants:**  Olivier Simonin, Jacques Saraydaryan, Fabrice Jumel.

To deal with navigation in highly populated environments, eg. flows of humans, we started to investigate the problem of mapping these flows. The challenge is to build such an information from robots perception while they move autonomously to perform their tasks. We also work on predicting humans location from the perceptions and the constructed flow-grid. This led us to define two models : i) a flow-grid mapping computing in each cell the probability to move in each of the $k$ possible directions (illustrated in figure 12 .a), ii) a pheromone-based model allowing to compute the current possible location of humans (flows), see figure 12 .b. We are currently measuring the efficiency of the proposed mapping compared to existing models (which do not model directions). First results will be submitted soon (to IROS 2016).

## 7.5. Multi-robot Motion-planning in dynamic environments

### 7.5.1. *Benchmarking and extension of multi-robot strategies*

#### 7.5.1.1. *Exploration of unknown and populated environments*
**Participants:**  Olivier Simonin, Nassim Kaldé (phd. Student, Larsen Inria Nancy), François Charpillet (inria Larsen, Nancy), Jan Faigl (ctu, Czech University Of Prague).

Exploration of unknown environment with a group of mobile robots consists mainly to compute a strategy that allows to visit efficiently the area while considering different constraints. These constraints can be trajectory coordination (between robots), presence of humans and limits on time, communication, and computational ressources allowed to robots. The exploration problem is related with mapping, surveillance (eg. patrolling) problems. In this context, O. Simonin and P. Lucidarme (University of Angers) published a general article on multi-robot mapping in the magazine Techniques de l'Ingénieur [5] (2015).

**Study of frontier-based strategies**

In this context, frontier-based approaches looks for an efficient allocation of the navigational goals which must be situated between the known and unknown areas (the frontiers). Goal candidate locations are repeatedly determined during the exploration. Then, the assignment of the candidates to the robots is solved as the task-allocation problem. A more frequent decision-making may improve performance of the exploration, but in a practical deployment of the exploration strategies, the frequency depends on the computational complexity of the task-allocation algorithm and available computational resources. Therefore, we proposed an evaluation framework to study exploration strategies independently on the available computational resources and we reported a comparison of the selected task-allocation algorithms deployed in multi-robot exploration [30]. This work is supported by the French-Czech PHC "Murotex".

**Exploration in populated environments**

In the context of the Phd of Nassim Kaldé, co-supervized by F. Chapillet (Inria Nancy, Larsen) and O. Simonin (Chroma), we study exploration in populated environments, in which pedestrian flows can severely impact performances. However, humans have adaptive skills for taking advantage of these flows while moving. Therefore, in order to exploit these human abilities, we propose a novel exploration strategy that explicitly allows for human-robot interactions. Our model for exploration in populated environments combines the classical frontier-based strategy with our interactive approach. For this purpose, we proposed an interaction model where robots can locally choose a human guide to follow and define a parametric heuristic to balance interaction and frontier assignments. This model is introduced in publication [3], where we evaluate to which extent human presence impacts the exploration model in terms of coverage ratio, travelled distance and elapsed time to completion. A simulator, based on V-REP and illustrated in figure 13 .a, has been developed to conduce the experimental measures.

#### 7.5.1.2. *Patrolling static and dynamic environments*
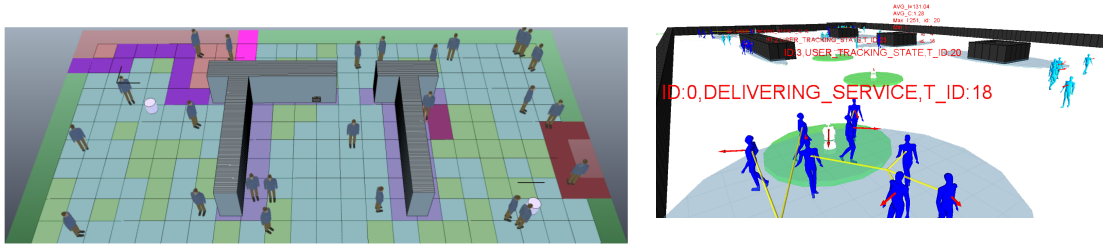**Participants:**  Olivier Simonin, Jacques Saraydaryan, Fabrice Jumel, Mihai Popescu, Herve Rivano (inria Urbanet).

*Figure 13. (a) Simulator to study exploration in populated environment based on V-REP (b) Simulator for dynamic patroling of people based on PedSim.*

**Patrolling moving people**

In the context of service robotics, we address the problem of serving people by a set of collaborating robots, that is to delever regularly services to moving people. We re-defined this problem as a dynamic patrolling task, that we called the robot-waiters problem, where robots have to regularly visit all the moving persons. In the publication [11], we proposed different criteria and metrics suitable to this problem, by considering not only the time to patrol all the people but also the equity of the delivery. We proposed and compared four algorithms, two are based on standard solutions to the static patrolling problem and two are defined according the specificity of patrolling moving entities. In order to limit robot traveled distances, the last approach introduces a clustering heuristic to identify groups among people. To compare algorithms and to prepare real experiments we developped a simulator combining a pedestrian model (PedSim) and a robotic model, illustrated in figure 13 .b. Experimental results show the efficiency of the specific new approaches over standard approaches. We also analysed the influence of the number of robots on the performances, for each approach.

We are currently developping new algorithms using the mapping and prediction of human flows based on the work presented in section 7.4.2 .

**Patrolling WSN**

In the multi-robot patrolling context, we investigated the problem of visiting regularly a set of fixed sensors by computing single-cycles on the graph formed by the WSN (Wireless sensors network). We set this problem as a graph covering with bounded hamiltonian cycles (in the M2R internship of Mihai-Ioan Popescu, now continuing as PhD student in Chroma). After giving insights of NP-hardness, we proposed a generic heuristic algorithm for solving the GCBHC. It works in two steps: the first one partitions the vertices, the second one computes hamiltonian cycles on each partition. We adapted the classic Multilevel Subgraph Partitioning algorithm to the specific requirements yielded by the networking metrics. To avoid the high complexity of this algorithm, we proposed another heuristic which exploits the geometric structure of the graph, the North-Eastern Neighbour heuristic. We implemented two classic hamiltonian cycle heuristics, one is based on Minimum Spanning Trees computations and the other on Christofides algorithm. Comparisons on randmly-generated graphs showed that the Christofides algorithm computes shorter cycles. An article presenting this work has been written and will be submitted soon.

## 7.5.2. Anytime algorithms for multi-robot cooperation

### 7.5.2.1. Observation of complex scenes

**Participants:** Olivier Simonin, Jilles Dibangoye, Laetitia Matignon (liris), Christian Wolf (liris), Jonathan Cohen (internship), Stefan Chitic.

Solving complex tasks with a fleet of robots requires to develop generic strategies that can decide in real time (or time-bounded) efficient and cooperative actions. This is particularly challenging in complex real environments. To this end, we explore anytime algorihms and adaptive/learning techniques.

The INSA BQR project "Crome" [0], led by O. Simonin, motivated the exploration of the joint-observation of complex (dynamic) scenes by a fleet of mobile robots. In our current work, the considered scenes are defined as a sequence of activities, performed by a person in a same place. Then, mobile robots have to cooperate to find a spatial configuration around the scene that maximizes the joint observation of the human pose skeleton. It is assumed that the robots can communicate but have no map of the environment and no external localisation.

To attack the problem, in cooperation with colleagues from vision (C. Wolf, Liris), we proposed an original concentric navigation model allowing to keep easily each robot camera towards the scene (see fig. 14 .a). This model is combined with an incremental mapping of the environment in order to limit the complexity of the exploration state space. We have also defined the marginal contribution of each robot observation, to facilitate stability in the search, while the exploration is guided by a meta-heuristics. We developed a simulator (fig. 14 .b) that allows to compare the variants of the approach and to show its features such as adaptation to the dynamic of the scene and robustness to the noise in the observations. Preliminary results have been presented in [8].

We have also developed an experimental framework, using Turtlebot2 robots, presented in figure 14 .c. Experimental measures and validation are in progress.
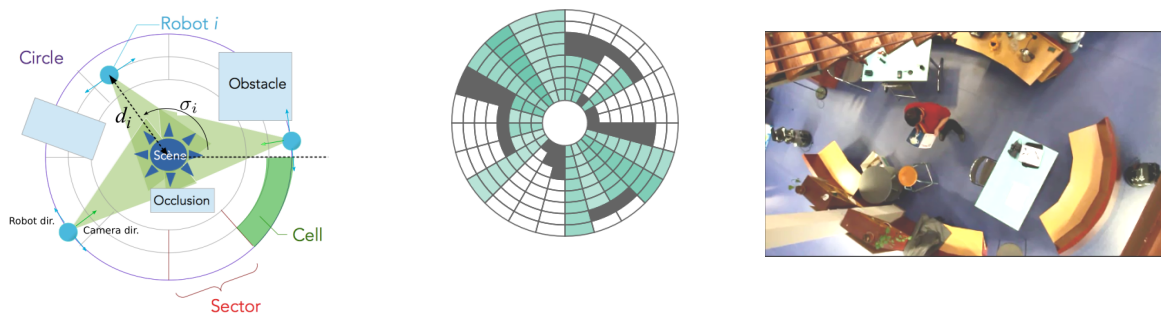


*Figure 14. Illustrations (a) Concentric navigation model, (b) Simulator and (c) experimental setup with Turtlebot 2.*

### 7.5.2.2. Middleware for multi-robot systems deployment

**Participants:** Stefan Chitic, Julien Ponge (citi, Dynamid), Olivier Simonin.

Multi-robots systems (MRS) require dedicated tools and models to face the complexity of their design and deployment (there is no or very limited tools/middleware for MRS). In this context, we addressed the problem of neighbors and service discovery in an ad-hoc network formed by a fleet of robots. Robots needs a protocol that is able to constantly discover new robots in their coverage area. This led us to propose a robotic middleware, SDfR, that is able to provide service discovery. This protocol is an extension of the Simple Service Discovery Protocol (SSDP) used in Universal Plug and Play (UPnP) to dynamic networks generated by the mobility of the robots. Even if SDfR is platform independent, we proposed a ROS integration in order to facilitate the usage. We evaluated a series of overhead benchmarking across static and dynamic scenarios. Eventualy, we experimented some use-cases where our proposal was successfully tested with Turtlebot 2 robots. Results have been presented to the national conference CAR'2015 and will appear in the international conference ICAART 2016 (accepted).

---

[0]Coordination d'une flottille de robots mobiles pour l'analyse multi-vue de scènes complexes

### 7.5.3. Sequential decision-making under uncertainty

Sequential decision-making under uncertainty is a core area of artificial intelligence, optimization, operations research, machine learning, and robotics. It involves one or multiple decision makers (or agents or robots) reasoning about the course of actions to achieve collective or self-interested goals while accounting both for the outcomes of current decisions and for future decision-making opportunities. Markov models (e.g., Markov decision processes and Markov games) have emerged as normative frameworks for optimizing decision under uncertainty. These models encompass a wide range of real-world applications: controlling intelligent vehicles; optimizing the production and distribution of energy resources; protecting endangered species; making telecommunication protocols faster and safer; monitoring and assisting elderly patients at home; designing robotic exploration technologies for search and rescue; but also many other applications. Decentralized partially observable Markov decision processes have emerged as the fundamental model to address multiple decision makers' decision-theoretic planning and learning problems. In that direction, we investigate generic, highly scalable and adaptable planning and learning algorithms to apply eventually in multi-robot planning tasks.

*7.5.3.1. Structural results for cooperative decentralized control problems*
**Participants:** Jilles S. Dibangoye, Olivier Simonin, Olivier Buffet (inria Nancy, Ex Maia Team), Mamoun Idrissi (internship, Insa de Lyon).

The intractability in cooperative, decentralized control problems is mainly due to prohibitive memory requirements in both optimal policies and value functions. The complexity analysis has emerged as the standard method to estimating the memory needed for solving a given computational problem, but complexity results may be somewhat limited. Our work [13] introduces a general methodology, called the structural analysis, for the design of optimality-preserving concise policies and value functions, which will eventually lead to the development of efficient theory and algorithms. For the first time, we showed that memory requirements for policies and value functions may be asymmetric, resulting in cooperative, decentralized control problems with exponential reductions in memory requirements. To apply this theoretical in robotics, we investigate during M. Idrissi's internship the robotic coverage of unknown areas.

*7.5.3.2. State-of-the-art algorithms for optimally solving Dec-POMDPs*
**Participants:** Jilles S. Dibangoye, Christopher Amato (univ. New Hampshire), Olivier Buffet (inria Nancy, Ex Maia Team), François Charpillet (inria Nancy, Larsen Team), Martin Pugnet (master Student, U. Claude Bernard Lyon).

Decentralized partially observable Markov decision processes (Dec-POMDPs) provide a general model for decision-making under uncertainty in cooperative decentralized settings but are difficult to solve optimally (NEXP-Complete). As a new way of solving these problems, we introduced the idea of transforming a Dec-POMDP into a continuous-state deterministic MDP with a piecewise-linear and convex value function. This approach makes use of the fact that planning can be accomplished in a centralized offline manner while execution can still be decentralized. This new Dec-POMDP formulation, which we call an occupancy MDP, allows powerful POMDP and continuous-state MDP methods to be used for the first time. To provide scalability, we refine this approach by combining heuristic search and compact representations that exploit the structure present in multi-agent domains, without losing the ability to converge to an optimal solution. In particular, in [14], we introduce a feature-based heuristic search value iteration (FB-HSVI) algorithm that relies on feature-based compact representations, point-based updates, and efficient action selection. However, scalability remains limited when the number of agents or problem variables becomes large. To overcome this limitation, we show that, under certain separability conditions of the optimal value function, the scalability of this approach can increase considerably. This separability is present when there is the locality of interaction between agents, which can be exploited to improve performance. A theoretical analysis demonstrates that FB-HSVI terminates in finite time with an optimal solution. We include an extensive empirical analysis using well-known benchmarks, thereby confirming that our approach provides significant scalability improvements compared to the state of the art. We push even further the envelope, during Martin's internship, assuming we only have access to an incomplete model of the world. This more realistic assumption that would ease application to robotics leads us directly to learning algorithms inspired from FB-HSVI.

*7.5.3.3. Distributed projected gradient-descent algorithm applied to smart grids*

**Participants:** Jilles S. Dibangoye, Arnaud Doniec (uria – Ecole Des Mines de Douai, France), H. Fakham, F. Colas And X. Guillaud (l2ep – Arts Et Métiers Paristech, France).

In a smart grid context, the increasing penetration of embedded generation units leads to a greater complexity in the management of production units. In this work, we focus on the impact of the introduction of decentralized generation for the unit commitment (UC) problem. Unit commitment problems consist in finding the optimal schedules and amounts of power to be generated by a set of generating units in response to an electricity demand forecast. While this problem has received a significant amount of attention, classical approaches assume that these problems are centralized and deterministic. However, these two assumptions are not realistic in a smart grid context. Indeed, finding the optimal schedules and amounts of power to be generated by multiple distributed generator units is not trivial since it requires to deal with distributed computation, privacy, stochastic planning, etc. Our contribution focuses on smart grid scenarios where the main source of complexity comes from the proliferation of distributed generating units. In solving this issue, we consider distributed stochastic unit commitment problems. In [2], we introduce a novel distributed gradient descent algorithm which allows us to circumvent classical assumptions. This algorithm is evaluated through a set of experiments on real-time power grid simulator.

## EXMO Project-Team

# 7. New Results

## 7.1. Ontology matching and alignments

We pursue our work on ontology matching and alignment support [4] with contributions to evaluation and the use of algebras of relations within alignments.

### 7.1.1. Evaluation

**Participant:** Jérôme Euzenat [Correspondent].

Since 2004, we run the Ontology Alignment Evaluation Initiative (OAEI) which organises evaluation campaigns for assessing the degree of achievement of actual ontology matching algorithms [3].

This year, we also handed out the organisation of OAEI 2015 to Ernesto Jiménez Ruiz (University of Oxford). We used again our generator for generating new version of benchmarks. The Alignment API was used for manipulating alignments and evaluating results [8].

The participating systems and evaluation results were presented in the 10th Ontology Matching workshop [13], held Bethleem (PA US). More information on OAEI can be found at http://oaei.ontologymatching.org/.

### 7.1.2. Algebras of alignment relations

**Participants:** Armen Inants [Correspondent], Jérôme Euzenat.

Qualitative calculi are central in qualitative binary constraint satisfaction problems. All formalisms developed so far are homogeneous – they assume a single universe. We had previously shown the advantages of using a homogeneous qualitative calculus for expressing ontology alignment relations between concepts.

They make it possible to aggregate alignments disjunctively or conjunctively and to propagate alignments within a network of ontologies. The previously considered algebra of relations contains taxonomical relations between classes only. We have tackled the problem of combining two or more calculi over disjoint universes into a single calculus [9]. The problem is important because ontology matching deals with various kinds of ontological entities: concepts, individuals, properties. We have designed an algorithm for combining two homogeneous calculi with different universes into a single calculus. This has been applied to alignment relations [9] combining algebras for relations between concepts and individuals. It is, first, able to deal with empty classes, and, second, incorporates all qualitative taxonomical relations that occur between individuals and concepts, including the relations "is a" and "is not". We have proved that this algebra is coherent with respect to the simple semantics of alignments.

The proposed algebras of relations and others have been integrated within the Alignment API (§6.1 ).

This work is part of the PhD of Armen Inants.

## 7.2. Data interlinking

The web of data uses semantic web technologies to publish data on the web in such a way that they can be interpreted and connected together. It is thus important to be able to establish links between these data, both for the web of data and for the semantic web that it contributes to feed. We consider this problem from different perspectives.

### 7.2.1. Interlinking cross-lingual RDF data sets

**Participants:** Tatiana Lesnikova [Correspondent], Jérôme David, Jérôme Euzenat.

RDF data sets are being published with labels that may be expressed in different languages. Even systems based on graph structure, ultimately rely on anchors based on language fragments. In this context, data interlinking requires specific approaches in order to tackle cross-lingualism. We proposed a general framework for interlinking RDF data in different languages and implemented two approaches: one approach is based on machine translation, the other one takes advantage of multilingual references, such as BabelNet. This year we investigated the second approach [10], finding that results were not as good as the translation approach. We also conducted evaluations on TheSoz, Agrovoc and Eurovoc thesauri.

This work is part of the PhD of Tatiana Lesnikova developed in the LINDICLE project (§9.1.1 ).

### 7.2.2. *An iterative import-by-query approach to data interlinking*
**Participant:** Manuel Atencia Arcas [Correspondent].

We modelled the problem of data interlinking as a reasoning problem on possibly decentralised data. We described an import-by-query algorithm that alternates steps of sub-query rewriting and of tailored querying of data sources [11]. It only imports data as specific as possible for inferring or contradicting target owl:sameAs assertions. Experiments conducted on a real-world dataset have demonstrated in practice the feasibility and usefulness of this approach for data interlinking and disambiguation purposes.

Additionally, and in line with the problem of dealing with uncertainty in linked data, we have proposed a probabilistic mechanism of trust that allow peers in a semantic peer-to-peer network to select the peers that are better suited to answer their queries, when query reformulation based on alignments may be unsatisfactory due to unsoundness or incompleteness of alignments [5].

This work was carried out in collaboration with Mustafa Al-Bakri and Marie-Christine Rousset (LIG).

### 7.2.3. *Link key extraction*
**Participants:** Jérôme David [Correspondent], Manuel Atencia Arcas, Jérôme Euzenat.

Ontologies do not necessarily come with key descriptions, and never with link key assertions (§3.3 ). Keys can be extracted from data by assuming that keys holding for specific data sets, may hold universally.

Following the work of last year on link key extraction [1] and the characterisation of the approach in formal concept analysis, we have fully characterised the results of our algorithm as formal concepts. We have also plans for extending both the approach and its formal concept analysis description through (i) applying it to full link keys as described in §3.3 , (ii) applying it to join and hierarchical key extraction, and (iii) applying it to hierarchical key extraction.

This work has been developed partly in the LINDICLE project (§9.1.1 ). Formal concept analysis aspects are considered with Amedeo Napoli (Orpailleur, LORIA).

## 7.3. Dynamic aspects of networks of ontologies

Huge quantities of data described by ontologies and linked together are made available. These are generated in an independent manner by autonomous providers such as individuals or companies. They are heterogeneous and their joint exploitation requires connecting them.

However, data and knowledge have to evolve facing changes in what they represent, changes in the context in which they are used and connections to new data and knowledge sources. As their production and exchange are growing larger and more connected, their evolution is not anymore compatible with manual curation and maintenance. We work towards their continuous evolution as it is critical to their sustainability.

Two different approaches are currently explored.

### 7.3.1. *Evolution of ontology networks and linked data*
**Participants:** Adam Sanchez Ayte [Correspondent], Jérôme David, Jérôme Euzenat.

We are considering the global evolution of knowledge represented by interdependent ontologies, data, alignments and links. Our goal is to be able to maintain such a structure with respect to the processes which are involved in its construction: logical inference, ontology matching, link key extraction, link generation, etc.

Our initial work is focused on how data and ontology changes cause alignment evolution, in particular when the alignment have been produced through instance-based matching using links between data. In this regard, we are developing techniques for circumscribing the elements and relationships affected by the change as well as evaluating the need for change propagation, i.e, most of the time a simple change will not trigger link key recomputation (§7.2.3 ).

This work is part of the PhD thesis of Adam Sanchez Ayte developed in the LINDICLE project (§9.1.1 ).

### 7.3.2. *Revision in networks of ontologies*
**Participant:** Jérôme Euzenat [Correspondent].

We reconsidered the belief revision problem in the context of networks of ontologies (§3.2 ): given a set of ontologies connected by alignments, how to evolve them such that they account for new information. In networks of ontologies, inconsistency may come from two different sources: local inconsistency in a particular ontology or alignment, and global inconsistency between them. Belief revision is well-defined for dealing with ontologies; we have investigated how it can apply to networks of ontologies. We formulated revision postulates for alignments and networks of ontologies based on an abstraction of existing semantics of networks of ontologies. We showed that revision operators cannot be simply based on local revision operators on both ontologies and alignments. We adapted the partial meet revision framework to networks of ontologies and show that it indeed satisfies the revision postulates [7]. Finally, we considered strategies based on network characteristics for designing concrete revision operators.

<center>**IMAGINE Project-Team**</center>

# 6. New Results

## 6.1. User-centered Models for Shapes and Shape Assemblies

- **Scientist in charge**: Stefanie Hahmann.
- **Other permanent researchers**: Marie-Paule Cani, Jean-Claude Léon, Damien Rohmer.

Our goal, is to develop responsive shape models, i.e. 3D models that respond in the expected way under any user action, by maintaining specific application-dependent constraints (such as a volumetric objects keeping their volume when bent, or cloth-like surfaces remaining developable during deformation, etc). We are extending this approach to composite objects made of distributions and/or combination of sub-shapes of various dimensions.

### 6.1.1. Developable Surfaces

**Participants:** Antoine Begault, Marie-Paule Cani, Stefanie Hahmann, Damien Rohmer, Camille Schreck.
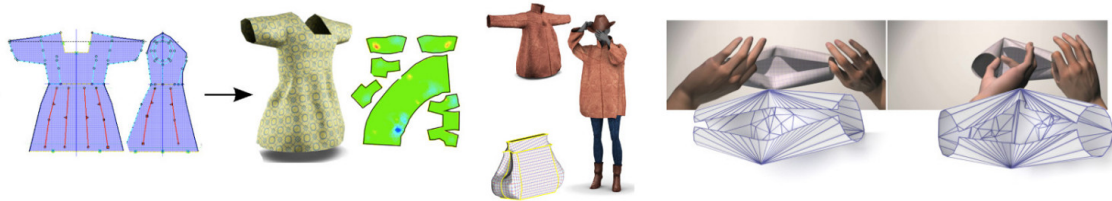


*Figure 5. Example of developable model generation from sketches in [10] and for virtual paper interaction [14].*

Developable surfaces are surfaces which can be unflattened on a plane without being stretched nor squeezed. In other words, they can be made from 2D pattern without change of lengths. They are usually hard to model efficiently as the length condition is non linear. We developed this year two different applications for developable surfaces, once applied for leather product designer, and the other one to virtual paper deformation.

We developed a method to generate 3D models for garments and leather products from designer sketches. Given two or three orthogonal sketched views depicting the silhouette, the seams, and the folds, we automatically compute a 3D developable surface and the corresponding 2D patterns which fits the silhouette and exhibits the designed folds. Our method can handle complex cases where the 2D silhouette actually correspond to a non planar and discontinuous curve on the 3D surface. We also proposed a new efficient approach to improves the developability of the resulting surface while preserving the pre-designed folds. This work has been published in ACM Transactions on Graphics [10], and we presented it in SIGGRAPH Asia in November.

Within the PhD work of Camille Shreck, we developed the first interactive 3D virtual model of crumpled paper. Deforming virtual paper is especially challenging to model efficiently has crumpling can be seen as singularities on the surface, leading therefore to non smooth surfaces which do not fit well to standard physically based deformation model. We proposed in this work a new geometrical representation of surface especially adapted to model non smooth developable surfaces as a set of planes, cylinders, and generalized cones meeting at the discontinuities of the surface. Our model can dynamically adapts to the surface deformation and to new crumples, while been associated to an optimal mesh triangulation containing very few triangles. Our interactive deformation model interleaves a standard Finite Element Model on the coarse

triangular mesh to guide the general deformation, with a geometrical steps adapting our surface structure to optimally sample the degrees of freedom of the crumpled paper. This work as been accepted for publication in ACM Transaction on Graphics [14], has been presented at the conference WomEncourage [29], and as a communication in AFIG [33].

### 6.1.2. Procedural models for shape assemblies

**Participants:** Marie-Paule Cani, Damien Rohmer, Ulysse Vimont.
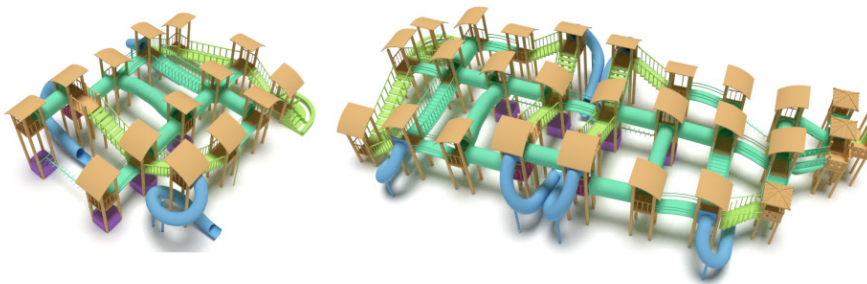


*Figure 6. Example of shape assemblies before and after subpart-remplacement in [11].*

A popular mode of shape synthesis involves mixing and matching parts from different objects to form a coherent whole. In collaboration with the University College London, Universiteit Utrecht, and KAUST, we proposed a method to automatically detect replaceable subparts within a complex assembly. In this work, we model the geometrical assembly as a graph where each node represent a single component, and the edges represents inter-part connectivity. Our method analyses this graph to detect similar inter-part connectivity enabling to exchange or mix sub-structures to synthesise new geometrical models. This work has been published in Eurographics [11].

### 6.1.3. Toward Functional CAD assemblies

**Participants:** Pablo Covès, Harold Vilmart, Robin Roussel, Damien Rohmer, Marie-Paule Cani, Jean-Claude Léon.
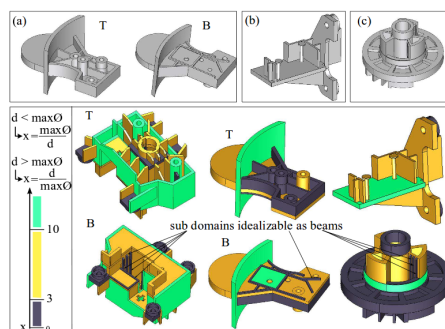


*Figure 7. Example of shape idealization in [5].*

We chose to focus on man-made objects to tackle the topic of shape assemblies. This is two-folds since CAD models of virtual industrial prototypes provide an excellent, real-size test-bed for our methods. Moreover, this is perfectly fitting the demand from industrial partners such as EDF and Airbus Group. On a complementary basis, we have initiated a partnership with UCL (University College London) to address function-preserving assembly deformation.

Assemblies representing products are most often reduced to a collection of independent CAD models representing each component. To our knowledge, there has been no approach proposed to generate CAD assembly models from 3D scans. An approach is initiated with a partnership with LIRIS (R. Chaine and J. Digne) and EDF in the framework of a Rhône-Alpes region project (Potasse) starting with the PhD of P. Coves.

Following the work of [43],[5], [38], partnership with Inria GRAPHIK team (F. Ulliana) has been set up and a deductive logic framework has been coupled to the SALOME plateform with the insertion of an ontology describing a subset of a product structure. This partnership is developed with the internship of H. Vilmart to evolve toward an intrinsic, knowledge-based representation of a product structure that takes into account the isometries of components using our prior work about symmetry analyses [42]. The description of components through this product structure aims at supporting the generation of CAD assembly models from 3D scans to be able to derive functionally meaningful constraints of relative positions of components extracted from scans.

In the scope of the ERC Expressive, a partnership has been set up with N. Mitra (UCL) with the starting PhD of R. Roussel addressing function-preserving assembly deformation.

## 6.2. Motion & Sound Synthesis

- **Scientist in charge**: François Faure.
- **Other permanent researchers**: Marie-Paule Cani, Damien Rohmer, Rémi Ronfard.

Animating objects in real-time is mandatory to enable user interaction during motion design. Physically-based models, an excellent paradigm for generating motions that a human user would expect, tend to lack efficiency for complex shapes due to their use of low-level geometry (such as fine meshes). Our goal is therefore two-folds: first, develop efficient physically-based models and collision processing methods for arbitrary passive objects, by decoupling deformations from the possibly complex, geometric representation; second, study the combination of animation models with geometric responsive shapes, enabling the animation of complex constrained shapes in real-time. The last goal is to start developing coarse to fine animation models for virtual creatures, towards easier authoring of character animation for our work on narrative design.

### 6.2.1. Real-time physically-based models

**Participants:** Marie-Paule Cani, Francois Faure, Pierre-Luc Manteaux, Richard Malgat, Matthieu Nesme.
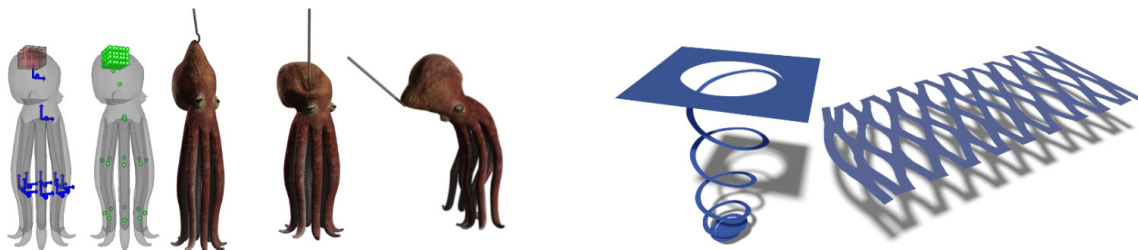


*Figure 8. Left: Mixing a coarse frame-based simulation to a local FEM patch from [24]. Right: Frame based simulation for surface cutting in [25].*

We keep on improving fundamental tools in physical simulation, such as new insight on constrained dynamics [15] at Siggraph. This allows more stable simulations of thin inextensible objects. A new extension of our volumetric contact approach (Siggraph 2010 and 2012) has been proposed [17] to apply rotational reaction to contact according to the shape of the contact area.

We have proposed an original approach to multi-resolution simulation, in which arbitrary deformation fields at different scales can be combined in a physically sound way[24]. This contrasts with the refinement of a given technique, such as hierarchical splines or adaptive meshes.

Following the success of frame-based elastic models (Siggraph 2011), a real-time animation framework provided in SOFA and currently used in many of our applications with external partners, we proposed an extension to the cutting of surface objects [25], in collaboration with Berkeley, where Pierre-Luc Manteaux spent 4 months at the end of 2014.

### 6.2.2. *Simulating paper material with sound*

**Participants:**  Marie-Paule Cani, Pierre-Luc Manteaux, Damien Rohmer, Camille Schreck.



*Figure 9. Left: Example of our paper tearing model in [23]. Right: Our sound synthesis for paper crumpling in [34].*

Extending our results on animating paper crumpling, we proposed to synthesise the sound associated to paper material. We proposed a real time model dedicated to paper tearing. In this work, we model the specific case when two hands are tearing a flat sheet of paper on a table, in this case we synthesise procedurally the geometrical deformation of the sheet using conical surface, the tearing using a procedural noise map, and the tearing sound as a modified white noise depending on the speed of action. This work has been published in Motion in Games conference [23].

We are also developing a sound synthesis method for paper crumpling. The geometrical surface deformation is analysed to drive a procedurally synthesized friction sound and a data driven crumpling sound. We are currently developing this work and did a first communication to AFIG conference [34].

### 6.2.3. *Animating anatomy*

**Participants:**  Armelle Bauer, Ali Hamadi Dicko, Francois Faure, Olivier Palombi, Damien Rohmer.

A real-time spine simulation model leveraging the multi-model capabilities of SOFA was presented in an international conference on biomechanics [7]. We also used a biomechanical model to regularize real-time motion capture and display, and performed live demos at the Emerging Technologies show of Siggraph Asia, Kobe, Japan [41].

We are developing an ontology-based virtual human embryo development model. In one side, a dedicated ontology stores the anatomical knowledge about organs' geometry, relations, and development rules. On the other side, we synthesize an animated visual 3D model using the informations of the ontology. This work can be seen as a first step toward interactive development anatomy teaching, or simulation, based on an ontology storing existing medical knowledge. This work has been published in the Journal of Biomedical Semantics [12].
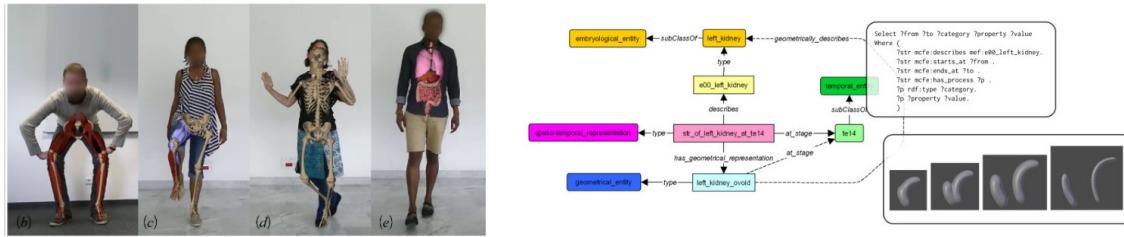
*Figure 10. Left: Example of the Living Book of Anatomy in [41]. Right: Example of kidney development ontology and its 3D modeling in [12].*

## 6.3. Knowledge-based Models for Narrative Design

- **Scientist in charge**: Rémi Ronfard.

- **Other permanent researchers**: Marie-Paule Cani, François Faure, Jean-Claude Léon, Olivier Palombi.

Our long term goal is to develop high-level models helping users to express and convey their own narrative content (from fiction stories to more practical educational or demonstrative scenarios). Before being able to specify the narration, a first step is to define models able to express some a priori knowledge on the background scene and on the object(s) or character(s) of interest. Our first goal is to develop 3D ontologies able to express such knowledge. The second goal is to define a representation for narration, to be used in future storyboarding frameworks and virtual direction tools. Our last goal is to develop high-level models for virtual cinematography such as rule-based cameras able to automatically follow the ongoing action and semi-automatic editing tools enabling to easily convey the narration via a movie.

### 6.3.1. Virtual direction tools

**Participants:** Adela Barbulescu, Rémi Ronfard.



*Figure 11. Live demo at the EXPERIMENTA exhibition in Grenoble.*

During the third year of Adela Barbulescu's PhD thesis, we proposed a solution for converting a neutral speech animation of a virtual actor (talking head) to an expressive animation. Using a database of expressive audiovisual speech recordings, we learned generative models of audiovisual prosody for 16 dramatic attitudes (seductive, hesitant, jealous, scandalized, etc.) and proposed methods for transferring them to novel examples. Our results demonstrate that the parameters which describe an expressive performance present person-specific signatures and can be generated using spatio-temporal trajectories; parameters such as voice spectrum can be obtained at frame-level, while voice pitch, eyebrow raising or head movement depend both on the frame and the temporal position at phrase-level. This work was presented at the first joint conference on facial animation and audio-visual speech processing [16] and in a live demo at the EXPERIMENTA exhibition in Grenoble, and was seen by 1200 visitors.

### 6.3.2. *Virtual cinematogaphy*
**Participants:** Quentin Galvane, Rémi Ronfard.
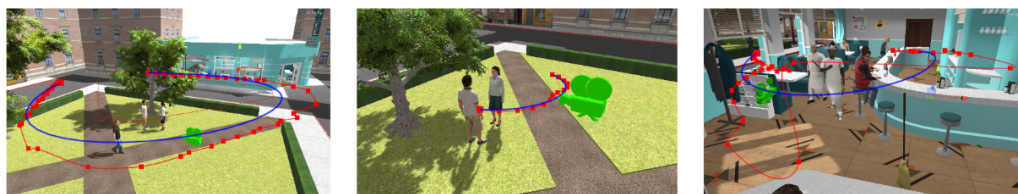


*Figure 12. Automatic camera trajectory in [18].*

During the third year of Quentin Galvane's Phd thesis, we proposed a solution for planning complex camera trajectories in crowded animation scenes [2] Galvane [18]. This work was done in a collaboration with Marc Christie in Rennes.

We also published new results from Vineet Gandhi's PhD thesis (defended in 2014) on the generation of cinematographic rushes from single-view recordings of theatre performances [32]. In that paper, we demonstrate how to use our algorithms to generate a large range of dynamic shot compositions from a single static view, a process which we call "vertical editing". Our patent application on this topic was reviewed positively and is being extended.

Those techniques were used to automatically generated cinematographically pleasant rushes from a monitor camera during rehearsals at Theatre des Celestins, as part of ANR project "Spectacle-en-Lignes". Results of the projects are described in two papers [28], Steiner [30] and we presented them to a professional audience during the Avignon theatre festival. This work was done in a collaboration with the Institut de Recherche et d'Innovation (IRI) at Centre Pompidou and the SYLEX team at LIRIS.

### 6.3.3. *Film editing & narrative design*
**Participants:** Quentin Galvane, Rémi Ronfard.

We proposed a new computational model for film editing at the AAAI artificial intelligence conference, which is based on semi-Makov chains [20]. Our model significantly extends previous work by explicitly taking into account the crucial aspect of timing (pacing) in film editing. Our proposal is illustrated with a reconstruction of a famous scene of the movie "Back to the future" in 3D animation, and a comparison of our automatic film editing algorithms with the director's version. Results are further discussed in two companion papers [31], Galvane [19]. This work was done in a collaboration with Marc Christie in Rennes. Future work is being planned to extend this important work to the case of live-action video (as described in the previous section) and to generalize for the case non-linear film editing including temporal ellipses and flashbacks.

*Figure 13. Example of automatic movie sequence editing in [20].*

# 6.4. Creating and Interacting with Virtual Prototypes

- **Scientist in charge**: Jean-Claude Léon.
- **Other permanent researchers**: Marie-Paule Cani, Olivier Palombi, Damien Rohmer, Rémi Ronfard.

The challenge is to develop more effective ways to put the user in the loop during content authoring. We generally rely on sketching techniques for quickly drafting new content, and on sculpting methods (in the sense of gesture-driven, continuous distortion) for further 3D content refinement and editing. The objective is to extend these expressive modeling techniques to general content, from complex shapes and assemblies to animated content. As a complement, we are exploring the use of various 2D or 3D input devices to ease interactive 3D content creation.

## 6.4.1. Sculpting shape hierarchies
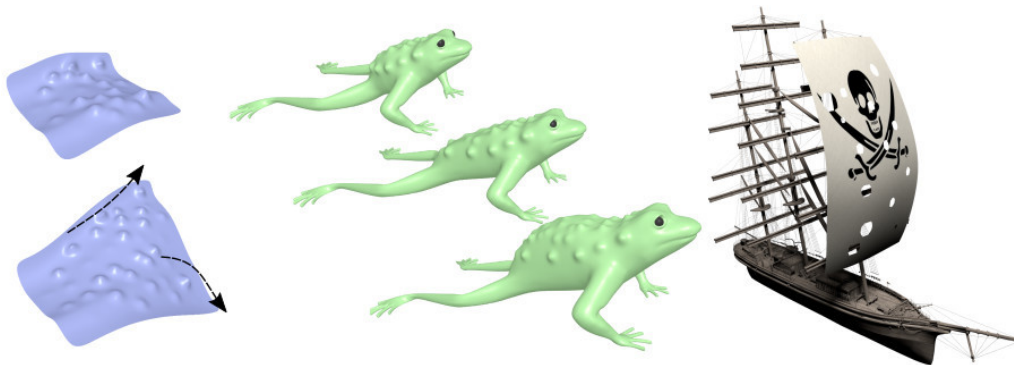


*Figure 14. Our method in [13] allows surface stretching while the geometrical details are continuously duplicated.*

Sculpting paradigm has been successfully applied to deform simple smooth surfaces. More complex objects representing virtual characters or real-life objects are however modeled as hierarchy of shapes with elements, sub-elements and details. Applying sculpting deformation to such objects is challenging as every parts of the hierarchy should stay coherent through the deformation.

When an object can be represented as a smooth underlying surface and a set of singular details, we proposed a real-time deformation approach enabling to freely stretch or squeeze the 3D object while continuously maintaining the details' appearance. Instead of stretching or squeezing the details the same way than the smooth underlying surface, we duplicate or merge them smoothly while ensuring that the distribution of details has the same caracteristic than the original one. We published this work in Shape Modeling International [13].

In the case of more general object hierarchies, we are developing a new methodology to apply generic deformation into complex assemblies while preserving their properties in extending the shape grammar approach into our new *deformation grammar*. We presented our preliminary results as a communication in the GTMG conference [35].

### 6.4.2. Sketching and sculpting Virtual Worlds

**Participants:**  Marie-Paule Cani, Guillaume Cordonnier, Ulysse Vimont.
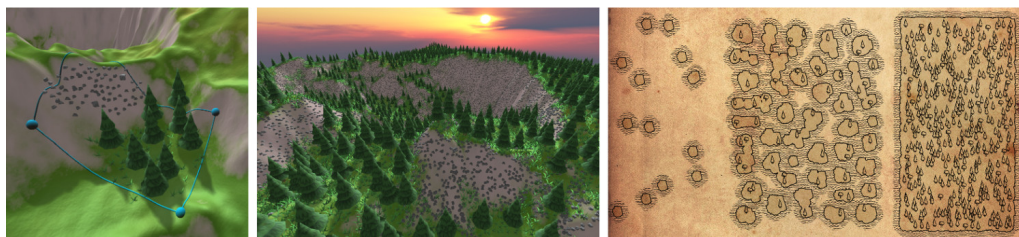


*Figure 15. Example of maps generated using our World Brush method [8].*

Modeling virtual worlds is particular challenging: the fractal-like distribution of details in terrain shapes makes them easy to identify, but very difficult to design using standard modeling software, even for expert users. Moreover, virtual worlds involve distributions of different categories of contents over terrains, such as vegetation, houses, roads or rivers. Efficiently modeling these sets of elements, which are statistically correlated, is indeed a challenge.

This year, our contributions to tackle these issues were two-folds:

Firstly, we investigated the use of a plate tectonics metaphor to generate plausible terrains from a simple vector map representing the location of the main rivers and mountain picks. The method uses a Voronoi tesselation of pick locations to automatically generate tectonic plates which themselves drive terrain folds. Hydraulic erosion is then used to further sculpt the terrain and add details, while the specified rivers are considered to maintain consistency with the input map. This work was published in [27]. A more accurate modeling of large scale fluvial erosion and plates tectonics phenomena was investigated in Guillaume Cordonnier's master thesis and is the object of his PhD, which started in October 2015.

Secondly, we proposed a paint-based interface to tackle the problem of easily populating a terrain with distributions of objects (trees, rocks, grass, houses, etc) or of graph-like structures such as rivers and roads. The key point of our solution is to learn statistics about distributions of elements and their correlation with other distributions, with graph structures, or with terrain slope, and store the resulting histograms as "colors" in a palette interface. After creating a few local distribution manually, the user selects them with a pipette tool, and is able to reuse them with a brush. We also provided a gradient tool to interpolate between two such "colors" an move tool enabling, for instance to move groups of trees and rocks over a terrain while maintaining the adequate correlation with local slope, and a deformation interface based on seam carving enabling to seamlessly stretch or compress a region of virtual world. This work, a collaboration between Arnaud Emilien when defended his phD in December 2014, Ulysse Vimont, Marie-Paule Cani, and Bedrich Benes from Purdue University, was published at Siggraph 2015 [8].

### 6.4.3. Sketching and sculpting Motion

**Participants:**  Marie-Paule Cani, Martin Guay, Kevin Jordao, Rémi Ronfard.

Sketching and sculpting methods were restricted so far to the design of static shapes. One of our research goals has been to extend these interaction metaphors to motion design. This year, this included three specific contributions.
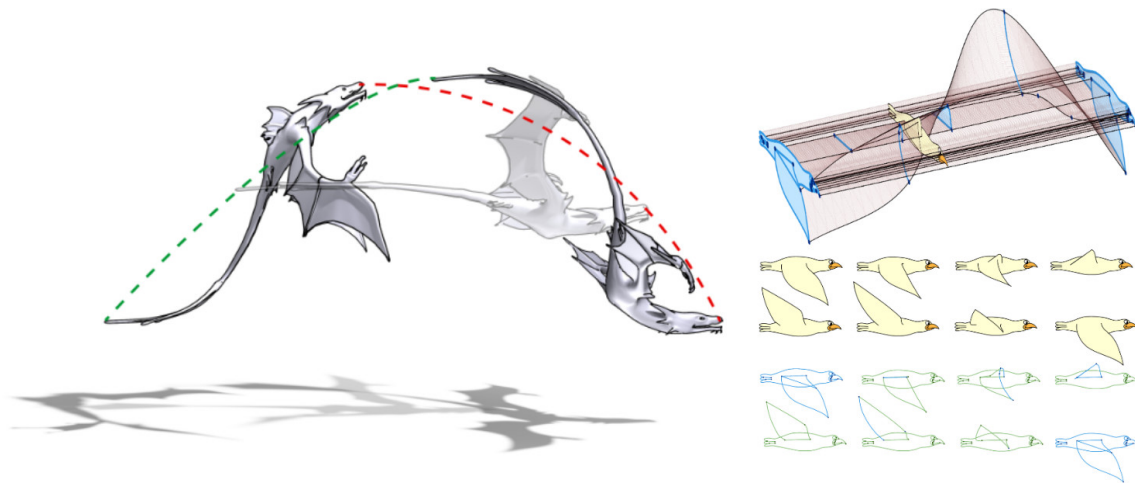
*Figure 16. Left: Result of our Sculpting Motion method [9]. Right: Result of our Vector Graphics method enabling to design vectorial animation [6].*

Firstly, to handle sketch-based representation of motion in the 2D case, we extended the static vector graphics complex data structure, which we had introduced at Siggraph last year, to vector graphics animations with time-varying topology [6]. This second paper was presented at Siggraph again this year. The proposed data structure is augmented with a rich set of editing operations, which can be used to quickly interpolate 2D drawings with different topologies. This work was done within a collaboration with Boris Dalstein and Michiel van de Panne from UBC, Canada.

Secondly, following a first method enabling to sculpt crowd animations (Jordao, Eurographics 2014), we developed a painting interface enabling to specify both density and main directions of motion in an animated crowd. The resulting system is still based on crowd-patches, i.e. the crowd motion is an assembly of local trajectories defined in interconnected patches. Our new painting system, called Crowd-Art, uses discrete changes in loop trajectories to evolve the number of in/out constraints in each patch until the requested density and directions are best matched. See [22]. This concluded Kevin Jordao's PhD thesis, co-advided by Julien Pettre from the MimeTIC team and in collaboration with Marc Christie, defended in December 2015.

Lastly, we developed the first expressive interface to interactively sketch and progressively sculpt and refine character motion. Our solution is based on a space-time sketching metaphor: The user sketches a single space-time stroke, which is used to initialize a series of dynamic lines of action, serving as intermediates to animate the character's model. Motion and shape deformation can be immediately replayed from this single stroke, since it sets at the same time shape, trajectory and speed (defined from the drawing speed). Thanks to visual feedback, the user can easily refine the resulting motion by editing specific lines of actions at fixed times, or by composing several motions together. This work, published at Siggraph, is one of the first methods enabling arbitrary motion to be defined from scratch by a beginner [9]. Together to another work enabling to add dynamics to character motion [21], this concluded Martin Guay's PhD thesis, defended in June 2015.

<p align="center" style="color:red"><strong>LEAR Project-Team</strong></p>

# 7. New Results

## 7.1. Visual recognition in images

### 7.1.1. *Weakly Supervised Object Localization with Multi-fold Multiple Instance Learning*

**Participants:** Ramazan Cinbis, Cordelia Schmid, Jakob Verbeek.

Object category localization is a challenging problem in computer vision. Standard supervised training requires bounding box annotations of object instances. This time-consuming annotation process is sidestepped in weakly supervised learning. In this case, the supervised information is restricted to binary labels that indicate the absence/presence of object instances in the image, without their locations. In [26], we propose to follow a multiple-instance learning approach that iteratively trains the detector and infers the object locations in the positive training images. Our main contribution is a multi-fold multiple instance learning procedure, which prevents training from prematurely locking onto erroneous object locations. Compared to state-of-the-art weakly supervised detectors, our approach better localizes objects in the training images, which translates into improved detection performance. Figure 1 illustrates the iterative object localization process on several example images. The technical report [26] is a journal paper under review after minor revision which extends a previous conference publication by adding experiments with CNN features, and a refinement procedure for the object location inference. These additions improve over related work that has appeared since the publication of the original paper.
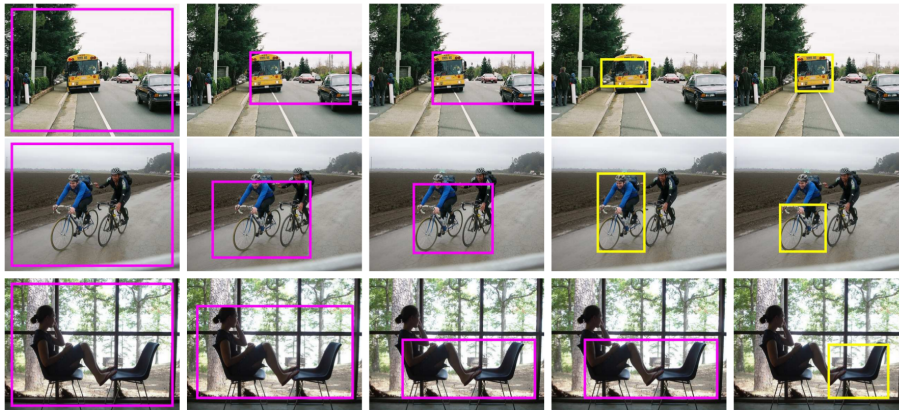


*Figure 1. Illustration of our iterative object localization process on several example images, from initialization (left) to final localization (right). Yellow bounding boxes indicate that the object location hypothesis is in agreement with the ground-truth, for pink boxes the hypothesis is incorrect.*

### 7.1.2. *Patch-level spatial layout for classification and weakly supervised localization*

**Participants:** Valentina Zadrija [University of Zagreb], Josip Krapac [University of Zagreb], Jakob Verbeek, Sinisa Segvic [University of Zagreb].

In [24] we propose a discriminative patch-level spatial layout model suitable for learning object localization models with weak supervision. We start from a block-sparse model of patch appearance based on the normalized Fisher vector representation. The appearance model is responsible for i) selecting a discriminative subset of visual words, and ii) identifying distinctive patches assigned to the selected subset. These patches are further filtered by a sparse spatial model operating on a novel representation of pairwise patch layout. We have evaluated the proposed pipeline in image classification and weakly supervised localization experiments on a public traffic sign dataset. The results show significant advantage of the proposed spatial model over state of the art appearance models.

### 7.1.3. *Approximate Fisher Kernels of non-iid Image Models for Image Categorization*
**Participants:** Ramazan Cinbis, Cordelia Schmid, Jakob Verbeek.

The bag-of-words (BoW) model treats images as sets of local descriptors and represents them by visual word histograms. The Fisher vector (FV) representation extends BoW, by considering the first and second order statistics of local descriptors. In both representations local descriptors are assumed to be identically and independently distributed (iid), which is a poor assumption from a modeling perspective. It has been experimentally observed that the performance of BoW and FV representations can be improved by employing discounting transformations such as power normalization. In [5], an expanded version of a previous conference publication, we introduce non-iid models by treating the model parameters as latent variables which are integrated out, rendering all local regions dependent. Using the Fisher kernel principle we encode an image by the gradient of the data log-likelihood w.r.t. the model hyper-parameters. Our models naturally generate discounting effects in the representations; suggesting that such transformations have proven successful because they closely correspond to the representations obtained for non-iid models. To enable tractable computation, we rely on variational free-energy bounds to learn the hyper-parameters and to compute approximate Fisher kernels. Our experimental evaluation results validate that our models lead to performance improvements comparable to using power normalization, as employed in state-of-the-art feature aggregation methods.

### 7.1.4. *Local Convolutional Features with Unsupervised Training for Image Retrieval*
**Participants:** Mattis Paulin, Matthijs Douze, Zaid Harchaoui, Julien Mairal, Florent Perronnin [Facebook], Cordelia Schmid.

Patch-level descriptors underlie several important computer vision tasks, such as stereo-matching or content-based image retrieval. We introduce a deep convolutional architecture that yields patch-level descriptors, as an alternative to the popular SIFT descriptor for image retrieval. The proposed family of descriptors, called Patch-CKN[17], adapt the recently introduced Convolutional Kernel Network (CKN), an unsupervised framework to learn convolutional architectures. We present a comparison framework to benchmark current deep convolutional approaches along with Patch-CKN for both patch and image retrieval (see Fig. 3 for our pipeline), including our novel "RomePatches" dataset. Patch-CKN descriptors yield competitive results compared to supervised CNNs alternatives on patch and image retrieval.

## 7.2. Learning and statistical models

### 7.2.1. *A Universal Catalyst for First-order Optimization*
**Participants:** Hongzhou Lin, Julien Mairal, Zaid Harchaoui.

In this paper [16], we introduce a generic scheme for accelerating first-order optimization methods in the sense of Nesterov, which builds upon a new analysis of the accelerated proximal point algorithm. Our approach consists of minimizing a convex objective by approximately solving a sequence of well-chosen auxiliary problems, leading to faster convergence. This strategy applies to a large class of algorithms, including gradient descent, block coordinate descent, SAG, SAGA, SDCA, SVRG, Finito/MISO, and their proximal variants. For all of these methods, we provide acceleration and explicit support for non-strongly convex objectives. In addition to theoretical speed-up, we also show that acceleration is useful in practice, as illustrated in Figure 4 , especially for ill-conditioned problems where we measure significant improvements.
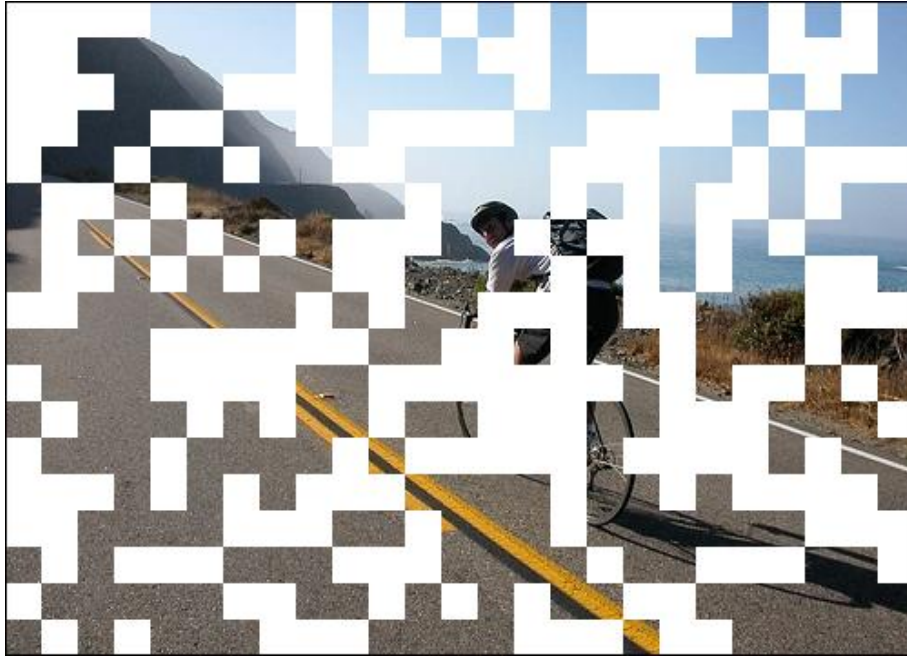
*Figure 2. Illustration of why local image patches are not independent: we can easily guess the image content in the masked areas.*
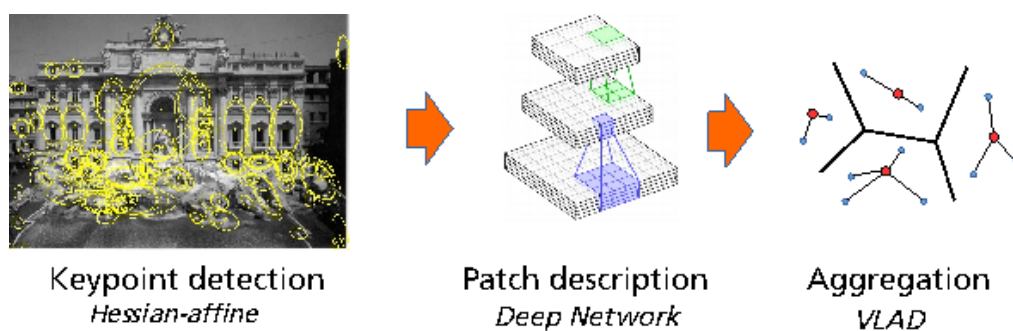


*Figure 3. Image retrieval pipeline. Interest points are extracted with the Hessian-affine detector (left), encoded in descriptor space using convolutional features (middle), and aggregated into a compact representation using VLAD-pooling (right).*
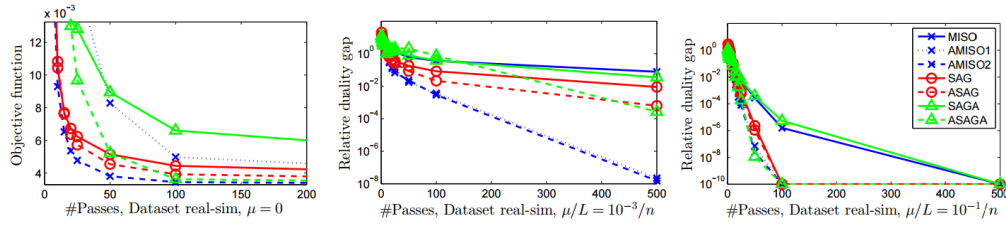
*Figure 4. Objective function value (or duality gap) for different number of passes performed over each dataset. The legend for all curves is on the top right. AMISO, ASAGA, ASAG refer to the accelerated variants of MISO, SAGA, and SAG, respectively.*

### 7.2.2. Incremental Majorization-Minimization Optimization with Application to Large-Scale Machine Learning

**Participant:** Julien Mairal.

In this paper [7], we study optimization methods consisting of iteratively minimizing surrogates of an objective function, as illustrated in Figure 5 . We introduce a new incremental scheme that experimentally matches or outperforms state-of-the-art solvers for large-scale optimization problems typically arising in machine learning.
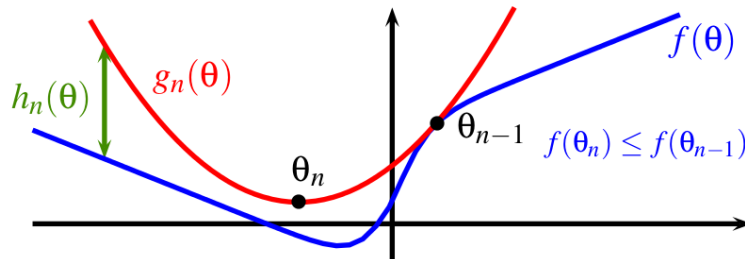


*Figure 5. Illustration of the basic majorization-minimization principle. We compute a surrogate $g_n$ of the objective function $f$ around a current estimate $\theta_{n-1}$. The new estimate $\theta_n$ is a minimizer of $g_n$. The approximation error $h_n$ is smooth.*

### 7.2.3. Coordinated Local Metric Learning

**Participants:** Shreyas Saxena, Jakob Verbeek.

Mahalanobis metric learning amounts to learning a linear data projection, after which the $\ell_2$ metric is used to compute distances. In [20], we develop local metric learning techniques which allow more flexible metrics, not restricted to linear projections, see 6 . Most of these methods partition the data space using clustering, and for each cluster a separate metric is learned. Using local metrics, however, it is not clear how to measure distances between data points assigned to different clusters. In this paper we propose to embed the local metrics in a global low-dimensional representation, in which the $\ell_2$ metric can be used. With each cluster we associate a

linear mapping that projects the data to the global representation. This global representation directly allows computing distances between points regardless to which local cluster they belong. Moreover, it also enables data visualization in a single view, and the use of $\ell_2$-based efficient retrieval methods. Experiments on the Labeled Faces in the Wild dataset show that our approach improves over previous global and local metric learning approaches.
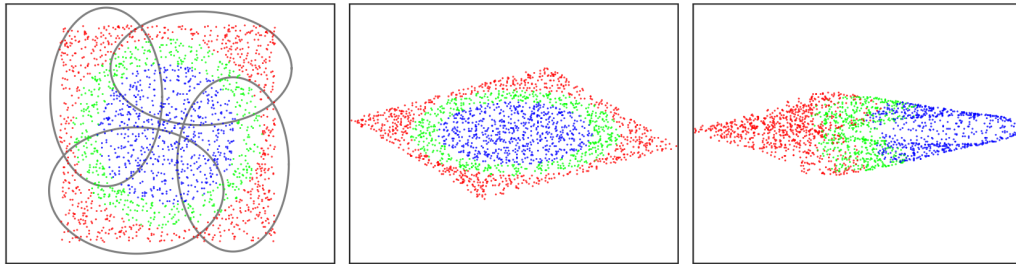


*Figure 6. Synthetic dataset with color coded class labels, and the GMM used by our CLML local metric (left). Data projection given by a global Mahalanobis metric (middle) and our local CLML metric (right). The pairwise training constraints are better respected by CLML.*

### 7.2.4. A convex formulation for joint RNA isoform detection and quantification from multiple RNA-seq samples

**Participants:** Elsa Bernard [Institut Curie, Ecoles des Mines-ParisTech], Laurent Jacob [CNRS, LBBE Laboratory], Julien Mairal, Jean-Philippe Vert [Institut Curie, Ecoles des Mines-ParisTech].

Detecting and quantifying isoforms from RNA-seq data is an important but challenging task. The problem is often ill-posed, particularly at low coverage. One promising direction is to exploit several samples simultaneously. In this paper [4], we propose a new method for solving the isoform deconvolution problem jointly across several samples. We formulate a convex optimization problem that allows to share information between samples and that we solve efficiently, as illustrated in Figure 7 . We demonstrate the benefits of combining several samples on simulated and real data, and show that our approach outperforms pooling strategies and methods based on integer programming. Our convex formulation to jointly detect and quantify isoforms from RNA-seq data of multiple related samples is a computationally efficient approach to leverage the hypotheses that some isoforms are likely to be present in several samples. The software and source code are available at http://cbio.ensmp.fr/flipflop.

### 7.2.5. Adaptive Recovery of Signals by Convex Optimization

**Participants:** Zaid Harchaoui, Anatoli Juditsky [Univ. Grenoble], Arkadi Nemirovski [Georgia Tech], Dimitry Ostrovsky [Univ. Grenoble].

In [13], we present a theoretical framework for adaptive estimation and prediction of signals of unknown structure in the presence of noise. The framework allows to address two intertwined challenges: (i) designing optimal statistical estimators; (ii) designing efficient numerical algorithms. In particular, we establish oracle inequalities for the performance of adaptive procedures, which rely upon convex optimization and thus can be efficiently implemented. As an application of the proposed approach, we consider denoising of harmonic oscillations

### 7.2.6. Semi-proximal Mirror-Prox for Nonsmooth Composite Minimization
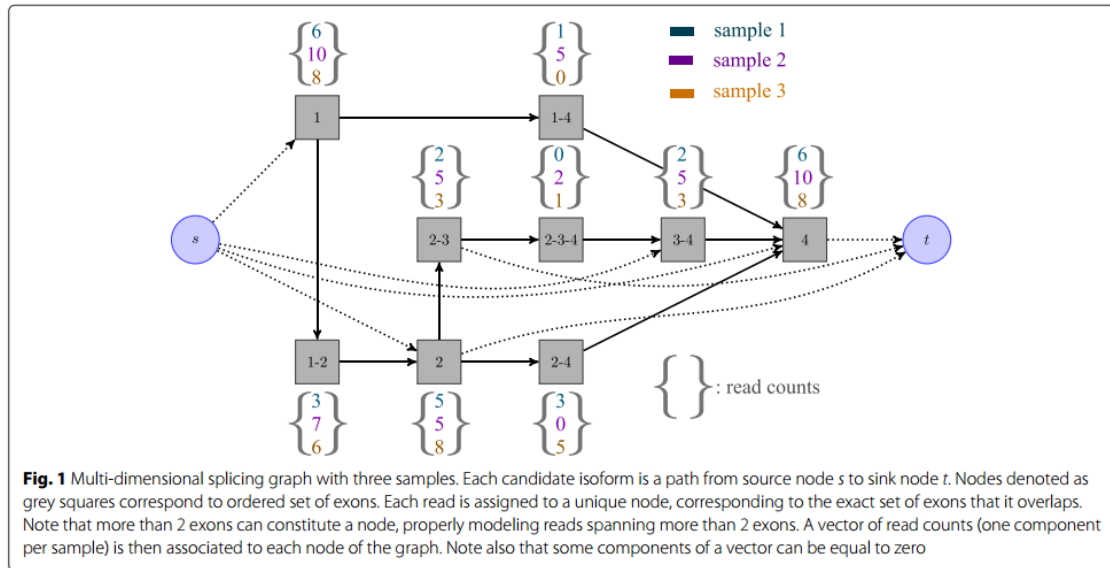
**Participants:** Niao He [Georgia Tech], Zaid Harchaoui.

**Fig. 1** Multi-dimensional splicing graph with three samples. Each candidate isoform is a path from source node *s* to sink node *t*. Nodes denoted as grey squares correspond to ordered set of exons. Each read is assigned to a unique node, corresponding to the exact set of exons that it overlaps. Note that more than 2 exons can constitute a node, properly modeling reads spanning more than 2 exons. A vector of read counts (one component per sample) is then associated to each node of the graph. Note also that some components of a vector can be equal to zero

*Figure 7. Graph on which we perform network flow optimization. Nodes represent observed reads, and paths on the graph correspond to isoforms.*

In [28], we propose a new first-order optimisation algorithm to solve high-dimensional non-smooth composite minimisation problems. Typical examples of such problems have an objective that decomposes into a non-smooth empirical risk part and a non-smooth regularisation penalty. The proposed algorithm, called Semi-Proximal Mirror-Prox, leverages the Fenchel-type representation of one part of the objective while handling the other part of the objective via linear minimization over the domain. The algorithm stands in contrast with more classical proximal gradient algorithms with smoothing, which require the computation of proximal operators at each iteration and can therefore be impractical for high-dimensional problems. We establish the theoretical convergence rate of Semi-Proximal Mirror-Prox, which exhibits the optimal complexity bounds, for the number of calls to linear minimization oracle. We present promising experimental results showing the interest of the approach in comparison to competing methods.

## 7.3. Recognition in video

### 7.3.1. *Beat-Event Detection in Action Movie Franchises*

**Participants:** Danila Potapov, Matthijs Douze, Jerome Revaud, Zaid Harchaoui, Cordelia Schmid.

While important advances were recently made towards temporally localizing and recognizing specific human actions or activities in videos, efficient detection and classification of long video chunks belonging to semantically-defined categories such as "pursuit" or "romance" remains challenging.

In our work [30], we introduce a new dataset, Action Movie Franchises, consisting of a collection of Hollywood action movie franchises. We define 11 non-exclusive semantic categories — called beat-categories — that are broad enough to cover most of the movie footage. The corresponding beat-events are annotated as groups of video shots, possibly overlapping. We propose an approach for localizing beat-events based on classifying shots into beat-categories and learning the temporal constraints between shots, as shown in Figure 8 . We show that temporal constraints significantly improve the classification performance. We set up an evaluation protocol for beat-event localization as well as for shot classification, depending on whether movies from the same franchise are present or not in the training data.

*Figure 8. A 5-minute extract from the proposed Action Movie Franchises dataset, ground truth annotation and output of different methods. Each color stands for a different event category: green —pursuit, blue —battle, yellow —victory-good, green —despair-good, pink —romance, gray —victory-bad, cadet blue —good-argue-good. Hashes mark difficult examples. The color code for the classifier evaluation is: white = true positive, gray = ignored, black = false positive.*

### 7.3.2. EpicFlow: Edge-Preserving Interpolation of Correspondences for Optical Flow

**Participants:**  Jerome Revaud, Philippe Weinzaepfel, Zaid Harchaoui, Cordelia Schmid.

In this paper [18], we propose a novel approach for optical flow estimation, targeted at large displacements with significant occlusions. It consists of two steps: i) dense matching by edge-preserving interpolation from a sparse set of matches; ii) variational energy minimization initialized with the dense matches. The sparse-to-dense interpolation relies on an appropriate choice of the distance, namely an edge-aware geodesic distance. This distance is tailored to handle occlusions and motion boundaries – two common and difficult issues for optical flow computation. We also propose an approximation scheme for the geodesic distance to allow fast computation without loss of performance. Subsequent to the dense interpolation step, standard one-level variational energy minimization is carried out on the dense matches to obtain the final flow estimation. The proposed approach, called Edge-Preserving Interpolation of Correspondences *(EpicFlow)* is fast and robust to large displacements. An overview is given in Figure 9 . EpicFlow significantly outperforms the state of the art on MPI-Sintel and performs on par on Kitti and Middlebury.
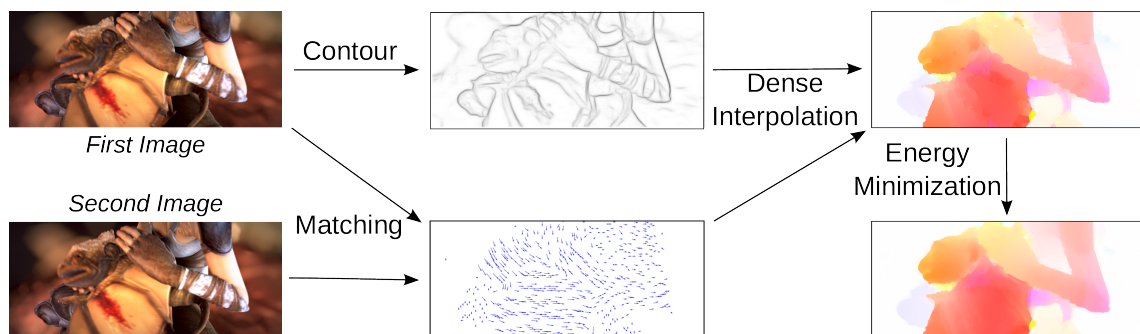


*Figure 9. Overview of EpicFlow. Given two images, we compute matches using DeepMatching and the edges of the first image using SED. We combine these two cues to densely interpolate matches and obtain a dense correspondence field. This is used as initialization of a one-level energy minimization framework.*

### 7.3.3. *DeepMatching: Hierarchical Deformable Dense Matching*

**Participants:**  Jerome Revaud, Philippe Weinzaepfel, Zaid Harchaoui, Cordelia Schmid.

In this paper [31], we introduce a novel matching algorithm, called DeepMatching, to compute dense correspondences between images. DeepMatching relies on a hierarchical, multi-layer, correlational architecture designed for matching images and was inspired by deep convolutional approaches, see Figure 10 . The proposed matching algorithm can handle non-rigid deformations and repetitive textures and efficiently determines dense correspondences in the presence of significant changes between images. We evaluate the performance of DeepMatching, in comparison with state-of-the-art matching algorithms, on the Mikolajczyk, the MPI-Sintel and the Kitti datasets. DeepMatching outperforms the state-of-the-art algorithms and shows excellent results in particular for repetitive textures. We also propose a method for estimating optical flow, called DeepFlow, by integrating DeepMatching in the large displacement optical flow (LDOF) approach of Brox et al. Compared to existing matching algorithms, additional robustness to large displacements and complex motion is obtained thanks to our matching approach. DeepFlow obtains competitive performance on public benchmarks for optical flow estimation.
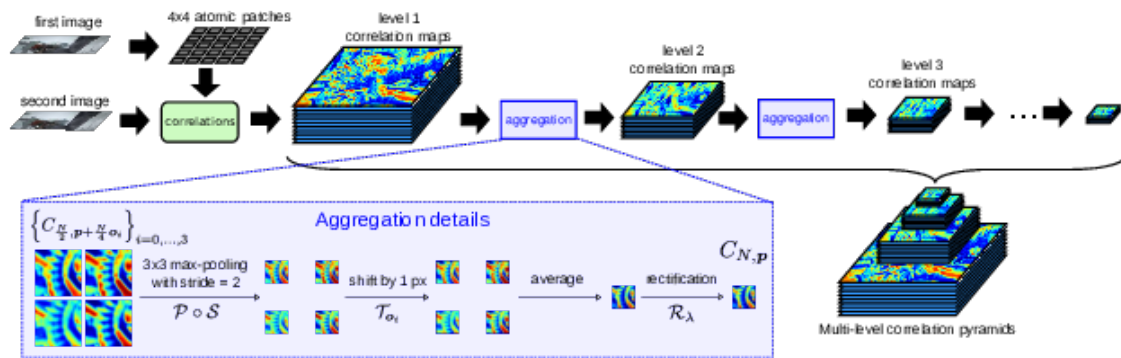


*Figure 10. Overview of the bottom-up part of DeepMatching, which builds the multi-level correlation pyramid, from which matches are then extracted.*

### 7.3.4. *Learning to Detect Motion Boundaries*

**Participants:**  Philippe Weinzaepfel, Jerome Revaud, Zaid Harchaoui, Cordelia Schmid.

In this paper [23], we propose a learning-based approach for motion boundary detection. Precise localization of motion boundaries is essential for the success of optical flow estimation, as motion boundaries correspond to discontinuities of the optical flow field. The proposed approach allows to predict motion boundaries, using a structured random forest trained on the ground-truth of the MPI-Sintel dataset, see Figure 11 . The random forest leverages several cues at the patch level, namely appearance (RGB color) and motion cues (optical flow estimated by state-of-the-art algorithms). Experimental results show that the proposed approach is both robust and computationally efficient. It significantly outperforms state-of-the-art motion-difference approaches on the MPI-Sintel and Middlebury datasets. We compare the results obtained with several state-of-the-art optical flow approaches and study the impact of the different cues used in the random forest. Furthermore, we introduce a new dataset, the YouTube Motion Boundaries dataset (**YMB**), that comprises 60 sequences taken from real-world videos with manually annotated motion boundaries. On this dataset, our approach, although trained on MPI-Sintel, also outperforms by a large margin state-of-the-art optical flow algorithms.

### 7.3.5. *Learning to track for spatio-temporal action localization*

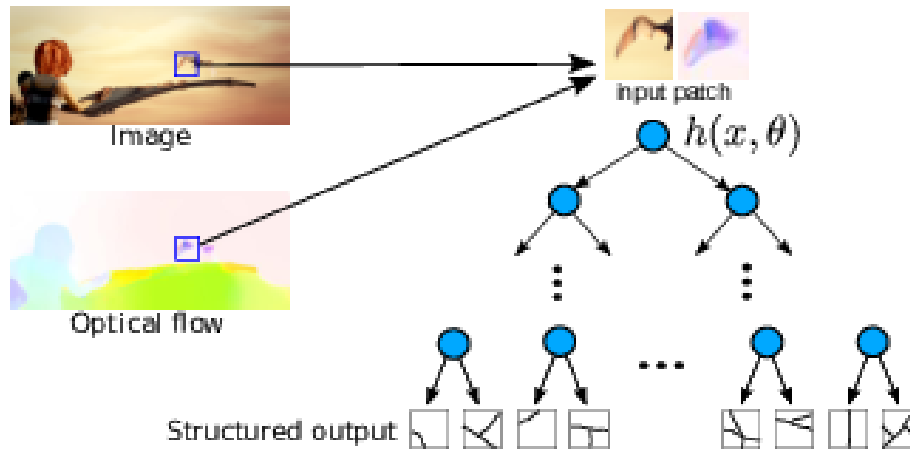**Participants:**  Philippe Weinzaepfel, Zaid Harchaoui, Cordelia Schmid.

*Figure 11. A structured random forest, taking as input a patch of an image and corresponding optical flow, outputs a motion boundaries patch. These motion boundaries patches are then aggregated to build the motion boundaries map for the whole image.*

In this paper [22], we propose an effective approach for spatio-temporal action localization in realistic videos. The approach first detects proposals at the frame-level and scores them with a combination of static and motion CNN features. It then tracks high-scoring proposals throughout the video using a tracking-by-detection approach. Our tracker relies simultaneously on instance-level and class-level detectors. The tracks are scored using a spatio-temporal motion histogram, a descriptor at the track level, in combination with the CNN features. Finally, we perform temporal localization of the action using a sliding-window approach at the track level. An overview of our approach is given in Figure 12 . We present experimental results for spatio-temporal localization on the UCF-Sports, J-HMDB and UCF-101 action localization datasets, where our approach outperforms the state of the art with a margin of 15%, 7% and 12% respectively in mAP.

### 7.3.6. *A robust and efficient video representation for action recognition*
**Participants:**  Heng Wang, Dan Oneata, Cordelia Schmid, Jakob Verbeek.

In [9] we present a state-of-the-art video representation and apply it to efficient action recognition and detection. We first propose to improve the popular dense trajectory features by explicit camera motion estimation. Local feature trajectories consistent with the homography are considered as due to camera motion, and thus removed. This results in significant improvement on motion-based HOF and MBH descriptors. We further explore the recent Fisher vector as an alternative feature encoding approach to the standard bag-of-words histogram, and consider different ways to include spatial layout information in these encodings. We present a large and varied set of evaluations, considering (i) classification of short basic actions on six datasets, (ii) localization of such actions in featurelength movies, and (iii) large-scale recognition of complex events. We find that our improved trajectory features significantly outperform previous dense trajectories, and that Fisher vectors are superior to bag-of-words encodings for video recognition tasks. In all three tasks, we show substantial improvements over the state-of-the-art results. This journal paper combines and extends earlier conference papers.

### 7.3.7. *Circulant temporal encoding for video retrieval and temporal alignment*
**Participants:**  Jerome Revaud, Matthijs Douze, Hervé Jégou [Inria Rennes, Facebook AI Research], Cordelia Schmid, Jakob Verbeek.
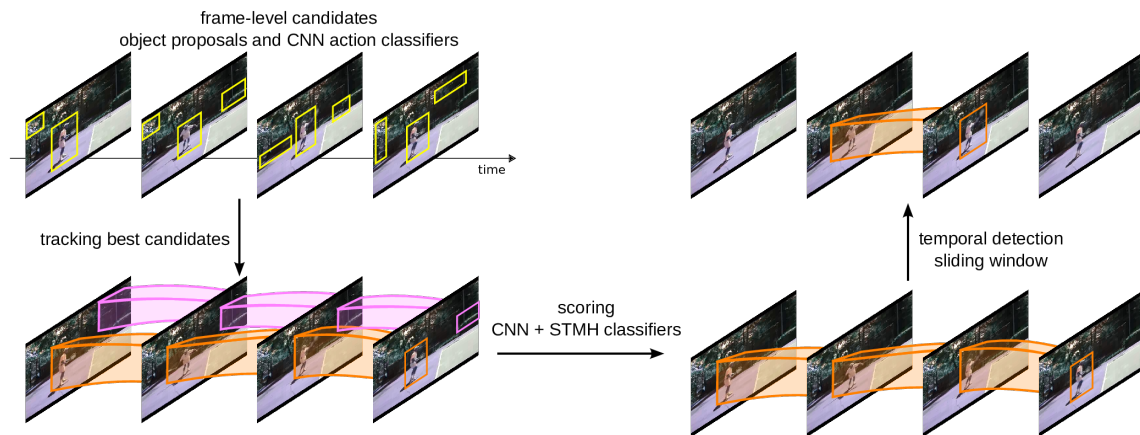
*Figure 12. Overview of our spatio-temporal action localization approach. We detect frame-level object proposals and score them with CNN action classifiers. The best candidates, in term of scores, are tracked throughout the video. We then score the tracks with CNN and spatio-temporal motion histogram (STMH) classifiers. Finally, we perform a temporal sliding window for detecting the temporal extent of the action.*

In [6] we address the problem of specific video event retrieval. Given a query video of a specific event, e.g., a concert of Madonna, the goal is to retrieve other videos of the same event that temporally overlap with the query. Our approach encodes the frame descriptors of a video to jointly represent their appearance and temporal order. It exploits the properties of circulant matrices to efficiently compare the videos in the frequency domain. This offers a significant gain in complexity and accurately localizes the matching parts of videos. The descriptors can be compressed in the frequency domain with a product quantizer adapted to complex numbers. In this case, video retrieval is performed without decompressing the descriptors. The second problem we consider is the temporal alignment of a set of videos. We exploit the matching confidence and an estimate of the temporal offset computed for all pairs of videos by our retrieval approach. Our robust algorithm aligns the videos on a global timeline by maximizing the set of temporally consistent matches. The global temporal alignment enables synchronous playback of the videos of a given scene. This journal paper extends an earlier conference paper.

### 7.3.8. Pose Estimation and Segmentation of Multiple People in Stereoscopic Movies

**Participants:** Guillaume Seguin [Willow], Karteek Alahari, Josef Sivic [Willow], Ivan Laptev [Willow].

The work in [8] presents a method to obtain a pixel-wise segmentation and pose estimation of multiple people in stereoscopic videos, as shown in Figure 13 . This task involves challenges such as dealing with unconstrained stereoscopic video, non-stationary cameras, and complex indoor and outdoor dynamic scenes with multiple people. We cast the problem as a discrete labelling task involving multiple person labels, devise a suitable cost function, and optimize it efficiently. The contributions of our work are two-fold: First, we develop a segmentation model incorporating person detections and learnt articulated pose segmentation masks, as well as colour, motion, and stereo disparity cues. The model also explicitly represents depth ordering and occlusion. Second, we introduce a stereoscopic dataset with frames extracted from feature-length movies "StreetDance 3D" and "Pina". The dataset contains 587 annotated human poses, 1158 bounding box annotations and 686 pixel-wise segmentations of people. The dataset is composed of indoor and outdoor scenes depicting multiple people with frequent occlusions. We demonstrate results on our new challenging dataset, as well as on the H2view dataset from (Sheasby et al. ACCV 2012).

*Figure 13. We segment multiple people in the scene, estimate their poses and relative front-to-back order, denoted by the numbers in the image below, in every frame of a video sequence.*

### 7.3.9. Encoding Feature Maps of CNNs for Action Recognition

**Participants:** Xiaojiang Peng, Cordelia Schmid.

In [29] We describe our approach for action classification in the THUMOS Challenge 2015. Our approach is based on two types of features, improved dense trajectories and CNN features, as illustrated in Figure 14 . For trajectory features, we extract HOG, HOF, MBHx, and MBHy descriptors and apply Fisher vector encoding. For CNN features, we utilize a recent deep CNN model, VGG19, to capture appearance features and use VLAD encoding to encode/pool convolutional feature maps which shows better performance than average pooling of feature maps and full-connected activation features.
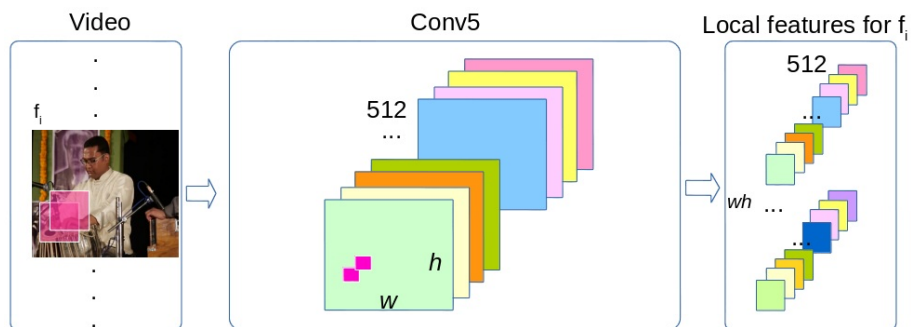


*Figure 14. Local features from convolutional feature maps. Each pixel (pink square in the middle image) in the Conv5 feature map is actually a feature for the corresponding patch in original frame. We obtain w\*h 512-D features for frame $f_i$.*

### 7.3.10. Online Object Tracking with Proposal Selection

**Participants:** Yang Hua, Karteek Alahari, Cordelia Schmid.

Tracking-by-detection approaches are some of the most successful object trackers in recent years. Their success is largely determined by the detector model they learn initially and then update over time. However, under challenging conditions where an object can undergo transformations, e.g., severe rotation, these methods are found to be lacking. In [14], we address this problem by formulating it as a proposal selection task and making two contributions. The first one is introducing novel proposals estimated from the geometric transformations undergone by the object, and building a rich candidate set for predicting the object location. The second one is devising a novel selection strategy using multiple cues, i.e., detection score and edgeness score computed from state-of-the-art object edges and motion boundaries. We extensively evaluate our approach on the visual object tracking 2014 challenge and online tracking benchmark datasets, and show the best performance. Sample results are shown in Figure 15 . Our tracker based on this method has recently won the visual object tracking challenge (VOT-TIR) organized as part of ICCV 2015 in Santiago, Chile.
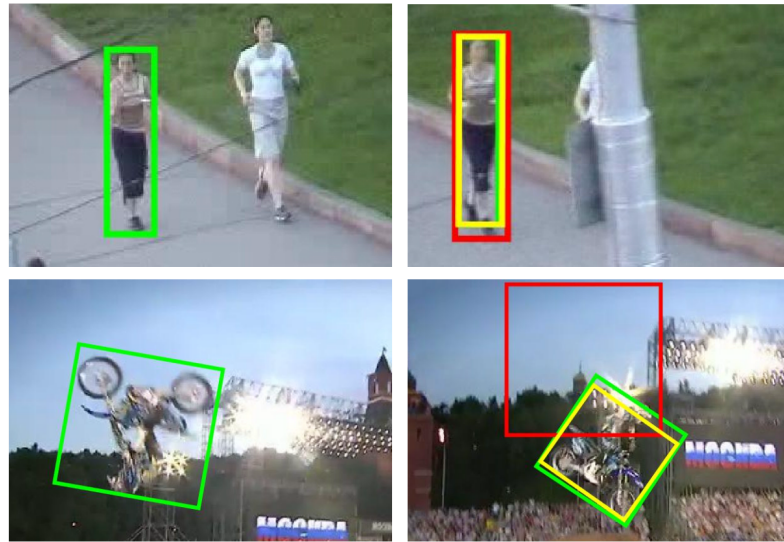
*Figure 15. Sample frames (cropped) from the jogging (top row) and motocross (bottom row) sequences. The ground truth annotation (green) in the first frame (left) is used to train our tracker and the winner of VOT2014 challenge. We show these two tracking results (right) on another frame in the sequence. Our method (yellow) successfully tracks objects undergoing deformations unlike winner of VOT2014 challenge (red).*

<p style="text-align:center;color:red;font-weight:bold;">MAVERICK Project-Team</p>

# 6. New Results

## 6.1. Single Scattering in participating media with refractive boundaries

**Participant:** Nicolas Holzschuch [contact].



**(a)** Our algorithm (169 s)    **(b)** Point samples [WZHB09]    **(c)** Photon mapping
(27 samples, 170 s)    (650 000 photons, 167 s)
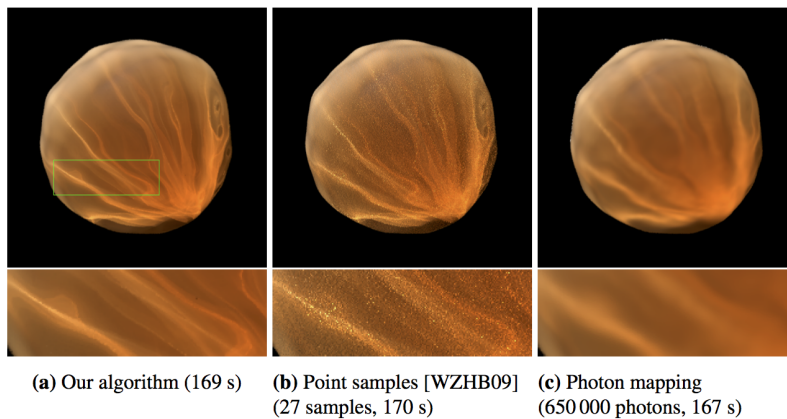
*Figure 2. Single scattering: comparison between our algorithm and existing methods (equal computation time) on a translucent sphere illuminated by a point light source from behind.*

Volume caustics are high-frequency effects appearing in participating media with low opacity, when refractive interfaces are focusing the light rays (see Figure 2 ). Refractions make them hard to compute, since screen locality does not correlate with spatial locality in the medium. We have developed a new method for accurate computation of single scattering effects in a participating media enclosed by refractive interfaces. Our algorithm is based on the observation that although radiance along each camera ray is irregular, contributions from individual triangles are smooth. Our method gives more accurate results than existing methods, faster. It uses minimal information and requires no precomputation or additional data structures. This paper was published in the *Computer Graphics Forum* journal [3] and presented at the *Eurographics Symposium on Rendering*.

## 6.2. Diffraction effects in reflectance properties

**Participant:** Nicolas Holzschuch [contact].

Reflectance properties express how objects in a virtual scene interact with light. They control the appearance of the object: whether it looks shiny or not, it has a metallic or plastic appearance. The reflectance model (BRDF) is essential for photorealistic pictures. Measured reflectance provide high realism, at the expense of memory cost. Parametric models are compact, but it is difficult to find the right parameters from measured reflectance.
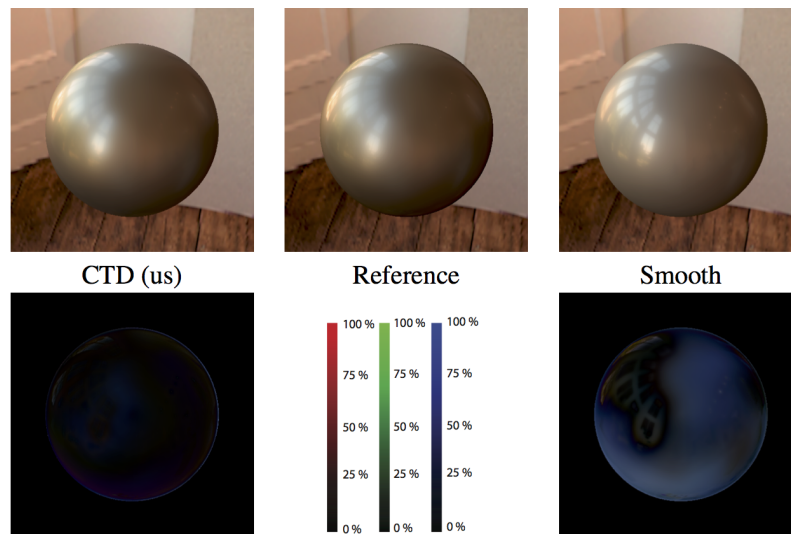
*Figure 3. Comparison between our model combining reflection and diffraction (left), measured data (center) and state-of-the-art model (right).*

Many parametric models are based on a physical representation of the surface micro-geometry and how it interacts with incoming light. The Cook-Torrance model assumes that light follows the principles of optical geometry: it is reflected by the surface micro-geometry but also potentially occluded. The diffraction model assumes that the micro-geometry diffracts the incoming light. This reflectance model has an intrinsic wavelength dependency. Previous experiments have shown that fitting measured materials to parametric models is hard. Heuristic models based on either Cook-Torrance or diffraction are complex, with many parameters. Our research has shown that both effects (optical geometry and diffraction) are present in most measured materials [6]. Based on this knowledge, we have proposed a new reflectance model, that accurately represents measured reflectance [10]. This model combines optical geometry for the specular peak and diffraction effects for wide-angle scattering.

## 6.3. Efficient and Accurate Spherical Kernel Integrals using Isotropic Decomposition

**Participant:** Cyril Soler [contact].

Spherical filtering is fundamental to many problems in image synthesis, such as computing the reflected light over a surface or anti-aliasing mirror reflections over a pixel. This operation is challenging since the profile of spherical filters (e.g., the view-evaluated BRDF or the geometry-warped pixel footprint, above) typically exhibits both spatial-and-rotational-variation at each pixel, precluding precomputed solutions. We accelerate complex spherical filtering tasks using isotropic spherical decomposition (ISD), decomposing spherical filters into a linear combination of simpler isotropic kernels. Our general ISD is flexible to the choice of the isotropic kernels, and we demonstrate practical realizations of ISD on several problems in rendering: shading and prefiltering with spatially-varying BRDFs, anti-aliasing environment mapped mirror reflections, and filtering of noisy reflectance data. Compared to previous basis-space rendering solutions, our shading solution generates ground truth-quality results at interactive rates, avoiding costly reconstruction and large approximation errors. This paper was published in *ACM Transactions on Graphics* [4] and presented at Siggraph Asia 2015.
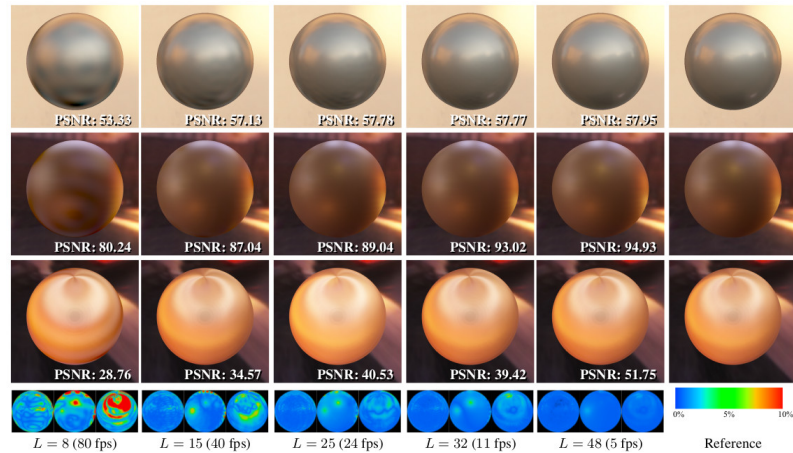
*Figure 4. Convergence of spherical ISD shading for increasing L. Top to bottom: isotropic alum-bronze (with* `pisa` *illumination), isotropic gold-paint and anisotropic yellow-satin (both using* `grace cathedral` *illumination). Reference images were ray-traced using 300K samples per pixel. False color error images in the bottom row visually illustrate the convergence of our RZH approximation to the reference rendering. Note that the dark region in the center of the spheres on the last row of renderings is indeed part of the underlying reflectance input data.*

## 6.4. Color transfer guided by summary statistics

**Participants:** Benoit Arbelot, Thomas Hurtut, Romain Vergne [contact], Joëlle Thollot.



Color transfer                                    Colorization

*Figure 5. Our framework allows for automatic local color transfer (left) and colorization (right) based on textural properties.*

We have targeted two related color manipulation problems: *Color transfer* for modifying an image colors and *colorization* for adding colors to a greyscale image. Automatic methods for these two applications propose to modify the input image using a reference that contains the desired colors. Previous approaches usually do not target both applications and suffer from two main limitations: possible misleading associations between input and reference regions and poor spatial coherence around image structures. In this paper, we propose a unified framework that uses the textural content of the images to guide the color transfer and colorization. Our method introduces an edge-aware texture descriptor based on region covariance, allowing for local color

transformations. We show that our approach is able to produce results comparable or better than state-of-the-art methods in both applications. This work was presented at the AFIG conference [7]. An extended version is available as a research report [9].

## 6.5. Programmable 2D Arrangements for Element Texture Design

**Participants:** Hugo Loi, Thomas Hurtut, Romain Vergne, Joëlle Thollot [contact].

We introduce a programmable method for designing stationary 2D arrangements for element textures, namely textures made of small geometric elements. These textures are ubiquitous in numerous applications of computer-aided illustration. Previous methods, whether they be example-based or layout-based, lack control and can produce a limited range of possible arrangements. Our approach targets technical artists who will design an arrangement by writing a script. These scripts are using three types of operators: partitioning operators for defining the broad-scale organization of the arrangement, mapping operators for controlling the local organization of elements, and merging operators for mixing different arrangements. These operators are designed so as to guarantee a stationary result meaning that the produced arrangements will always be repetitive. We show (see Figure 10 ) that this simple set of operators is sufficient to reach a much broader variety of arrangements than previous methods. Editing the script leads to predictable changes in the synthesized arrangement, which allows an easy iterative design of complex structures. Finally, our operator set is extensible and can be adapted to application-dependent needs. This work is available as a research report [11]

## 6.6. Piecewise polynomial Reconstruction of Scalar Fields from Simplified Morse-Smale Complexes

**Participants:** Léo Allemand-Giorgis, Georges-Pierre Bonneau [contact].

Morse-Smale (MS) complexes have been proposed to visualize topological features of scalar fields defined on manifold domains. Herein, three main problems have been addressed in the past: (a) efficient computation of the initial combinatorial structure connecting the critical points; (b) simplification of these combinatorial structures; (c) reconstruction of a scalar field in accordance to the simplified Morse-Smale complex. The present paper faces the third problem by proposing a novel approach for computing a scalar field coherent with a given simplified MS complex that privileges the use of piecewise polynomial functions. Based on techniques borrowed from shape preserving design in Computer Aided Geometric Design, our method constructs the surface cell by cell using piecewise polynomial curves and surfaces. We present the benefit and limitations of using polynomials for reconstruction surfaces from topological data. This research was published in a book chapter [8].
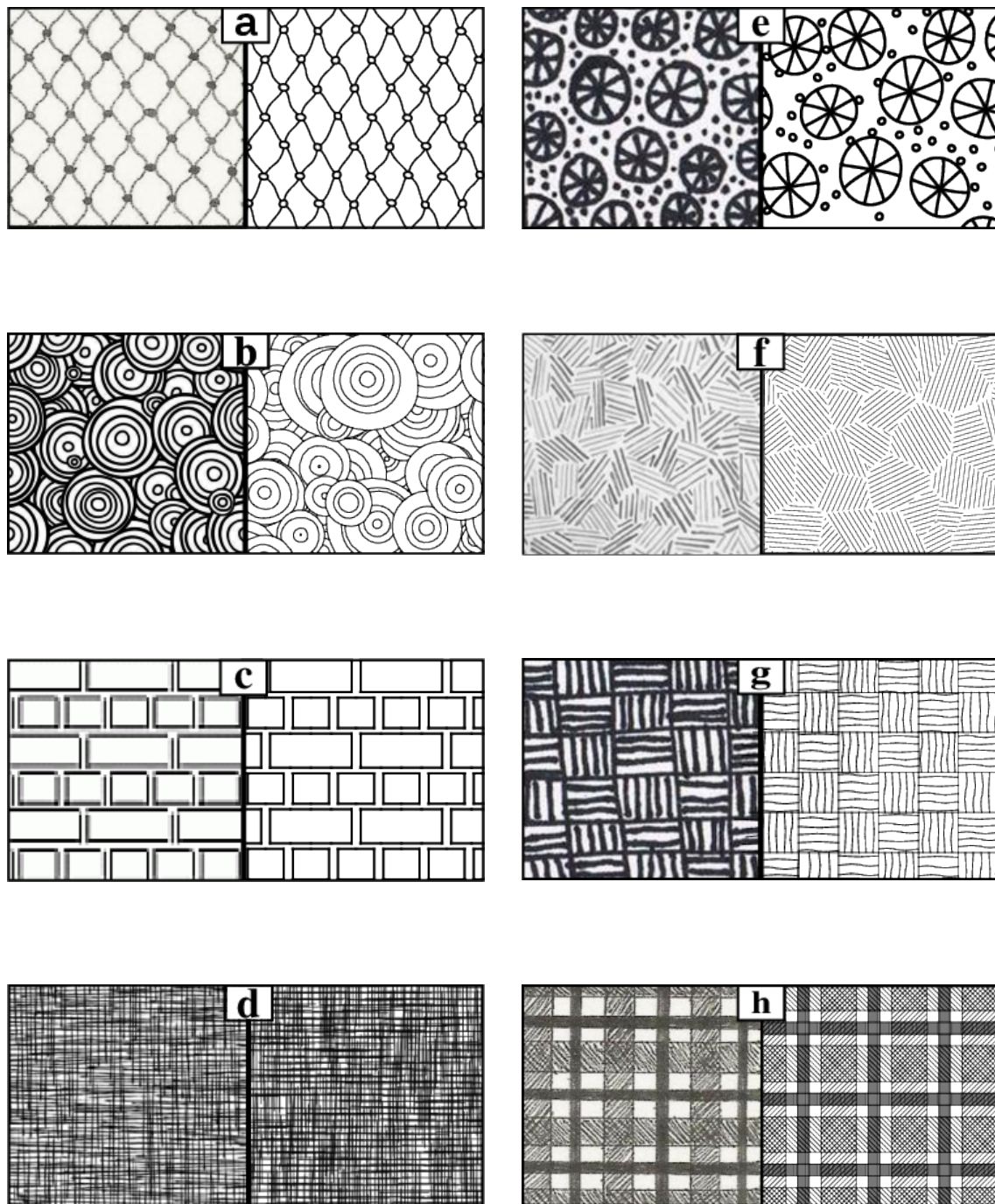
*Figure 6. **Element textures commonly used.** These textures can be found in professional art (d,g,h), casual art (a,e,f), technical productions such as Computer-Assisted Design illustration tools (c), and textile industry (b). For each example, we show a hand-drawn image (left), and our synthesized reproduction of its geometric arrangement (right). (a,b,c) Classic regular distributions with contact, overlap and no adjacency between elements respectively. (d) Overlap of two textures creating cross hatching. (e) Non overlapping combination of two textures. (f,g,h) Complex element textures with clusters of elements. — Image credit: (d,g,h) "Rendering in Pen and Ink: The Classic Book On Pen and Ink Techniques for Artists, Illustrators, Architects, and Designers" ; (a,e) Profusion Art [profusionart.blogspot.com ]; (f) Hayes' Art Classes [hayesartclasses.blogspot.com ]; (c) CompugraphX [www.compugraphx.com ]; (b) 123Stitch [www.123stitch.com ].*
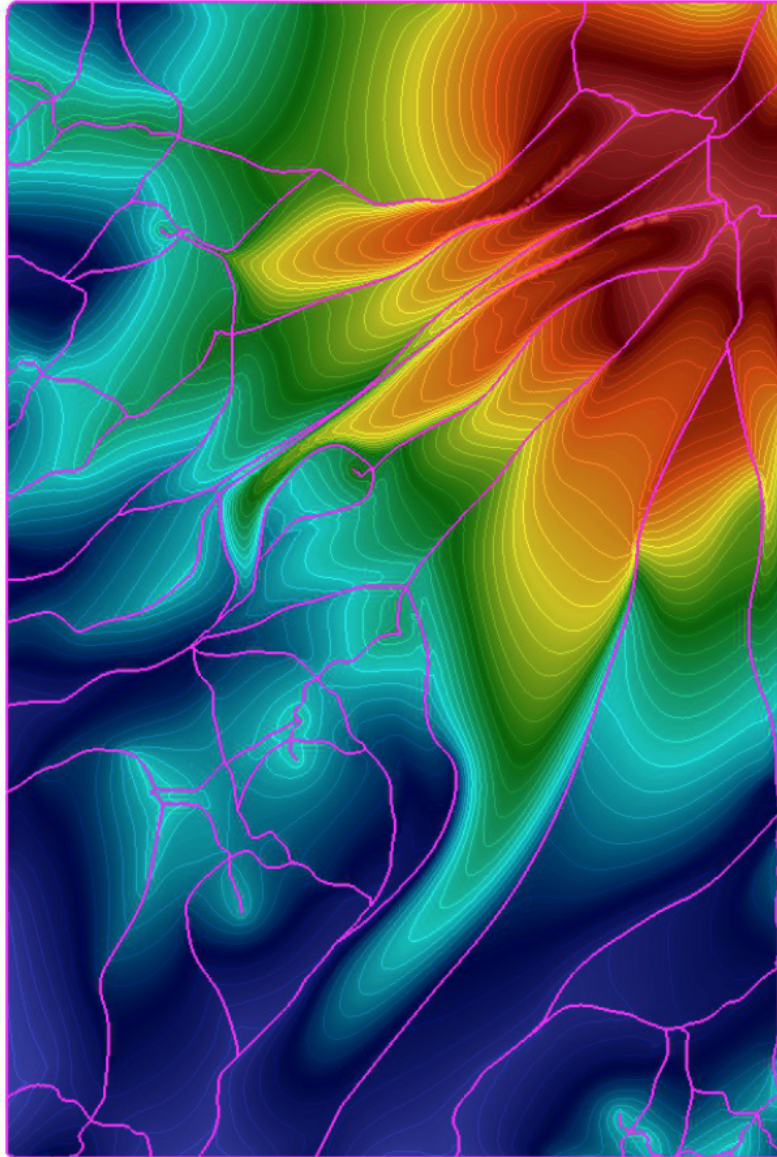
*Figure 7. A function is reconstructed from its Morse-Smale complex (in purple). Inside the cells the function is monotonic so that no critical points are inserted, as can be seen from the isocontours in white. This technique is useful in Visualization whenever critical points in the data are important.*

# 7. New Results

## 7.1. QuickCSG: Arbitrary and Faster Boolean Combinations of N Solids

While studied over several decades, the computation of boolean operations on polyhedra is almost always addressed by focusing on the case of two polyhedra. For multiple input polyhedra and an arbitrary boolean operation to be applied, the operation is decomposed over a binary CSG tree, each node being processed separately in quasilinear time. For large trees, this is both error prone due to intermediate geometry and error accumulation, and inefficient because each node yields a specific overhead. We introduce a fundamentally new approach to polyhedral CSG evaluation, addressing the general N-polyhedron case. We propose a new vertex-centric view of the problem, which both simplifies the algorithm computing resulting geometric contributions, and vastly facilitates its spatial decomposition. We then embed the entire problem in a single KD-tree, specifically geared toward the final result by early pruning of any region of space not contributing to the final surface. This not only improves the robustness of the approach, it also gives it a fundamental speed advantage, with an output complexity depending on the output mesh size instead of the input size as with usual approaches. Complemented with a task-stealing parallelization, the algorithm achieves breakthrough performance, one to two orders of magnitude speedups with respect to state-of-the-art CPU algorithms, on boolean operations over two to several dozen polyhedra. The algorithm is also shown to outperform recent GPU implementations and approximate discretizations, while producing a topologically exact output without redundant facets. This algorithm was published as Inria research report [16].
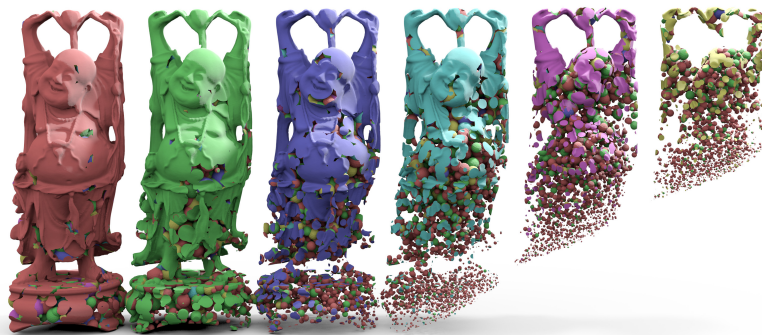


*Figure 3. Intersection of 6 Buddhas with the union of 100,000 spheres (total 24 million triangles). Computed in 8 seconds on a desktop machine [16]*

## 7.2. An Efficient Volumetric Framework for Shape Tracking

Recovering 3D shape motion using visual information is an important problem with many applications in computer vision and computer graphics, among other domains. Most existing approaches rely on surface-based strategies, where surface models are fit to visual surface observations. While numerically plausible, this paradigm ignores the fact that the observed surfaces often delimit volumetric shapes, for which deformations are constrained by the volume inside the shape. Consequently, surface-based strategies can fail when the observations define several feasible surfaces, whereas volumetric considerations are more restrictive with

respect to the admissible solutions. In this work, we investigate a novel volumetric shape parametrization to track shapes over temporal sequences. In constrast to Eulerian grid discretizations of the observation space, such as voxels, we consider general shape tesselations yielding more convenient cell decompositions, in particular the Centroidal Voronoi Tesselation. With this shape representation, we devise a tracking method that exploits volumetric information, both for the data term evaluating observation conformity, and for expressing deformation constraints that enforce prior assumptions on motion. Experiments on several datasets demonstrate similar or improved precisions over state-of-the-art methods, as well as improved robustness, a critical issue when tracking sequentially over time frames. This work was accepted as **oral** at CVPR 2015 (less than 3% acceptance rate) [8].
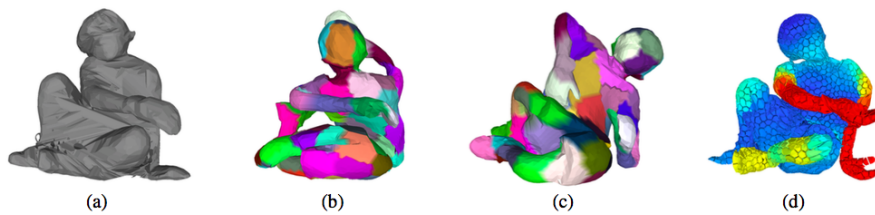


*Figure 4. Frames of the* GOALKEEPER *dataset acquired on the Kinovis platform. (a) Visual hull input. (b) Tracking result of Cagniart et al. 2010. (c) Allain et al. 2014. (d) This method [8]. Note the improved angular shapes and the improved robustness.*

## 7.3. Sparse Multi-View Consistency for Object Segmentation

Multiple view segmentation consists in segmenting objects simultaneously in several views. A key issue in that respect and compared to monocular settings is to ensure propagation of segmentation information between views while minimizing complexity and computational cost. In this work, we first investigate the idea that examining measurements at the projections of a sparse set of 3D points is sufficient to achieve this goal. The proposed algorithm softly assigns each of these 3D samples to the scene background if it projects on the background region in at least one view, or to the foreground if it projects on foreground region in all views. Second, we show how other modalities such as depth may be seamlessly integrated in the model and benefit the segmentation. The paper exposes a detailed set of experiments used to validate the algorithm, showing results comparable with the state of art, with reduced computational complexity. We also discuss the use of different modalities for specific situations, such as dealing with a low number of viewpoints or a scene with color ambiguities between foreground and background. This work was published as article in the PAMI journal [3].

## 7.4. Building Statistical Shape Spaces for 3D Human Modeling

Statistical models of 3D human shape and pose learned from scan databases have developed into valuable tools to solve a variety of vision and graphics problems. Unfortunately, most publicly available models are of limited expressiveness as they were learned on very small databases that hardly reflect the true variety in human body shapes. In this paper, we contribute by rebuilding a widely used statistical body representation from the largest commercially available scan database, and making the resulting model available to the community (visit http://humanshape.mpi-inf.mpg.de). As preprocessing several thousand scans for learning the model is a challenge in itself, we contribute by developing robust best practice solutions for scan alignment that quantitatively lead to the best learned models. We make implementations of these preprocessing steps also publicly available. We extensively evaluate the improved accuracy and generality of our new model, and show its improved performance for human body reconstruction from sparse input data. This work was published as Max Planck research report [17].

*Figure 5. Three views of the* PLANT *dataset as processed by our method for mutli-view silhouette extraction [3].*
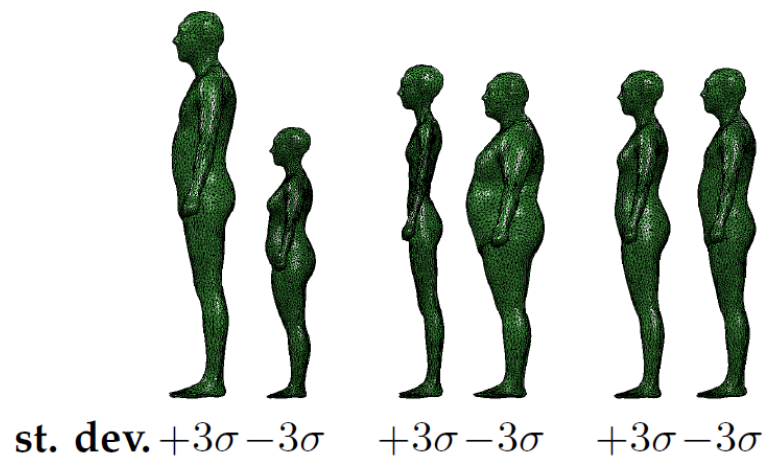


*Figure 6. Visualization of the first three principal components learned from a large database of posture-normalized 3D human body scans [17].*

## 7.5. A Groupwise Multilinear Correspondence Optimization for 3D Faces

Multilinear face models are widely used to model the space of human faces with expressions. For databases of 3D human faces of different identities performing multiple expressions, these statistical shape models decouple identity and expression variations. To compute a high-quality multilinear face model, the quality of the registration of the database of 3D face scans used for training is essential. Meanwhile, a multilinear face model can be used as an effective prior to register 3D face scans, which are typically noisy and incomplete. Inspired by the minimum description length approach, we propose the first method to jointly optimize a multilinear model and the registration of the 3D scans used for training. Given an initial registration, our approach fully automatically improves the registration by optimizing an objective function that measures the compactness of the multilinear model, resulting in a sparse model. We choose a continuous representation for each face shape that allows to use a quasi-Newton method in parameter space for optimization. We show that our approach is computationally significantly more efficient and leads to correspondences of higher quality than existing methods based on linear statistical models. This allows us to evaluate our approach on large standard 3D face databases and in the presence of noisy initializations. This work was published at the ICCV conference [9].

## 7.6. A statistical shape space model of the palate surface trained on 3D MRI scans of the vocal tract

We describe a minimally-supervised method for computing a statistical shape space model of the palate surface. The model is created from a corpus of volumetric magnetic resonance imaging (MRI) scans collected from 12 speakers. We extract a 3D mesh of the palate from each speaker, then train the model using principal component analysis (PCA). The palate model is then tested using 3D MRI from another corpus and evaluated using a high-resolution optical scan. We find that the error is low even when only a handful of measured coordinates are available. In both cases, our approach yields promising results. It can be applied to extract the palate shape from MRI data, and could be useful to other analysis modalities, such as electromagnetic articulography (EMA) and ultrasound tongue imaging (UTI). This work was published at the 18th International Congress of Phonetic Sciences [11].

## 7.7. Toward User-specific Tracking by Detection of Human Shapes in Multi-Cameras

Human shape tracking consists in fitting a template model to temporal sequences of visual observations. It usually comprises an association step, that finds correspondences between the model and the input data, and a deformation step, that fits the model to the observations given correspondences. Most current approaches find their common ground with the Iterative-Closest-Point (ICP) algorithm, which facilitates the association step with local distance considerations. It fails when large deformations occur, and errors in the association tend to propagate over time. In this paper, we propose a discriminative alternative for the association, that leverages random forests to infer correspondences in one shot. It allows for large deformations and prevents tracking errors from accumulating. The approach is successfully integrated to a surface tracking framework that recovers human shapes and poses jointly. When combined with ICP, this discriminative association proves to yield better accuracy in registration, more stability when tracking over time, and faster convergence. Evaluations on existing datasets demonstrate the benefits with respect to the state-of-the-art. This work was published at CVPR 2015 [12].

## 7.8. Video based Animation Synthesis with the Essential Graph

We propose a method to generate animations using video-based mesh sequences of elementary movements of a shape. New motions that satisfy high-level user-specified constraints are built by recombining and interpolating the frames in the observed mesh sequences. The interest of video based meshes is to provide real full shape information and to enable therefore realistic shape animations. A resulting issue lies, however, in the difficulty

to combine and interpolate human poses without a parametric pose model, as with skeleton based animations. To address this issue, our method brings two innovations that contribute at different levels: Locally between two motion sequences, we introduce a new approach to generate realistic transitions using dynamic time warping; More globally, over a set of motion sequences, we propose the essential graph as an efficient structure to encode the most realistic transitions between all pairs of input shape poses. Graph search in the essential graph allows then to generate realistic motions that are optimal with respect to various user-defined constraints. We present both quantitative and qualitative results on various 3D video datasets. They show that our approach compares favourably with previous strategies in this field that use the motion graph. This work was published at the 3DV 2015 conference [10].
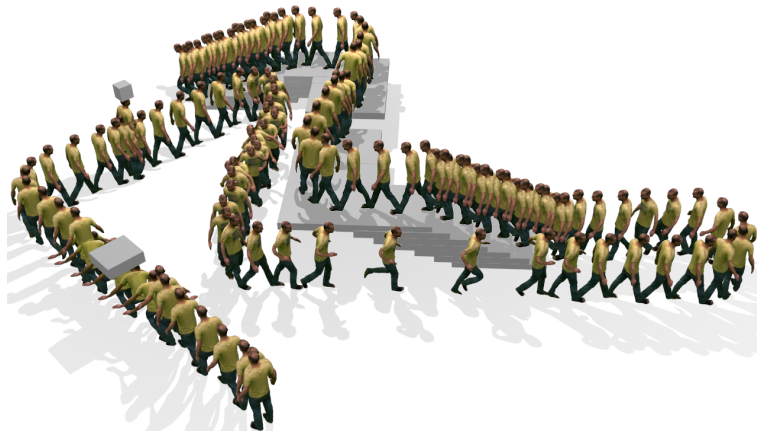


*Figure 7. Example of 4D animation generated using by combining recorded 4D sequences [10].*

## 7.9. Implicit B-Spline Surface Reconstruction

This paper presents a fast and flexible curve/surface reconstruction technique based on implicit b-spline. This representation does not require any parameterization and it is locally supported. This fact has been exploited in this paper to propose a reconstruction technique through solving a sparse system of equations. This method is further accelerated to reduce the dimension to the active control lattice. Moreover, the surface smoothness and user interaction are allowed for controlling the surface. Finally, a novel weighting technique has been introduced in order to blend small patches and smooth them in the overlapping regions. The whole framework is very fast and efficient and can handle large cloud of points with low computational cost. The experimental results show the flexibility and accuracy of the proposed algorithm to describe objects with complex topologies. Comparisons with other fitting methods highlight the superiority of the proposed approach in the presence of noise and missing data. This work was published as journal article in IEEE Transactions on Image Processing [6].

## 7.10. A Bayesian Approach to Multi-view 4D Modeling

This paper considers the problem of automatically recovering temporally consistent animated 3D models of arbitrary shapes in multi-camera setups. An approach is presented that takes as input a sequence of frame-wise reconstructed surfaces and iteratively deforms a reference surface such that it fits the input observations. This approach addresses several issues in this field that include: large frame-to-frame deformations, noise, missing data, outliers and shapes composed of multiple components with arbitrary geometries. The problem is cast as a geometric registration with two major features. First, surface deformations are modeled using

mesh decomposition into elements called patches. This strategy ensures robustness by enabling flexible regularization priors through inter-patch rigidity constraints. Second, registration is formulated as a Bayesian estimation that alternates between probabilistic datal-model association and deformation parameter estimation. This accounts for uncertainties in the acquisition process and allows for noise, outliers and missing geometries in the observed meshes. In the case of marker-less 3D human motion capture, this framework can be specialized further with additional articulated motion constraints. Extensive experiments on various 4D datasets show that complex scenes with multiple objects of arbitrary nature can be processed in a robust way. They also demonstrate that the framework can capture human motion and provides visually convincing as well as quantitatively reliable human poses. This work was published as journal article in International Journal on Computer Vision (IJCV) [4].

## 7.11. A Hierarchical Approach for Regular Centroidal Voronoi Tessellations

In this paper we consider Centroidal Voronoi Tessellations (CVTs) and study their regularity. CVTs are geometric structures that enable regular tessellations of geometric objects and are widely used in shape modeling and analysis. While several efficient iterative schemes, with defined local convergence properties, have been proposed to compute CVTs, little attention has been paid to the evaluation of the resulting cell decompositions. In this paper, we propose a regularity criterion that allows us to evaluate and compare CVTs independently of their sizes and of their cell numbers. This criterion allows us to compare CVTs on a common basis. It builds on earlier theoretical work showing that second moments of cells converge to a lower bound when optimising CVTs. In addition to proposing a regularity criterion, this paper also considers computational strategies to determine regular CVTs. We introduce a hierarchical framework that propagates regularity over decomposition levels and hence provides CVTs with provably better regularities than existing methods. We illustrate these principles with a wide range of experiments on synthetic and real models.

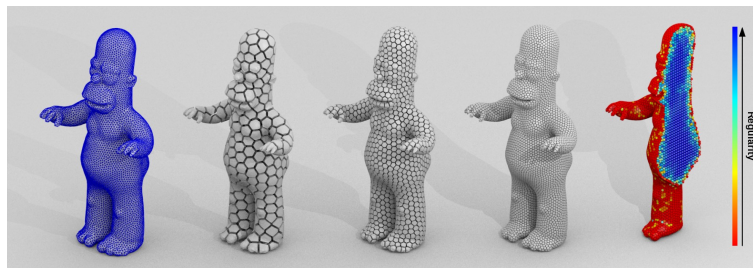This work was published as a journal article in Computer Graphics Forum [7].



*Figure 8. Hierarchical computation of a centroidal Voronoi tessellation from a 3D mesh [7]. Inside cells are very regular.*

## 7.12. Just Noticeable Distortion Profile for Flat-Shaded 3D Mesh Surfaces

It is common that a 3D mesh undergoes some lossy operations (e.g., compression, watermarking and transmission through noisy channels), which can introduce geometric distortions as a change in vertex position. In most cases the end users of 3D meshes are human beings; therefore, it is important to evaluate the visibility of introduced vertex displacement. In this paper we present a model for computing a Just Noticeable Distortion (JND) profile for flat-shaded 3D meshes. The proposed model is based on an experimental study of the properties of the human visual system while observing a flat-shaded 3D mesh surface, in particular the contrast sensitivity function and contrast masking. We first define appropriate local perceptual properties on 3D meshes. We then detail the results of a series of psychophysical experiments where we have measured the threshold

needed for a human observer to detect the change in vertex position. These results allow us to compute the JND profile for flat-shaded 3D meshes. The proposed JND model has been evaluated via a subjective experiment, and applied to guide 3D mesh simplification as well as to determine the optimal vertex coordinates quantization level for a 3D model.

This work was published as a journal article in IEEE Transactions on Visualization and Computer Graphics [5].
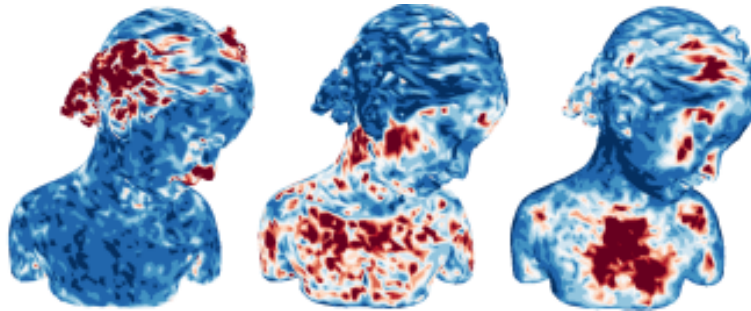


*Figure 9. Just noticeable distortion profile in a light independent mode (left, middle) or with a light fixed in front of the model (right), for vertex displacements in the normal direction (left, right) or in the tangent direction (middle) [5].*

<p style="text-align: center; color: red;">**PERCEPTION Project-Team**</p>

# 6. New Results

## 6.1. Supervised Audio-Source Localization

We addressed the problem of localizing audio sources using binaural measurements. After proposing an unsupervised method [20], we proposed a supervised formulation that simultaneously localizes multiple sources at different locations [22]. The approach is intrinsically efficient because, contrary to prior work, it relies neither on source separation, nor on monaural segregation. The method starts with a training stage that establishes a locally-linear Gaussian regression [21] between the directional coordinates of all the sources and the auditory features extracted from binaural measurements. While fixed-length wide-spectrum sounds (white noise) are used for training to reliably estimate the model parameters, we show that the testing (localization) can be extended to variable-length sparse-spectrum sounds (such as speech), thus enabling a wide range of realistic applications. Indeed, we demonstrate that the method can be used for audio-visual fusion, namely to map speech signals onto images and hence to spatially align the audio and visual modalities, thus enabling to discriminate between speaking and non-speaking faces. We release a novel corpus of real-room recordings that allow quantitative evaluation of the co-localization method in the presence of one or two sound sources. Experiments demonstrate increased accuracy and speed relative to several state-of-the-art methods. More recently the method has been extended to an arbitrary number of microphones [35], [34]. Moreover, we have started to develop a method that extracts the direct path on an acoustic wave in order to enable robust audio-source localization in reverberant environments [40].

Websites:
https://team.inria.fr/perception/research/acoustic-learning/
https://team.inria.fr/perception/research/binaural-ssl/
https://team.inria.fr/perception/research/local-rtf/

## 6.2. Multichannel Audio-Source Separation

We address the problem of separating audio sources from time-varying convolutive mixtures. We proposed an unsupervised probabilistic framework based on the local complex-Gaussian model combined with non-negative matrix factorization. The time-varying mixing filters are modeled by a continuous temporal stochastic process. This model extends the case of static filters which corresponds to static audio sources. While static filters can be learn in advance, e.g. [37], time-varying filters cannot and therefore the problem is more complex. We present a variational expectation-maximization (VEM) algorithm that employs a Kalman smoother to estimate the time-varying mixing matrix, and that jointly estimates the source parameters. The sound sources are then separated by Wiener filters constructed with the estimators provided by the VEM algorithm. Extensive experiments on simulated data show that the proposed method outperforms a block-wise version of a state-of-the-art baseline method. This work is part of the PhD topic of Dionyssos Kounades Bastian and is conducted in collaboration with Sharon Gannot (Bar Ilan University) and Xavier Alameda Pineda (University of Trento). It received the best student paper award at WASPAA'15 [31]. An extended version has been submitted to IEEE Transactions on Audio, Speech, and Language Processing [39].

## 6.3. Audio-Visual Speaker Tracking and Recognition

Any multi-party conversation system benefits from speaker diarization, that is, the assignment of speech signals among the participants. More generally, in HRI and CHI scenarios it is important to recognize the speaker over time. We propose to address speaker diarization and speaker recognition using both audio and visual data. We cast the diarization problem into a tracking formulation whereby the active speaker is detected and tracked over time. A probabilistic tracker exploits the spatial coincidence of visual and auditory observations and infers a single latent variable which represents the identity of the active speaker. Visual and auditory observations are fused using our recently developed weighted-data mixture model [38], while several options for the speaking turns dynamics are fulfilled by a multi-case transition model. The modules that translate raw audio and visual data into image observations are also described in detail. The performance of the proposed trackers [29], [30] are tested on challenging data-sets that are available from recent contributions which are used as baselines for comparison. Currently we are developing a variational framework for the on-line tracking of multiple persons [36].

Websites:
https://team.inria.fr/perception/research/speakerloc/
https://team.inria.fr/perception/research/speechturndet/
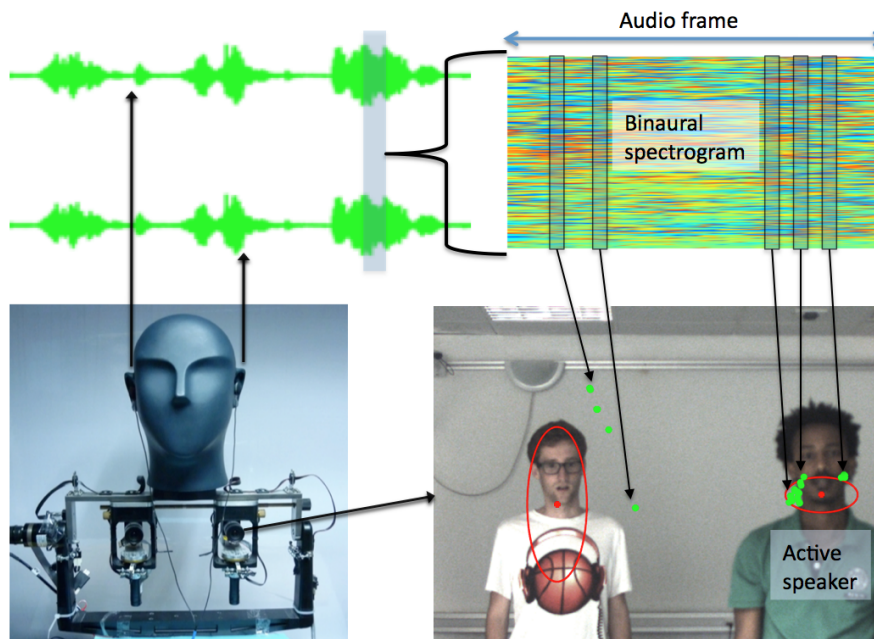https://team.inria.fr/perception/research/avdiarization/



*Figure 6. This figures illustrates the general principle of our audio-visual speaker tracking and diarization method. The auditory and visual data are recorded with two microphones and one camera. The audio signals are segmented into frames and each frame (vertical grey rectangle) is transformed into a binaural spectrogram [20]. This spectrogram is composed of a sequence of binaural vectors (vertical rectangles) and each binaural vector is mapped onto a sound-source direction which corresponds to a point in the image plane (green dots) [22]. The proposed audio-visual tracker associates people detected in the image sequence with these sound directions via audio-visual clustering [38] that is combined with an active-speaker transition model.*

## 6.4. Head Pose Estimation

Head pose estimation is an important task, because it provides information about cognitive interactions that are likely to occur. Estimating the head pose is intimately linked to face detection. We addressed the problem of head pose estimation with three degrees of freedom (pitch, yaw, roll) from a single image and in the presence of face detection errors. Pose estimation is formulated as a high-dimensional to low-dimensional mixture of linear regression problem [21]. We propose a method that maps HOG-based descriptors, extracted from face bounding boxes, to corresponding head poses. To account for errors in the observed bounding-box position, we learn regression parameters such that a HOG descriptor is mapped onto the union of a head pose and an offset, such that the latter optimally shifts the bounding box towards the actual position of the face in the image. The performance of the proposed method is assessed on publicly available datasets. The experiments that we carried out show that a relatively small number of locally-linear regression functions is sufficient to deal with the non-linear mapping problem at hand. Comparisons with state-of-the-art methods show that our method outperforms several other techniques. This work is part of the PhD of Vincent Drouard and it received the best student paper award (second place) at the IEEE ICIP'15 [28]. Currently we investigate a temporal extension of this model.

Website:
https://team.inria.fr/perception/research/head-pose/

## 6.5. High-Resolution Scene Reconstruction

We addressed the problem of range-stereo fusion for the construction of high-resolution depth maps. In particular, we combine low-resolution depth data with high-resolution stereo data, in a maximum a posteriori (MAP) formulation. Unlike existing schemes that build on MRF optimizers, we infer the disparity map from a series of local energy minimization problems that are solved hierarchically, by growing sparse initial disparities obtained from the depth data. The accuracy of the method is not compromised, owing to three properties of the data-term in the energy function. Firstly, it incorporates a new correlation function that is capable of providing refined correlations and disparities, via sub-pixel correction. Secondly, the correlation scores rely on an adaptive cost aggregation step, based on the depth data. Thirdly, the stereo and depth likelihoods are adaptively fused, based on the scene texture and camera geometry. These properties lead to a more selective growing process which, unlike previous seed-growing methods, avoids the tendency to propagate incorrect disparities. The proposed method gives rise to an intrinsically efficient algorithm, which runs at 3FPS on 2.0MP images on a standard desktop computer. The strong performance of the new method is established both by quantitative comparisons with state-of-the-art methods, and by qualitative comparisons using real depth-stereo data-sets [23]. This work is funded by the ANR project MIXCAM.

Website:
https://team.inria.fr/perception/research/dsfusion/

## 6.6. Hyper-Spectral Image Analysis

As an extension to our work on high-dimensional regression [21] we addressed the problem of analyzing hyper-spectral data. In particular we addressed the problem of recovering physical properties (parameters) form hyper-spectral low-resolution images, i.e. at large planetary scales. This involves resolving inverse problems which can be addressed within machine learning, with the advantage that, once a relationship between physical parameters and spectra has been established in a data-driven fashion, the learned relationship can be used to estimate physical parameters for new hyper-spectral observations. Within this framework, we propose a spatially-constrained and partially-latent regression method which maps high-dimensional inputs (hyper-spectral images) onto low-dimensional responses (physical parameters such as the local chemical composition of the soil). The proposed regression model comprises two key features. Firstly, it combines a Gaussian mixture of locally-linear mappings (GLLiM) with a partially-latent response model. While the former makes high-dimensional regression tractable, the latter enables to deal with physical parameters that cannot be observed or, more generally, with data contaminated by experimental artifacts that cannot be
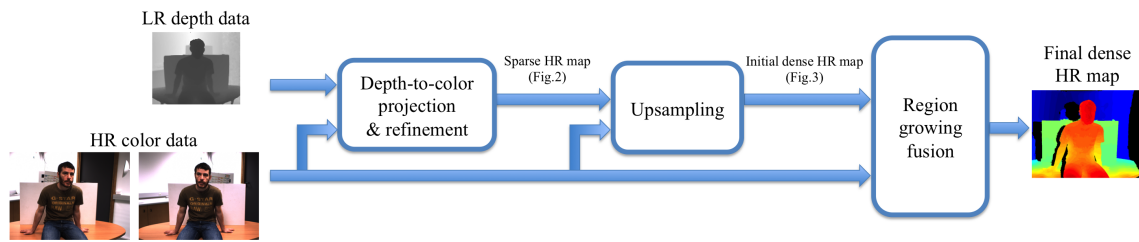
*Figure 7. The pipeline of the proposed depth-stereo fusion method [23]. The low-resolution (LR) depth data are projected onto the color data and refined to yield a high-resolution (HR) sparse disparity map. Starting from these disparity seeds, an upsampling process provides an initial HR dense disparity map. Both the HR seeds and the initial dense disparity map are then used by the region-growing depth-stereo fusion to produce the final HR depth map. A prominent feature of our method is that fusion takes place at several data processing stages.*

explained with noise models. Secondly, spatial constraints are introduced in the model through a Markov random field (MRF) prior which provides a spatial structure to the Gaussian-mixture hidden variables [19]. Experiments conducted on a database composed of remotely sensed observations collected from the Mars planet by the Mars Express orbiter demonstrate the effectiveness of the proposed model.

## 6.7. Gaussian Mixture Regression for Acoustic-Articulatory Inversion

The team expertise in latent-variable mixture models was applied to the problem of adaptation of an acoustic-articulatory model of a reference speaker to the voice of another speaker, using a limited amount of audio-only data [25]. In the context of pronunciation training, a virtual talking head displaying the internal speech articulators (e.g., the tongue) could be automatically animated by means of such a model using only the speaker's voice. In this study, the articulatory-acoustic relationship of the reference speaker is modeled by a gaussian mixture model (GMM). To address the speaker adaptation problem, we propose a new framework called cascaded Gaussian mixture regression (C-GMR), and derive two implementations. The first one, referred to as Split-C-GMR, is a straightforward chaining of two distinct GMRs: one mapping the acoustic features of the source speaker into the acoustic space of the reference speaker, and the other estimating the articulatory trajectories with the reference model. In the second implementation, referred to as Integrated-C-GMR, the two mapping steps are tied together in a single probabilistic model. For this latter model, we present the full derivation of the exact EM training algorithm, that explicitly exploits the missing data methodology of machine learning. Other adaptation schemes based on maximum-a posteriori (MAP), maximum likelihood linear regression (MLLR) and direct cross-speaker acoustic-to-articulatory GMR are also investigated. Experiments conducted on two speakers for different amount of adaptation data show the interest of the proposed C-GMR techniques. This work was done in collaboration with Thomas Hueber and Gérard Bailly from Gipsa Lab and with Xavier Alameda-Pineda from University of Trento and former team member.

# PRIMA Project-Team

# 6. New Results

## 6.1. Attention-Based Navigation

**Participants:** Thierry Fraichard, Remi Paulin, Patrick Reignier.

The domain of service-robots is growing fast and has become the focus of many researchers and industrials alike. Application areas are extremely broad, from logistics to handicap assistance. A large proportion of such robots are expected to share humans' living space and thus must be endowed with navigation capabilities that exceed the standard requirements pertaining to autonomous navigation such as motion safety. In a human populated environment, optimality does not boil down to minimizing resources such as time or distance traveled anymore, the robot motion must abide by social rules and move in a manner which is **appropriate**.

Most of the approaches proposed so far rely upon the definition of so-called **social spaces**, *i.e.* regions in the environment that, for different reasons, the persons consider as psychologically theirs. Such social spaces are primarily characterized using either the position of the person, *e.g.* "Personal space" [36], or the activity he is currently engaged in, *e.g.* "Interaction Space" [41] and "Activity Space" [45]. The most common approach is then to define costmaps on such social spaces: the higher the cost, the less desirable it is for the robot to be at the corresponding position. The costmaps are ultimately used for motion planning and navigation purposes.

While improving upon the standard "non social" navigation methods, this type of approach intrinsically ignores the correlations between interactions as well as the influence of the robot on those interactions. It thus fails to capture several important features of social navigation, such as the distraction and surprise caused to the surrounding individuals. To overcome those limits, we suggest using the psychological concept of attention, which plays a central role when humans navigate around each other. This concept brings a new degree of control over the motion of the robot, namely the invasive and distracting character of the robot motion, which have so far proven hard to tackle with the conventional tools such as social spaces. Beside leading appropriate motion, attention-based navigation enables interaction through motion by predicting the quantity of attention the human will give to the robot.

Building upon a computational model of attention that was earlier proposed in [47], we have developed the novel concept of **attention field**. The attention field is straightforward to define: it is a measure of the amount of attention that a given person would allocate to the robot, should the robot be in a given position/state. It is a mapping from the state space of the robot to $\mathbb{R}$. We use this attention field in order to carefully control the degree of distraction caused by the robot to the individuals in its surroundings. By monitoring the variations of attentional resources that it causes, we also control the amount of surprise caused by the robot which must be kept to a minimum since it is a cause of discomfort. Furthermore this approach enables us to tackle more complex situations where more than one person is involved such as the task of delivering a private message to an individual, or else joining a group (an example of interaction through motion). Rather than navigating on a single global costmap, this new approach provides for each path several measures of the distraction and surprise caused by the robot on a given individual. Those quantities are then multi-optimised in order to find a path that satisfies all the given requirements for fulfilling the robot's task as well as minimizing the discomfort for individuals who are not directly involved in an interaction with the robot.

In 2015, we have developed a variant of the well-known differential evolution algorithm which deals with optimizing continuous trajectories under multiple constraints. The performance of our approach is now being compared with trajectories obtained by relying only on social spaces. Besides the traditional qualitative approaches to evaluate the discomfort caused by the robot motion, we work on defining more quantitative measures that would enable us to further validate our approach.

## 6.2. SPOK: End User Programming for Smart Homes

**Participant:** Alexandre Demeure.

As part of the CATRENE project AppsGate, we have developed SPOK, an End User Development Environment, that enables inhabitants to control and program their smart Homes via a web interface. The current version of SPOK includes an editor for editing programs using a pseudo-natural language and an interpreter. A multi-syntax editor as well as additional services such as a debugger and a simulator are expected for the second version.

A multi-syntax editor will allow users to build syntactically correct programs using the syntax that is most appropriate to them or by using a combination of them. These syntaxes include pseudo-natural language (i.e. a constrained natural language) and graphical iconic syntax (as exemplified by Scratch [Maloney et al. 2010]). The interaction techniques used to enter programs may be menu-based, free typing, as well as by demonstration in the physical home or by the way of the simulator. The simulator is the dual digital representation of the real home. It is intended to serve also as a debugger for testing and correcting end-user programs.

Whatever syntax used by end-users, programs are translated into syntactic abstract trees whose leaves reference services provided by the Core HMI and/or by the Extended HMI Middleware. The interpreter, executes end-user programs, using the corresponding syntactic abstract trees as input.

In order to support a dynamically extensible grammar as well as to provide end-users with feedforward at the user interface of the editor, the grammar used by the editor is split into 2 parts: the root grammar and the device specific grammars. The root grammar specifies the generic structures of an end-user program: loops, conditions, etc. The device specific grammars are separated from the root grammar to be able to dynamically build the final grammar to be compliant with what is currently installed and detected by the AppsGate server. Each device type brings with it its own events, status and actions. These grammatical elements are injected into the root grammar when generating the parser and for compiling end-user programs.

The language used by end-users to express their programs is a pseudo-natural language using the rule-based programming paradigm. The left hand side of a rule is composed of events and conditions, and the right hand side specifies the actions to be taken when the left hand-side is true or becomes true. A program may include several rules that can be executed either in parallel or sequentially. Once entered, programs are translated into syntactic abstract trees. The interpreter, executes end-user programs, using the corresponding syntactic abstract trees as input. SPOK is implemented as a mix of OSGi and ApAM components where ApAM is in turn a middleware that runs on top of OSGi.

## 6.3. Qualitative approaches for building energy management

**Participant:** Patrick Reignier.

Reducing housing energy costs is a major challenge of the 21st century. In the near future, the main issue for building construction is the thermal insulation, but in the longer term, the issues are those of renewable energy (solar, wind, etc.) and smart buildings. Home automation system basically consists of household appliances linked via a communication network allowing interactions for control purposes. Thanks to this network, a load management mechanism can be carried out: it is called distributed control. An optimal home energy management system is still a goal to aim for, because lots of aspects are still not completely fulfilled. Most of the energy systems respect only the energy needs, but they don't tackle the user needs or satisfaction. Energy systems also have a lack when it comes to the dynamicity of the environments (the system ability to adapt).The problem is similar for the existing HMI (Human User Interface) of those Home Automation Systems where only experts can understand the data coming from the sensors and most important, the energy plan coming from management system (How? and Why?). The goal of this study is to propose a house energy model that can be both used to predict at some level energy evolution and that can be understood by the end user. The house energy model is based on Fuzzy Cognitive Maps representing cause-effects relations. It is first designed by an expert and then automatically tuned to a particular house using machine learning approaches. Preliminary experiments have been done this year using the Predis datasets.

## 6.4. Situation Aware Services on Mobile Devices

**Participants:** James Crowley, Thibaud Flury.

Modern mobile devices, such as smart phones and tablets, combine a rich set of sensors, internet connectivity, with embedded computational power and memory. The PRIMA group has recently demonstrated that it is possible to construct embedded software that uses the full suite of mobile sensors to recognise activities and learn the daily routines of users.

A first proof of concept has recently been constructed using recognition of places and activities. The system was trained by having student volunteers carry a cell phone running a data acquisition program that recorded signals from accelerometer, gyroscope, ambient sound, ambient light, Cell tower, wifi, bluetooth, and GPS based geolocalisation. The data were labeled by the students with ground truth data about transportation modes, places, and activities. This data was then used to learn recognition routines. Recognition of places, activities, and transportation was used to construct probabilistic models of daily routines using PRIMA's situation modelling techniques, previous demonstrated in constructing situation aware services. The system was demonstrated by constructing a Twitter Bot (a robot that publishes on twitter) that published information about volunteers during their daily activity.

A professional quality software system named CAM - Context Aware Manager - is currently under construction, and will be licensed to the PRIMA startup SItu8ed, for use in context aware mobile services.

## 6.5. Perceiving mass in mixed reality

**Participants:** Sabine Coquillart, Paul Issartel.

In mixed reality, real objects can be used to interact with virtual objects. However, unlike in the real world, real objects do not encounter any opposite reaction force when pushing against virtual objects. The lack of reaction force during manipulation prevents users from perceiving the mass of virtual objects. Although this could be addressed by equipping real objects with force-feedback devices, such a solution remains complex and impractical. In this work, we present a technique to produce an illusion of mass without any active force-feedback mechanism. This is achieved by simulating the effects of this reaction force in a purely visual way. A first study demonstrates that our technique indeed allows users to differentiate light virtual objects from heavy virtual objects. In addition, it shows that the illusion is immediately effective, with no prior training. In a second study, we measure the lowest mass difference (JND) that can be perceived with this technique. The effectiveness and ease of implementation of our solution provides an opportunity to enhance mixed reality interaction at no additional cost.

## 6.6. Pseudo-haptic feedback

**Participants:** Sabine Coquillart, Jingtao Chen.

"Pseudo-haptic feedback" is a technique aiming to simulate haptic sensations without active haptic feedback devices. Peudo-haptic techniques have been used to simulate various haptic feedbacks such as stiffness, torques, and mass. In the framework of Jingtao Chen PhD thesis, a novel pseudo-haptic experiment has been set up. The aim of this experiment is to study the EMG signals during a pseudo-haptic task. A stiffness discrimination task similar to the one published in Lecuyer's PhD thesis has been chosen. The experimental set-up has been developed, as well as the software controlling the experiment. Pre-tests are under way. They will be followed by the tests with subjects.

<span style="color:red">**TYREX Project-Team**</span>

# 6. New Results

## 6.1. Expressive Logical Combinators

A popular technique for the analysis of web query languages relies on the translation of queries into logical formulas. These formulas are then solved for satisfiability using an off-the-shelf satisfiability solver. A critical aspect in this approach is the size of the obtained logical formula, since it constitutes a factor that affects the combined complexity of the global approach. In this work [21], we present logical combinators whose benefit is to provide an exponential gain in succinctness in terms of the size of the logical representation. This opens the way for solving a wide range of problems such as satisfiability and containment for expressive query languages in exponential-time, even though their direct formulation into the underlying logic results in an exponential blowup of the formula size, yielding an incorrectly presumed two-exponential time complexity. We illustrate this from a practical point of view on a few examples such as numerical occurrence constraints and tree frontier properties which are concrete problems found with semi-structured data [21].

## 6.2. Behavioural Types

Behavioural type systems ensure more than the usual safety guarantees of static analysis. They are based on the idea of "types-as-processes", providing dedicated type algebras for particular properties, ranging from protocol compatibility to race-freedom, lock-freedom, or even responsiveness.

Two successful, although rather different, approaches, are session types and process types. The former allows to specify and verify (distributed) communication protocols using specific type (proof) systems; the latter allows to infer from a system specification a process abstraction on which it is simpler to verify properties, using a generic type (proof) system. What is the relationship between these approaches? Can the generic one subsume the specific one? At what price? And can the former be used as a compiler for the latter?

In [15], we showed how communication protocols can be integrated into an object-oriented type system supporting *non-uniform objects*, i. e. objects where the sequences of method calls are restricted, such as a `File` where `read()` cannot be called after `close()`. In such a system, communication protocols can be enforced by giving appropriate non-uniform types to the socket objects. We defined a sound and complete type checking algorithm for a small distributed class-based object-oriented language with structural subtyping. Static typing guarantees that both sequences of messages on channels, and sequences of method calls on objects, conform to type-theoretic specifications, thus ensuring type-safety.

## 6.3. SPARQL Queries

Static analysis is a core task in query optimization and knowledge base verification. In [14], [24], we study static analysis techniques for SPARQL, the standard language for querying Semantic Web data. We are interested in developing techniques through reductions to the validity problem in logic

In [22], we investigate techniques for detecting SPARQL query update independence. A query is independent of an update when the execution of the update does not affect the result of the query. Determining independence is especially useful in the context of huge RDF repositories, where it permits to avoid expensive yet useless re-evaluation of queries. While this problem has been intensively studied for fragments of relational calculus, very few works exist for the standard query language for the semantic web. We report on our investigations on how a notion of independence can be defined in the SPARQL context.

## 6.4. Semantic Subtyping

In a programming language, subtyping represents a notion of safe substitutability (it is always safe to replace a value of some type with a value of a subtype). There are several ways such a relation can be formally defined. Semantic subtyping consists of giving a set-theoretic denotation to types and using set inclusion to define subtyping. Works by Benzaken, Castagna, Frisch and Xu have described how to define such relations for complex type algebras comprising recursive, product, function, intersection, union, and complement types together with type variables. In [17], we showed how to formalise such a relation in logic and decide it in EXPTIME, answering an open question, and discussed experiments made with the full implementation of the system in our solver (5.3 ).

## 6.5. Spatio-temporal validation of multimedia documents

A multimedia document authoring system should provide analysis and validation tools that help authors find and correct mistakes before document deployment. Although very useful, multimedia validation tools are not often provided. Spatial validation of multimedia documents may be performed over the initial position of media items before presentation starts. However, such an approach does not lead to good results when media item placement changes over time. Some document authoring languages allow the definition of spatio-temporal relationships among media items and they can be moved or resized during runtime. Current validation approaches do not verify dynamic spatio-temporal relationships. In [19], we present a novel approach for spatio-temporal validation of multimedia documents. We model the document state, extending the Simple Hypermedia Model (SHM), comprising media item positioning during the whole document presentation. Mappings between document states represent time lapse or user interaction. We also define a set of atomic formulas upon which the author's expectations related to the spatio-temporal layout can be described and analyzed.

## 6.6. XQuery and Static Typing

XQuery is a functional language dedicated to XML data querying and manipulation. As opposed to other W3C-standardized languages for XML (e.g. XSLT), it has been intended to feature strong static typing. Currently, however, some expressions of the language cannot be statically typed with any precision.

In [20], we argue that this is due to a discrepancy between the semantics of the language and its type algebra. We discuss how to handle this discrepancy by improving the type system. We describe a logic-based language of extended types able to represent inner tree nodes and show how it can dramatically increase the precision of typing for navigation expressions. We describe how inclusion between these extended types and the classical regular tree types can be decided, allowing a hybrid system combining both type languages. The result is a net increase in precision of typing.

In a previous work, we aimed at bridging the gap between path-based XML processing languages like XQuery and pattern-based such languages like CDuce. We extend the language CDuce into a succinct core $\lambda$-calculus that captures XQuery 3.0. The extensions we consider essentially allow CDuce to implement XPath-like navigational expressions by pattern matching and precisely type them. The elaboration of XQuery 3.0 into the extended CDuce provides a formal semantics and a sound static type system for XQuery 3.0 programs.

## 6.7. Efficiently Deciding $\mu$-calculus with Converse over Finite Trees

In [16], we present a sound and complete satisfiability-testing algorithm and its effective implementation for an alternation-free modal $\mu$-calculus with converse, where formulas are cycle-free and are interpreted over finite ordered trees. The time complexity of the satisfiability-testing algorithm is $2^{O}(n)$ in terms of formula size n. The algorithm is implemented using symbolic techniques (BDD). We present crucial implementation techniques and heuristics that we used to make the algorithm as fast as possible in practice. Our implementation is detailed in 5.3 .

## 6.8. Reasoning with Style

The Cascading Style Sheets (CSS) language constitutes a key component of web applications. It offers a series of sophisticated features to stylize web pages. Its apparent simplicity and power are however counterbalanced by the difficulty of debugging and maintaining style sheets, tasks for which developers still lack appropriate tools. In particular, significant portions of CSS code become either useless or redundant, and tend to accumulate over time. The situation becomes even worse as more complex features are added to the CSS language (e.g. CSS3 powerful selectors). A direct consequence is a waste of CPU that is required to display web pages, as well as the significant amount of useless traffic at web scale. Style sheets are designed to operate on a set of documents (possibly generated). However, existing techniques consist in syntax validators, optimizers and runtime debuggers that operate in one particular document instance. As such, they do not provide guarantees concerning all web pages in CSS refactoring, such as preservation of the formatting. This is partly because they are essentially syntactic and do not take advantage of CSS semantics to detect redundancies. In [18], we propose a set of automated refactoring techniques aimed at removing redundant and inaccessible declarations and rules, without affecting the layout of any document to which the style sheet is applied. We implemented a prototype that has been extensively tested with popular web sites (such as Google Sites, CNN, Apple, etc.). We show that significant size reduction can be obtained while preserving the code readability and improving maintainability.

## 6.9. A Comparative Analysis of Attitude Estimation

We investigate the precision of attitude estimation techniques in the context of pedestrian dead-reckoning with commodity smartphones. We propose a comparative analysis of state-of-the-art algorithms for attitude estimation in this setting. We provide an experimental setup with a precise ground truth obtained with a motion capture system. We precisely quantify the error in attitude estimation obtained with each technique. We discuss the obtained results and analyse advantages and limitations of current technology for further PDR research.