Activity Report 2013

Section New Results

<p style="text-align:center">**ARIC Project-Team**</p>

6. New Results

6.1. Cryptography and lattices

6.1.1. *Group signatures*

Group signatures are cryptographic primitives where users can anonymously sign messages in the name of a population they belong to. Gordon et al. (Asiacrypt 2010) suggested the first realization of group signatures based on lattice assumptions in the random oracle model. A significant drawback of their scheme is its linear signature size in the cardinality N of the group. A recent extension proposed by Camenisch et al. (SCN 2012) suffers from the same overhead.

F. Laguillaumie, A. Langlois, B. Libert (Technicolor), and D. Stehlé described in [24] the first lattice-based group signature schemes where the signature and public key sizes are essentially logarithmic in N (for any fixed security level). Their basic construction only satisfies a relaxed definition of anonymity (just like the Gordon et al. system) but readily extends into a fully anonymous group signature (i.e., that resists adversaries equipped with a signature opening oracle). They proved the security of their schemes in the random oracle model under the SIS and LWE assumptions.

6.1.2. *Classical hardness of learning with errors*

Z. Brakerski (Stanford U.), A. Langlois, C. Peikert (Georgia Institute of Technology), O. Regev (Courant Institute, New York U.), and D. Stehlé showed in [16] that the Learning with Errors (LWE) problem is classically at least as hard as standard worst-case lattice problems, even with polynomial modulus. Previously this was only known under quantum reductions. Their techniques capture the tradeoff between the dimension and the modulus of LWE instances, leading to a much better understanding of the landscape of the problem. The proof is inspired by techniques from several recent cryptographic constructions, most notably fully homomorphic encryption schemes.

6.1.3. *Improved Zero-knowledge Proofs of Knowledge for the ISIS Problem, and Applications*

In all existing efficient proofs of knowledge of a solution to the infinity norm Inhomogeneous Small Integer Solution ISIS_∞ problem, the knowledge extractor outputs a solution vector that is only guaranteed to be $\widetilde{O}(n)$ times longer than the witness possessed by the prover. As a consequence, in many cryptographic schemes that use these proof systems as building blocks, there exists a gap between the hardness of solving the underlying ISIS_∞ problem and the hardness underlying the security reductions. Together with S. Ling, K. Nguyen, and H. Wang (Nanyang Technological University, Singapore), D. Stehlé generalized in [26] Stern's protocol to obtain two statistical zero-knowledge proofs of knowledge for the ISIS_∞ problem that remove this gap. Their result yields the potential of relying on weaker security assumptions for various lattice-based cryptographic constructions. As applications of their proof system, they introduced a concurrently secure identity-based identification scheme based on the worst-case hardness of the $\text{SIVP}_{\widetilde{O}(n^{1.5})}$ problem (in the L2 norm) in general lattices in the random oracle model, and an efficient statistical zero-knowledge proof of plaintext knowledge with small constant gap factor for Regev's encryption scheme.

6.1.4. *Decoding by Embedding: Correct Decoding Radius and DMT Optimality*

In lattice-coded multiple-input multiple-output (MIMO) systems, optimal decoding amounts to solving the closest vector problem (CVP). Embedding is a powerful technique for the approximate CVP, yet its remarkable performance is not well understood. In [8], C. Ling (Imperial College, London), L. Luzzi (ENSEA, U. Cergy Pontoise), and D. Stehlé analyzed the embedding technique from a bounded distance decoding (BDD) viewpoint. They proved that the Lenstra, Lenstra and Lovász (LLL) algorithm can achieve $1/(2\gamma)$-BDD for $\gamma \approx O(2^{n/4})$, yielding a polynomial-complexity decoding algorithm performing exponentially better than Babai's which achieves $\gamma = O(2^{n/2})$. This substantially improves the existing result $\gamma = O(2^n)$ for embedding decoding. They also proved that BDD of the regularized lattice is optimal in terms of the diversity-multiplexing gain tradeoff (DMT).

6.1.5. A New View on HJLS and PSLQ: Sums and Projections of Lattices

The HJLS and PSLQ algorithms are the de facto standards for discovering non-trivial integer relations between a given tuple of real numbers. In [19], J. Chen, D. Stehlé, and G. Villard provided a new interpretation of these algorithms, in a more general and powerful algebraic setup: they view them as special cases of algorithms that compute the intersection between a lattice and a vector subspace. Further, they extracted from them the first algorithm for manipulating finitely generated additive subgroups of a Euclidean space, including projections of lattices and finite sums of lattices. They adapted the analyses of HJLS and PSLQ to derive correctness and convergence guarantees. They also investigated another approach based on embedding the input in a higher dimensional lattice and calling the LLL lattice reduction algorithm.

6.2. Certified computing and computer algebra

6.2.1. Polynomial system solving

Polynomial system solving is a core topic of computer algebra. While the worst-case complexity of this problem is known to be hopelessly large, the practical complexity for large families of systems is much more reasonable. Progress has been made in assessing precise complexity estimates in this area.

First, M. Bardet (U. Rouen), J.-C. Faugère (PolSys team), and B. Salvy studied the complexity of Gröbner bases computations, in particular in the generic situation where the variables are in simultaneous Noether position with respect to the system. They gave a bound on the number of polynomials of each degree in a Gröbner basis computed by Faugère's F5 algorithm in this generic case for the grevlex ordering (which is also a bound on the number of polynomials for a reduced Gröbner basis), and used it to bound the exponent of the complexity of the F5 algorithm [35].

Next, a fundamental problem in computer science is to find all the common zeroes of m quadratic polynomials in n unknowns over F_2. The cryptanalysis of several modern ciphers reduces to this problem. Up to now, the best complexity bound was reached by an exhaustive search in $4\log_2 n\, 2^n$ operations. In [1], M. Bardet (U. Rouen), J.-C. Faugère (PolSys team), B. Salvy, and P.-J. Spaenlehauer (CARAMEL team) gave an algorithm that reduces the problem to a combination of exhaustive search and sparse linear algebra. This algorithm has several variants depending on the method used for the linear algebra step. Under precise algebraic assumptions, they showed that the deterministic variant of their algorithm has complexity bounded by $O(2^{0.841n})$ when $m = n$, while a probabilistic variant of the Las Vegas type has expected complexity $O(2^{0.792n})$. Experiments on random systems showed that the algebraic assumptions are satisfied with probability very close to 1. They have also given a rough estimate for the actual threshold between their method and exhaustive search, which is as low as 200, and thus very relevant for cryptographic applications.

6.2.2. Linear differential equations

Creative telescoping algorithms compute linear differential equations satisfied by multiple integrals with parameters. Together with A. Bostan and P. Lairez (SpecFun team), B. Salvy described a precise and elementary algorithmic version of the Griffiths–Dwork method for the creative telescoping of rational functions. This leads to bounds on the order and degree of the coefficients of the differential equation, and to the first complexity result which is simply exponential in the number of variables. One of the important features of the algorithm is that it does not need to compute certificates. The approach is vindicated by a prototype implementation [15].

In [2], B. Salvy proved with A. Bostan (SpecFun team) and K. Raschel (U. Tours) that the sequence $(e_n^{\mathfrak{S}})_{n\geq 0}$ of excursions in the quarter plane corresponding to a nonsingular step set $\mathfrak{S} \subseteq \{0, \pm 1\}^2$ with infinite group does not satisfy any nontrivial linear recurrence with polynomial coefficients. Accordingly, in those cases, the trivariate generating function of the numbers of walks with given length and prescribed ending point is not D-finite. Moreover, they displayed the asymptotics of $e_n^{\mathfrak{S}}$. This completes the classification of these walks.

With F. Johansson and M. Kauers (RISC, Linz, Austria), M. Mezzarobba presented in [23] a new algorithm for computing hyperexponential solutions of ordinary linear differential equations with polynomial coefficients. The algorithm relies on interpreting formal series solutions at the singular points as analytic functions and evaluating them numerically at some common ordinary point. The numerical data is used to determine a small number of combinations of the formal series that may give rise to hyperexponential solutions.

6.2.3. *Exact linear algebra*

Transforming a matrix over a field to echelon form, or decomposing the matrix as a product of simpler matrices that reveal the rank profile, is a fundamental building block of computational exact linear algebra. For such tasks the best previously available algorithms were either rank sensitive (i.e., of complexity expressed in terms of the exponent of matrix multiplication and the rank of the input matrix) or in place (i.e., using essentially no more memory that what is needed for matrix multiplication). In [6] C.-P. Jeannerod, C. Pernet, and A. Storjohann (U. Waterloo, Canada) have proposed algorithms that are both rank sensitive and in place. These algorithms required to introduce a matrix factorization of the form $A = CUP$ with C a column echelon form giving the row rank profile of the input matrix A, U a unit upper triangular matrix, and P a permutation matrix.

6.2.4. *Certified multiple-precision evaluation of the Airy Ai function*

The series expansion at the origin of the Airy function $Ai(x)$ is alternating and hence problematic to evaluate for $x > 0$ due to cancellation. S. Chevillard (APICS team) and M. Mezzarobba showed in [20] how an arbitrary and certified accuracy can be obtained in that case. Based on a method recently proposed by Gawronski, Müller, and Reinhard, they exhibited two functions F and G, both with nonnegative Taylor expansions at the origin, such that $Ai(x) = G(x)/F(x)$. The sums are now well-conditioned, but the Taylor coefficients of G turn out to obey an ill-conditioned three-term recurrence. They then used the classical Miller algorithm to overcome this issue. Finally, they bounded all errors and proposed an implementation which, by allowing an arbitrary and certified accuracy, can be used for example to provide correct rounding in arbitrary precision.

6.2.5. *Standardization of interval arithmetic*

The IEEE 1788 working group is devoted to the standardization of interval arithmetic. V. Lefèvre and N. Revol are very active in this group. This year is the last year granted by IEEE for the preparation of a draft text of the standard. 2014 will be devoted to a ballot on the whole text, first by the standardization working group and then by a group of experts appointed by IEEE. In 2013, the definition of interval literals, of constructors, and of input and output has been adopted. The work now concentrates on portions of the final text [42].

6.2.6. *Parallel product of interval matrices*

The problem considered here is the multiplication of two matrices with interval coefficients. Parallel implementations by N. Revol and Ph. Théveny [10] compute results that satisfy the inclusion property, which is the fundamental property of interval arithmetic, and offer good performances: the product of two interval matrices is not slower than 15 times the product of two floating-point matrices.

6.2.7. *Numerical reproducibility*

What is called *numerical reproducibility* is the problem of getting the same result when the scientific computation is run several times, either on the same machine or on different machines. In [43], the focus is on interval computations using floating-point arithmetic: N. Revol identifies implementation issues that may invalidate the inclusion property, and presents several ways to preserve this inclusion property. This work has also been presented at several conferences [30], [29], [31].

6.3. Floating-point arithmetic

6.3.1. *Improved error bounds for complex floating-point arithmetic with a fused-multiply add*

Assuming that a fused multiply-add (FMA) instruction is available, C.-P. Jeannerod, N. Louvet, and J.-M. Muller [22] obtained sharp error bounds for various alternatives to Kahan's FMA-based algorithm for 2 x

2 determinants (which they had analyzed in [5]). They showed how to combine such variants with Kahan's original scheme in order to derive componentwise-accurate algorithms for complex floating-point division. Finally, they established sharp or reasonably sharp error bounds for each of these division algorithms.

C.-P. Jeannerod, P. Kornerup (U. of Southern Denmark), N. Louvet, and J.-M. Muller [36] studied the impact of the FMA on the normwise relative accuracy of complex floating-point multiplication. They showed that the classical normwise relative error bound $\sqrt{5}\,u$ (with u the unit roundoff) can be decreased further to $2u$, and that this new constant is best possible for several FMA-based multiplication algorithms.

J.-M. Muller analyzed in [41] another 2 x 2 determinant algorithm, due to Cornea, Harrison, and Tang, and showed that for radix 2 it admits a sharp relative error bound of the form $2u + O(u^2)$.

6.3.2. *Improved error bounds for numerical linear algebra*

C.-P. Jeannerod and S. M. Rump (Hamburg University of Technology) [7] showed that when evaluating sums of n real numbers in standard floating-point arithmetic, the usual fraction $\gamma_n = nu/(1 - nu)$, which has the form $nu + O(u^2)$ and requires $nu < 1$, can be replaced by nu without any restriction on n. Applications include simpler and more general error bounds for inner products, matrix-vector multiplication, and classical matrix multiplication.

In [45] they extended these results to LU and Cholesky factorizations as well as to triangular linear system solving by showing that the constants γ_n that appear classically in the backward error bounds for such problems can all be replaced by $O(u^2)$-free and unconditional constants nu. To get these new bounds the main ingredient is a general framework for bounding expressions of the form $|\rho - s|$, where s is the exact sum of a floating-point number and $n - 1$ real numbers, and where ρ is a real number approximating the computed sum $\widehat{s}$.

6.3.3. *On Ziv's rounding test*

F. de Dinechin, J.-M. Muller and S. Torres studied with C. Lauter (Univ. Paris 6) the rounding test introduced by Ziv in its libultim software [4]. This test determines if an approximation to the value $f(x)$ of an elementary function at a given point x suffices to return the floating-point number nearest to $f(x)$. They showed that the same test may be used for efficient implementation of floating-point operations with input and output operands of different formats. That test depends on a "magic constant" e and they also showed how to choose that constant to make the test reliable and efficient. Various cases are considered, depending on the availability of an FMA instruction, and on the range of $f(x)$.

6.3.4. *Various issues related to double roundings*

Double rounding is a phenomenon that may occur when different floating-point precisions are available on the same system. Although double rounding is, in general, innocuous, it may change the behavior of some useful floating-point algorithms. G. Melquiond (Toccata team), E. Martin-Dorel (then in the Marelle team), and J.-M. Muller analyzed in [9] the potential influence of double rounding on the Fast2Sum and 2Sum algorithms, on some summation algorithms, and Veltkamp's splitting. When performing divisions using Newton-Raphson (or similar) iterations on a processor with a floating-point fused multiply-add instruction, one must sometimes scale the iterations, to avoid over/underflow and/or loss of accuracy. This may lead to double-roundings, resulting in output values that may not be correctly rounded when the quotient falls in the subnormal range. J.-M. Muller showed in [13] how to avoid this problem.

6.3.5. *Comparison between binary and decimal floating-point numbers*

The IEEE 754-2008 standard for floating-point arithmetic arithmetic specifies binary as well as decimal formats. N. Brisebarre, C. Lauter (Univ. Paris 6), M. Mezzarobba, and J.-M. Muller introduced in [17] an algorithm that allows one to quickly compare a binary64 floating-point number and a decimal64 floating-point number, assuming the "binary encoding" of the decimal formats specified by the IEEE-754 standard is used. It is a two-step algorithm: a first pass, based on the exponents only, makes it possible to quickly eliminate most cases; then, when the first pass does not suffice, a more accurate second pass is required. They provide an implementation of several variants of their algorithm, and compare them.

6.3.6. *Conversions between binary and decimal floating-point numbers*

Conversion between binary and decimal floating-point representations is ubiquitous. Floating-point radix conversion means converting both the exponent and the mantissa. O. Kupriianova and C. Lauter (Univ. Paris 6) and J.-M. Muller developed in [38] an atomic operation for floating-point radix conversion with simple straight-line algorithm, suitable for hardware design. Exponent conversion is performed with a small multiplication and a lookup table. It yields the correct result without error. Mantissa conversion uses a few multiplications and a small lookup table that is shared amongst all types of conversions. The accuracy changes by adjusting the computing precision.

6.3.7. *Table-maker's dilemma*

Computing hardest-to-round cases of elementary functions is a key issue when one wants to develop an efficient and reliable implementation of such a function. The algorithms developed until now required a large amount of computation and produced a simple yes/no answer. In [40], G. Hanrot developed together with E. Martin-Dorel (Toccata team), M. Mayero (IUT Villetaneuse, LIPN), and L. Théry (Marelle team) a certificate-based approach of the SLZ algorithm where the execution produces certificates which can then be validated using Coq. This allows one to validate a posteriori the fact that for a given function, a given input precision p and bound p', there is no pair (x, y) of floating-point representable numbers in precision p such that $2^{-e_p(f(x))}|f(x) - y| \leq 2^{-p'}$. This approach has been tested on the exponential function over $[1/2, 1]$, with an input precision of 53 bits and $p' = 300$.

6.4. Hardware and FPGA arithmetic

6.4.1. *Reconfiguring arithmetic*

With B. Pasca (Altera), F. de Dinechin contributed a book chapter about of the opportunities and challenges of computer arithmetic for reconfigurable/FPGA computing [32]. The main point of this chapter is to look beyond the heritage of processor arithmetic. Using many examples from the FloPoCo project and others, it shows the benefits of merging and fusing standard operators, it introduces an open-ended space of non-standard operators, and illustrates the power of machine-generation of such arithmetic cores.

6.4.2. *The bit heap framework for fixed-point arithmetic*

N. Brunie, F. de Dinechin, and M. Istoan, with students G. Sergent, K. Illyes, and B. Popa, extended FloPoCo with a versatile framework for manipulating sums of weighted bits [28], [18]. Such bit heaps may be used to express and optimize at the bit level a wide range of operators (from adders and multipliers to polynomials, filters, and other coarse arithmetic cores). A single piece of code can then be used to generate an architecture for any of these operators.

6.4.3. *Elementary functions*

F. de Dinechin, with P. Echeverria and M. Lopez-Vallejo (U. Madrid) and B. Pasca (Altera), published a hardware architecture for the floating-point pow and powr functions of the IEEE-754-2008 standard [3]. These functions compute x^y, and differ only in the specification of special cases. The implementation, distributed in FloPoCo, is parameterized in exponent and significand size. It combines suitably modified exponential and logarithm units.

F. de Dinechin and M. Istoan, with student G. Sergent, compared several hardware algorithms for the implementation of sine, cosine, and combined sine/cosine [21]: unrolled CORDIC in two variants with several minor improvements, polynomial approximation, and an ad-hoc architecture based on trigonometric identities. A surprising result is that the ad-hoc architecture betters CORDIC even when its multipliers and tables are synthesized as logic.

6.4.4. *Contributions to processor architecture*

S. Collange (ALF team) and N. Brunie with G. Diamos (Nvidia) suggested improvements for the architecture of general-purpose graphical processing units [11]. As threads take different paths across the control-flow graph, SIMD lockstep execution is partially lost, and must be regained whenever possible in order to maximize the occupancy of SIMD units. Two techniques are described to handle SIMT control divergence and identify reconvergence points. The most advanced one operates in constant space and handles indirect jumps and recursion. In terms of performance, this solution is at least as efficient as state-of-the-art techniques in use in current GPUs.

N. Brunie and F. de Dinechin studied with B. de Dinechin (Kalray) the integration of a tightly coupled reconfigurable accelerator in a massively parallel multiprocessor [27]. For this purpose, they described an architecture exploration framework that produces an architecture along with the relevant compilation software. This framework was demonstrated on AES, SHA2, and a FIR filter.

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

6. New Results

6.1. Energy efficiency of large scale distributed systems

Participants: Ghislain Landry Tsafack Chetsa, Mohammed El Mehdi Diouri, Jean-Patrick Gelas, Olivier Glück, Laurent Lefèvre, François Rossigneux.

6.1.1. Analysis and Evaluation of Different External and Internal Power Monitoring Devices for a Server and a Desktop Machine

Large-scale distributed systems (e.g., datacenters, HPC systems, clouds, large-scale networks, etc.) consume and will consume enormous amounts of energy. Therefore, accurately monitoring the power and energy consumption of these systems is increasingly more unavoidable. The main novelty of this contribution [15] is the analysis and evaluation of different external and internal power monitoring devices tested using two different computing systems, a server and a desktop machine. Furthermore, we also provide experimental results for a variety of benchmarks which exercise intensively the main components (CPU, Memory, HDDs, and NICs) of the target platforms to validate the accuracy of the equipment in terms of power dispersion and energy consumption. We highlight that external wattmeters do not offer the same measures as internal wattmeters. Thanks to the high sampling rate and to the different measured lines, the internal wattmeters allow an improved visualization of some power fluctuations. However, a high sampling rate is not always necessary to understand the evolution of the power consumption during the execution of a benchmark.

6.1.2. Your Cluster is not Power Homogeneous

Future supercomputers will consume enormous amounts of energy. These very large scale systems will gather many homogeneous clusters. We analyze the power consumption of the nodes from different homogeneous clusters during different workloads. As expected, we observe that these nodes exhibit the same level of performance. However, we also show that different nodes from a homogeneous cluster may exhibit heterogeneous idle power energy consumption even if they are made of identical hardware. Hence, we propose an experimental methodology to understand such differences. We show that CPUs are responsible for such heterogeneity which can reach 20% in terms of energy consumption. So energy aware (Green) schedulers must take care of such hidden heterogeneity in order to propose efficient mapping of tasks. To consume less energy, we propose an energy-aware scheduling approach taking into account the heterogeneous idle power consumption of homogeneous nodes [20]. It shows that we are able to save energy up to 17% while exploiting the high power heterogeneity that may exist in some homogeneous clusters.

6.1.3. Energy Consumption Estimations of Fault Tolerance protocols

Energy consumption and fault tolerance are two interrelated issues to address for designing future exascale systems. Fault tolerance protocols used for checkpointing have different energy consumption depending on parameters like application features, number of processes in the execution and platform characteristics. Currently, the only way to select a protocol for a given execution is to run the application and monitor the energy consumption of different fault tolerance protocols. This is needed for any variation of the execution setting. To avoid this time and energy consuming process, we propose an energy estimation framework [16], [17], [7]. It relies on an energy calibration of the considered platform and a user description of the execution setting. We evaluate the accuracy of our estimations with real applications running on a real platform with energy consumption monitoring. Results show that our estimations are highly accurate and allow selecting the best fault tolerant protocol without pre-executing the application.

6.1.4. Energy Consumption Estimations of Data Broadcasting

Future supercomputers will gather hundreds of millions of communicating cores. The movement of data in such systems will be very energy consuming. We address the issue of energy consumption of data broadcasting in such large scale systems. To this end, in [19], [7], we propose a framework to estimate the energy consumed by different MPI broadcasting algorithms for various execution settings. Validation results show that our estimations are highly accurate and allow to select the least consuming broadcasting algorithm.

6.1.5. A Smart-Grid Based Framework for Consuming Less and Better in Extreme-Scale Infrastructures

As they will gather hundreds of million cores, future exascale supercomputers will consume enormous amounts of energy. Besides being very important, their power consumption will be dynamic and irregular. Thus, in order to consume energy efficiently, powering such systems will require a permanent negotiation between the energy supplier and one of its major customers represented by exascale platforms. We have designed SESAMES [18], [53], a smart and energy-aware service-oriented architecture manager that proposes energy-efficient services for exascale applications and provides an optimized reservation scheduling. The new features of this framework are the design of a smart grid and a multi-criteria green job scheduler. Simulation results show that with the proposed multi-criteria job scheduler, we are able to save up to 2.32 % in terms of energy consumption, 24.22 % in terms of financial cost and reduce up to 7.12 % the emissions of CO_2.

6.1.6. Clustered Virtual Home Gateway (vHGW)

This result is a joint work between Avalon team (J.P. Gelas, L. Lefevre) and Addis Abeba University (M. Tsibie and T. Assefa). The customer premises equipment (CPE), which provides the interworking functions between the access network and the home network, consumes more than 80% of the total power in a wireline access network. In the GreenTouch initiative (cf Section 7.3), we aim at a drastic reduction of the power consumption by means of a passive or quasi-passive CPE. Such approach requires that typical home gateway functions, such as routing, security, and home network management, are moved to a virtual home gateway (vHGW) server in the network. In our first prototype virtual home gateways of the subscribers were put in LXC containers on a unique GNU/Linux server. The container approach is more scalable than separating subscribers by virtual machines. We demonstrated a sharing factor of 500 to 1000 virtual home gateways on one server, which consumes about 150 W, or 150 to 300 mW per subscriber. Comparing this power consumption with the power of about 2 W for the processor in a thick client home gateway, we achieved an efficiency gain of 5-10x. The prototype was integrated and demonstrated at TIA 2012 in Dallas. In our current work, we propose the Clustered vHGWs Data center architecture to yield optimal energy conservation through virtual machine's migration among physical nodes based on the current subscriber's service access state, while ensuring SLA respective subscribers. Thus, optimized energy utilization of the data center is assured without compromising the availability of service connectivity and QoS preferences of respective subscribers.

6.1.7. Improving Energy Efficiency of Large Scale Systems without a priori Knowledge of Applications and Services

Unlike their hardware counterpart, software solutions to the energy reduction problem in large scale and distributed infrastructures hardly result in real deployments. At the one hand, this can be justified by the fact that they are application oriented. At the other hand, their failure can be attributed to their complex nature which often requires vast technical knowledge behind proposed solutions and/or thorough understanding of applications at hand. This restricts their use to a limited number of experts, because users usually lack adequate skills. In addition, although subsystems including the memory and the storage are becoming more and more power hungry, current software energy reduction techniques fail to take them into account. We propose a methodology for reducing the energy consumption of large scale and distributed infrastructures. Broken into three steps known as (i) phase identification, (ii) phase characterization, and (iii) phase identification and system reconfiguration; our methodology abstracts away from any individual applications as it focuses on the infrastructure, which it analyses the runtime behaviour and takes reconfiguration decisions accordingly.

The proposed methodology is implemented and evaluated in high performance computing (HPC) clusters of varied sizes through a Multi-Resource Energy Efficient Framework (MREEF). MREEF implements the proposed energy reduction methodology so as to leave users with the choice of implementing their own system reconfiguration decisions depending on their needs. Experimental results show that our methodology reduces the energy consumption of the overall infrastructure of up to 24% with less than 7% performance degradation. By taking into account all subsystems, our experiments demonstrate that the energy reduction problem in large scale and distributed infrastructures can benefit from more than "the traditional" processor frequency scaling. Experiments in clusters of varied sizes demonstrate that MREEF and therefore our methodology can easily be extended to a large number of energy aware clusters. The extension of MREEF to virtualized environments like cloud shows that the proposed methodology goes beyond HPC systems and can be used in many other computing environments.

6.1.8. Reservation based Usage for Energy Efficient Clouds: the Climate Architecture

The FSN XLcloud project (cf Section 7.1) strives to establish the demonstration of a High Performance Cloud Computing (HPCC) platform based on OpenStask, that is designed to run a representative set of compute intensive workloads, including more specifically interactive games, interactive simulations and 3D graphics. XLcloud is based on OpenStack, and Avalon is contributing to the energy efficiency part of this project. We have proposed and brought our contribution to Climate, a new resource reservation framework for OpenStack, developed in collaboration with Bull, Mirantis and other OpenStack contributors. Climate allows the reservation of both physical and virtual resources, in order to provide a mono-tenancy environment suitable for HPC applications. Climate chooses the most efficient hosts (flop/W). This metric is computed from the CPU / GPU informations, mixed with real power consumption measurements provided by the Kwapi framework. The user requirements may be loose, allowing Climate to choose the best time slot to place the reservation. Climate will be improved with standby mode features, to shut down automatically the unused hosts. The first release of Climate is planned at the end of January 2014, and we expect an incubation in the next version of OpenStack.

6.2. Simulation of Large Scale Distributed Systems

Participants: Frédéric Desprez, Jonathan Rouzaud-Cornabas, Frédéric Suter.

6.2.1. Toward Better Simulation of MPI Applications on Ethernet/TCP Networks

Simulation and modeling for performance prediction and profiling is essential for developing and maintaining HPC code that is expected to scale for next-generation exascale systems, and correctly modeling network behavior is essential for creating realistic simulations. In [11], we proposed an implementation of a flow-based hybrid network model that accounts for factors such as network topology and contention, which are commonly ignored by other approaches. We focused on large-scale, Ethernet-connected systems, as these currently compose 37.8% of the TOP500 index, and this share is expected to increase as higher-speed 10 and 100GbE become more available. The European Mont-Blanc project that studies exascale computing by developing prototype systems with low-power embedded devices will also use Ethernet-based interconnect. Our model is implemented within SMPI, an open-source MPI implementation that connects real applications to the SIMGRID simulation framework (cf Section 5.5). SMPI provides implementations of collective communications based on current versions of both OpenMPI and MPICH. SMPI and SIMGRID also provide methods for easing the simulation of large-scale systems, including shadow execution, memory folding, and support for both online and offline simulation. We validated our proposed model by comparing traces produced by SMPI with those from real world experiments, as well as with those obtained using other established network models. Our study shows that SMPI has a consistently better predictive power than classical LogP-based models for a wide range of scenarios including both established HPC benchmarks and real applications.

6.2.2. SimGrid: a Sustained Effort for the Versatile Simulation of Large Scale Distributed Systems

SIMGRID (cf Section 5.5) is a toolkit for the versatile simulation of large scale distributed systems, whose development effort has been sustained for the last fifteen years. Over this time period SIMGRID has evolved from a one-laboratory project in the U.S. into a scientific instrument developed by an international collaboration. The keys to making this evolution possible have been securing of funding, improving the quality of the software, and increasing the user base. We detailed in [55]how we have been able to make advances on all three fronts, on which we plan to intensify our efforts over the upcoming years.

6.2.3. *Simulating Multiple Clouds from a Client Point of View: SGCB an AWS Simulator*

Validating a new application over a Cloud is not an easy task and it can be costly over public Clouds. Simulation is a good solution if the simulator is accurate enough and if it provides all the features of the target Cloud. In [49], we have proposed an extension of the SIMGRID simulation toolkit to simulate the Amazon IaaS Cloud. Based on an extensive study of the Amazon platform and previous evaluations, we have integrated models into the SIMGRID Cloud Broker and exposed the same API as Amazon to the users. Our experimental results have shown that our simulator is able to simulate different parts of Amazon for different applications.

6.3. Active Data: A Data-Centric Approach to Data Life-Cycle Management

Participants: Gilles Fedak, Anthony Simonet.

Data-intensive science offers new opportunities for innovation and discoveries, provided that large datasets can be handled efficiently. Data management for data-intensive science applications is challenging; requiring support for complex data life cycles, coordination across multiple sites, fault tolerance, and scalability to support tens of sites and petabytes of data. In [28], we argue that data management for data-intensive science applications requires a fundamentally different management approach than the current ad-hoc task centric approach. We propose Active Data, a fundamentally novel paradigm for data life cycle management. Active Data follows two principles: data-centric and event-driven. We report on the Active Data programming model and its preliminary implementation, and discuss the benefits and limitations of the approach on recognized challenging data-intensive science use-cases.

6.4. HPC Component Model

Participants: Zhengxiong Hou, Vincent Lanore, Christian Perez.

6.4.1. *Auto-tuning of Stencil Based Applications*

We have finished designing a tuning approach for stencil applications on multi-core clusters [25]. We focused in particular on a 2D Jacobi benchmark application as well as memory bandwidth performance. The tuning approach includes data partitioning within one node, the selection of the number of threads within a multi-core node, a data partitioning for multi nodes, and the number of nodes for a multi-core cluster. This model is based on a set of experiments on machines of GRID'5000 and on the Curie supercomputer.

6.4.2. *Static 2D FFT Adaptation through a Component Model based on Charm++*

Adaptation algorithms for HPC applications can improve performance but their implementation is often costly in terms of development and maintenance. Component models such as Gluon++, which is built on top of Charm++, propose to separate the business code, encapsulated in components, and the application structure, expressed through a component assembly. Adaptation of component-based HPC applications can be achieved through the optimization of the assembly. We have studied such an approach with the adaptation to network topology and data size of a Gluon++ 2D FFT application. Preliminary experimental results obtained on the GRID'5000 platform show the suitability of the proposed approach.

6.4.3. Towards Scalable Reconfiguration in Component Models

Some HPC applications require reconfiguration of their architecture at runtime; examples include adapting to (cloud) resource elasticity, efficient distributed deployment, Adaptive Mesh Refinement (AMR), and load balancing. This class of applications raises challenges such as handling of concurrent reconfigurations and distributed architecture representation at runtime. To our knowledge, no existing programming model addresses those challenges in the general case with both high programmability and scalability. We have identified a list of specific subproblems and use-cases and we have devised a preliminary component model to address some of them.

6.5. Resource Management and Scheduling

Participants: Eddy Caron, Frédéric Desprez, Gilles Fedak, Jose Luis Lucas, Christian Perez, Jonathan Rouzaud-Cornabas, Frédéric Suter.

6.5.1. Resource Management Architecture for Fair Scheduling of Optional Computations

Most High-Performance Computing platforms require users to submit a pre-determined number of computation requests (also called jobs). Unfortunately, this is cumbersome when some of the computations are optional, i.e., they are not critical, but their completion would improve results. For example, given a deadline, the number of requests to submit for a Monte Carlo experiment is difficult to choose. The more requests are completed, the better the results are, however, submitting too many might overload the platform. Conversely, submitting too few requests may leave resources unused and misses an opportunity to improve the results.

In cooperation with IRIT (Toulouse), we have proposed a generic client-server architecture and an implementation in DIET, a production GridRPC middleware, which auto-tunes the number of requests [12]. Real-life experiments show significant improvement of several metrics, such as user satisfaction, fairness and the number of completed requests. Moreover, the solution is shown to be scalable.

6.5.2. Advanced Promethee-based Scheduler Enriched with User-Oriented Methods

Efficiently scheduling tasks in hybrid Distributed Computing Infrastructures (DCI) is a challenging pursue because the scheduler must deal with a set of parameters that simultaneously characterize the tasks and the hosts originating from different types of infrastructure. In [27], we propose a scheduling method for hybrid DCIs, based on advanced multi-criteria decision methods. The scheduling decisions are made using pairwise comparisons of the tasks for a set of criteria like expected completion time and price charged for computation. The results are obtained with an XtremWeb-like pull-based scheduler simulator using real failure traces for a combination of three types of infrastructure. We also show how such a scheduler should be configured to enhance user satisfaction regardless their profiles, while maintaining good values for makespan and cost. We validate our approach with a statistical analysis on empirical data and show that our proposed scheduling method improves performance by 12-17% compared to other scheduling methods. Experimenting on large time-series and using realistic scheduling scenarios lead us to conclude about time consistency results of the method.

6.5.3. Fair Resource Sharing for Dynamic Scheduling of Workflows on Heterogeneous Systems

Scheduling independent workflows on shared resources in a way that satisfy users Quality of Service is a significant challenge. In [37], we described methodologies for off-line scheduling, where a schedule is generated for a set of known workflows, and on-line scheduling, where users can submit workflows at any moment in time. We consider the on-line scheduling problem in more detail and present performance comparisons of state-of-the-art algorithms for a realistic model of a heterogeneous system.

6.5.4. Image Transfer and Storage Cost Aware Brokering Strategies for Multiple Clouds

Nowadays, Clouds are used for hosting a large range of services. But between different Cloud Service Providers, the pricing model and the price of individual resources are very different. Furthermore hosting a service in one Cloud is the major cause of service outage. To increase resiliency and minimize the monetary cost of running a service, it becomes mandatory to span it between different Clouds. Moreover, due to dynamicity of both the service and Clouds, it could be required to migrate a service at run time. Accordingly, this ability must be integrated into the multi-Cloud resource manager, *i.e.* the Cloud broker. But, when migrating a VM to a new Cloud Service Provider, the VM disk image must be migrated too. Accordingly, data storage and transfer must be taken into account when choosing if and where an application will be migrated.

In [47], we have extended a cost-optimization algorithm to take into account storage costs to approximate the optimal placement of a service. The data storage management consists in taking two decisions: where to upload an image, and keep it on-line during the experiment lifetime or delete it when unused. Based on our experimentations, we have shown that the storage cost of VM disk image must not be neglected as done in previous work. Moreover, we have shown that using the accurate combinations of storage policies can dramatically reduce the storage cost (from 90% to 14% of the total bill).

6.6. Security for Virtualization and Clouds

Participants: Eddy Caron, Arnaud Lefray, Jonathan Rouzaud-Cornabas.

6.6.1. Improving Users' Isolation in IaaS: Virtual Machine Placement with Security Constraints

Nowadays virtualization is used as the sole mechanism to isolate different users on Cloud platforms. Due to improper virtualization of micro-architectural components, data leak and modification can occur on public Clouds. Moreover, using the same attack vector (improper virtualization of micro-architectural components), it is possible to induce performance interferences, *i.e.* noisy neighbors. Using this approach, a VM can slow down and steal resources from concurrent VMs. In [43], we have proposed placement heuristics that take into account isolation requirements. We have modified three classical heuristics to take into account these requirements. Furthermore, we have proposed four new heuristics that take into account the hierarchy of the Cloud platforms and the isolation requirements. Finally, we have evaluated these heuristics and compare them with the modified classical ones. We have shown that our heuristics are performing at least as good as classical ones but are scaling better and are faster by a few order of magnitude than the classical ones.

6.6.2. Security for Cloud Environment through Information Flow Properties Formalization with a First-Order Temporal Logic

The main slowdown of Cloud activity comes from the lack of reliable security. The on-demand security concept aims at delivering and enforcing the client's security requirements. In [50], we have presented an approach, Information Flow Past Linear Time Logic (IF-PLTL), to specify how a system can support a large range of security properties. We have presented how to control those information flows from lower system events. We have given complete details over IF-PLTL syntax and semantics. Furthermore, that logic enables to formalize a large set of security policies. Our approach is exemplified with the Chinese Wall commercial-related policy. Finally, we have discussed the extension of IF-PLTL with dynamic relabeling to encompass more realistic situations through the dynamic domains isolation policy.

6.6.3. Security Metrics for the Cloud Computing and Security-aware Virtual Machine Placement

In a classic Cloud Computing scenario, a client connects to a provider platform/service and submits his computation requirements, sometimes known as Service Level Agree- ments (SLAs). Then, the platform executes the computation taking into account, in its allocation algorithms, criteria like data location, CPU usage or duration of a job. As security in Cloud Computing is a main concern, we propose to consider security as another criteria for jobs scheduling. Thus, two questions need to be answered. The first one is how a client

can describe his needs in terms of security level and the second one is how the scheduler could leverage the security to satisfy the client requirements? To provide an answer, a system of security metrics is essential. Indeed, with appropriate metrics, we can quantify and compare the security level of our resources. Moreover, a client can easily describe his security requirements and the scheduler can allocate the fitted resources using these metrics. Unfortunately, such system of metrics is not yet available. Consequently, we developed a system of security metrics specific to the Cloud Computing and scheduling algorithms using these metrics for a Security-Aware Virtual Machine (VM) placement.

6.7. Self-healing of Operational Issues for Grid Computing

Participant: Frédéric Desprez.

Many scientists now formulate their computational problems as scientific workflows. Workflows allow researchers to easily express multi-step computational task. However, their large scale and the number of middleware systems involved in these gateways lead to many errors and faults. Fair quality of service (QoS) can be delivered, yet with important human intervention. Automating such operations is challenging for two reasons. First, the problem is online by nature because no reliable user activity prediction can be assumed, and new workloads may arrive at any time. Therefore, the considered metrics, decisions and actions have to remain simple and to yield results while the application is still executing. Second, it is non-clairvoyant due to the lack of information about applications and resources in production conditions. Computing resources are usually dynamically provisioned from heterogeneous clusters, clouds or desktop grids without any reliable estimate of their availability and characteristics. Models of application execution times are hardly available either, in particular on heterogeneous computing resources.

In collaboration with Rafaël Silva and Tristan Glatard, we proposed a general self-healing process for autonomous detection and handling of operational incidents in scientific workflow executions on grids. Instances are modeled as Fuzzy Finite State Machines (FuSM) where state degrees of membership are determined by an external healing process. Degrees of membership are computed from metrics assuming that incidents have outlier performance, e.g. a site or a particular invocation behaves differently than the others. These metrics make little assumptions on the application or resource characteristics. Based on incident degrees, the healing process identifies incident levels using thresholds determined from the platform history. A specific set of actions is then selected from association rules among incident levels. The healing process is parametrized on real application traces acquired in production on the European Grid Infrastructure (EGI).

To optimize task granularity in distributed scientific workflows, we presented a method that groups tasks when the fineness degree of the application becomes higher than a threshold determined from execution traces. Controlling the granularity of workflow activities executed on grids is required to reduce the impact of task queuing and data transfer time. Our method groups tasks when the fineness degree of the application, which takes into account the ratio of shared data and the queuing/round-trip time ratio, becomes higher than a threshold determined from execution traces. The algorithm also de-groups task groups when new resources arrive. Results showed that under stationary load, our fineness control process significantly reduces the makespan of all applications. Under non-stationary load, task grouping is penalized by its lack of adaptation, but our de-grouping algorithm corrects it in case variations in the number of available resources are not too fast [21].

To address unfairness among workflow executions, we proposed an algorithm to fairly allocate distributed computing resources among workflow executions to multi-user platforms. We consider a non-clairvoyant, online fairness problem where the platform workload, task costs, and resource characteristics are unknown and not stationary. We define a novel metric that quantifies unfairness based on the fraction of pending work in a workflow. It compares workflow activities based on their ratio of queuing tasks, their relative durations, and the performance of resources where tasks are running, as information becomes available during the execution. Our method is implemented and evaluated on 4 different applications executed in production conditions on EGI. Results show that our method can very significantly reduce the standard deviation of the slowdown, and the average value of our metric [22].

BAMBOO Project-Team

6. New Results

6.1. Symbiont genome evolution and dynamics

The objective of this part of our work was to analyse genome rearrangements and dynamics. The results obtained were both algorithmic and biological.

In terms of algorithms, we developed a new method for repeat identification (RIME) [12], as well as an algorithm for finding the minimum number of three constrained versions of inversions that transform one given genome into another [25]. The constrained versions concerned symmetric, almost-symmetric and unitary inversions. The genome rearrangement algorithm is not exact: it is based on a greedy randomized search procedure to find such minimum number of constrained inversions.

The main set of biological results [4], [14] concerned trypanosomatids of the genera *Angomonas* and *Strigomonas* that live in a mutualistic association characterised by extensive metabolic cooperation with obligate endosymbiotic Betaproteobacteria. In contrast to their counterparts lacking symbionts, such trypanosomatids exhibit lower nutritional requirements and are autotrophic for essential amino acids and vitamins. Phylogenetic analyses showed that the cooperation in the first case is complemented by multiple horizontal gene transfers, from bacterial lineages to trypanosomatids, that appear to have occurred several times in the course of evolution. In contrast, but for three exceptions, such transfers are absent as concerns vitamin biosynthesis.

The above work was made possible in part because of the sequencing and annotation of the genomes whose metabolic pathways could then be inferred. We participated in these for some of the genomes involved in the above study [17].

6.2. Host-symbiont metabolic dialog

The methodological work done has covered one main question concerning what we called metabolic stories. Given a subset of metabolites representing those monitored as being under- or over-produced in some condition (e.g., interaction with a parasite) and a metabolic network represented as a compound graph, metabolic stories are maximal directed acyclic graphs (DAGs) that cover all the metabolites in the subset of interest, and have all sources and targets among these metabolites. One exact algorithm (TOUCHÉ, [24]) was developed to enumerate all metabolic stories that improved on our previous method (GOBBOLINO).

The algorithm above was validated on biological data [16] in a study of the response of yeast to cadmium exposure. We used this system as a proof of concept for our method and we showed that we are able to find a story that reproduces very well the current knowledge about the yeast response to cadmium. We further showed that this response is mostly based on enzyme activation. We also provided a framework for exploring the alternative pathways or side effects this local response is expected to have in the rest of the network. Finally, we discussed several interpretations for the changes we see and we suggest hypotheses which could in principle be experimentally tested.

6.3. Host-symbiont genetic dialog

Two sets of problems were addressed: (i) the development of algorithms for analysing NGS data especially RNAseq, and (ii) the development of algorithms for identifying small RNAs, notably microRNAs, and their targets.

The computational work on NGS is described in another section.

Computational work on small RNAs, initially miRNAs, led to the development of a new algorithmic method. This builds upon previously developed approaches, one which was applied to *Anopheles darlingi* for inferring miRNAs that however had a high rate of false positives, and a second that provided a way for navigating among all the candidates found. Recently however, we arrived at a better model for such inference in the double sense that the rate of false positives is smaller without losing in sensitivity, while the method is much faster. The paper presenting this work and the algorithm (MIRINHO) was submitted and is currently in revision.

6.4. Symbiont-host co-cladogenesis and co-evolution at the sequence and network levels

The problem here was to: (i) study the co-evolution of a set of hosts and their symbionts, and (ii) to understand the genetic architecture of a parasitic invasion by investigating the different phenotypes such invasion produces in the host.

Work on the first point took longer than initially planned but two papers are now submitted. In the first, titled "Co-phylogeny Reconstruction via an Approximate Bayesian Computation", we describe an algorithm (COALA) for estimating the frequency of co-evolutionary events based on a likelihood-free approach. 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 to estimate the frequency of the events in cases where the dataset consists of trees with a large number of taxa. We evaluate our method on simulated and on real datasets. We show that in both cases, for a same pair of host and parasite trees, different sets of frequencies for the events constitute equally probable solutions. Moreover, sometimes these sets lead to different parsimonious optimal reconciliations, in the sense of presenting a different number of the events. For this reason, it appears crucial to take this into account before attempting any further biological interpretation of such reconciliations. More generally, we also show 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.

In the second submitted paper related to the study of co-evolution and titled "EUCALYPT: Efficient tree reconciliation enumerator", we present a polynomial-delay algorithm for enumerating all optimal reconciliations. We show that in general many optimal solutions exist. We give an example where, for two pairs of host-parasite trees having each less than 40 leaves, the number of solutions is 2309, even when only time feasible solutions are kept. To facilitate their interpretation, those solutions are also classified in terms of the number of each event that they contain. This often enables to reduce considerably the number of different classes of solutions to examine further, but the number may remain high enough (16 for the same example). Depending on the cost vector, both numbers may increase considerably (for the same instance, to respectively 4080384 and 275).

Concerning the second question (genetic architecture of a parasitic invasion), one such phenotype is called "cytoplasmic incompatibility" (CI). Briefly, when a parasite invades a male host, it induces the death of the host's offspring unless the female is also infected. This has been explained by a toxin/antitoxin model that involves a toxin deposited by the parasites in the male's sperm inducing the death of the zygote unless neutralised by an antidote produced by the parasites in the egg. One toxin/antitoxin pair is usually linked to one genetic factor. Given a set of observed CIs, the question is how many genetic factors explain it. In its simplest form, this mathematically translates into, given a bipartite graph, finding its minimum biclique edge cover. One biclique corresponds to one factor. We had previously analysed the complexity of the problem and proposed an algorithm that was this year applied to a set of CI data from *Culex pipiens* [18].

6.5. NGS for biodiversity

In collaboration with the Laboratoire d'Écologie Alpine (LECA) at Grenoble where there is a strong expertise on DNA meta-barcoding, we had devised several tools for barcode design and analysis. ECOPRIMERS thus identified new barcode markers and their associated PCR primers within a DNA meta-barcoding approach. The algorithm was optimised two quality indexes measuring taxonomical range and discrimination to select the most efficient markers from a set of reference sequences, according to some experimental constraints such as marker length or specifically targeted taxa. We had also devised assembler algorithms directed to

organelles (mitochondria or chloroplasts). This year, in collaboration with the Inria project-team MISTIS, we developed a statistical modelling approach to investigate the spatial cross-correlations between different taxa identified by meta-barcoding of soil sample from French Guiana (this was selected as a conference paper at the "45ème Journées de Statistiques" 2013 that took place at Toulouse and is organised by the Société Française de Statistique. This approach allows to visualise the co-occurrence pattern as a "species interaction graph", and to study the mutual exclusion (competition) or inclusion (symbiosis) of different plant species.

6.6. NGS for genotypic variation detection

The computational work on NGS data concerned both algorithmic design and complexity analysis.

Based on the idea that each genotypic variation will correspond to a recognisable pattern in a de Bruijn graph constructed from a set of sequence reads, we had proposed a generic model for SNPs in DNA data, and then generalised it to the analysis of RNA. In this case, not only SNPs are present but also alternative splicing (AS) events, which, once again, generate a recognisable pattern in the de Bruijn Graph. We had therefore proposed a general model for all these variations (SNPs, indels and AS events) and introduced an exact algorithm (KISSPLICE) to extract all alternative splicing events. The algorithm also outputs candidate SNPs and indels. This year, we improved the algorithm [26]. As the problem relates to an old one in algorithmics (cycle enumeration), we also revisited it from a theoretical point of view [23].

The improved version of KISSPLICE [26] was used to analyse RNAseq data from two lines of *Asobara tabida* exhibiting different ovarian phenotypes in the absence of its endosymbiont *Wolbachia*. Although infected individuals of the two lines have similar phenotypes, numerous genes are differentially expressed between the two infected conditions. This could mean that two divergent strategies of tolerance have evolved. Preliminary results on the analysis of polymorphisms between these two lines suggest that differentially expressed genes tend to accumulate more variation. We are currently, via experiments done by the biologists in our team, testing the hypothesis that such genes are under strong selection pressure and may evolve through mutation accumulation, a process that could be related to assimilation.

A preliminary analysis of human data from the ENCODE project performed with KISSPLICE showed that an assembly-based method (without reference genome) is able to recover AS events that are missed by mapping-based methods (with a reference genome). Some of these events were experimentally validated, which represents the best type of proof we can provide to the biologists. The experimental part is made by our collaborator from the Inserm, Didier Auboeuf, in his team at the Centre National de Cancérologie of Lyon (CNCL), with whom we had an Inserm project, EXOMIC, funded for three years starting from 2012.

The identification of SNPs is also getting renewed interest even in the presence of a reference genome thanks to the possibility of re-sequencing many times the genome of a same or of very closely related species. The difficulty in the case of SNPs is to distinguish them from sequencing errors and from inexact repeats. We proposed a statistical test enabling to identify variations that are condition-specific, which enables to greatly enrich the list of potential SNP candidates. The paper on this test is in preparation. Its results as applied to the RNAseq data from two lines of *Asobara tabida* (see above) and to Drosophila species having diverged very recently were validated by, respectively, Fabrice Vavre and Cristina Vieira, both members of BAMBOO.

We also started addressing the problem that repeats (such as transposable elements for instance but not only) represent more in general for both local and global assemblers. We are thus developing a method that would enable to identify, in a de Bruijn graph built from RNAseq data, the vertices potentially corresponding to the borders of a repeated sequence. Preliminary results on simulated and real data show that the approach is promising (paper in preparation).

<div align="center">

BEAGLE Project-Team

</div>

6. New Results

6.1. Stochastic dynamics of gene expression

A number of studies have established that stochasticity in gene expression may play an important role in many biological phenomena but the molecular mechanisms at stake are still poorly understood. By joint experimental and computational approaches, we explored the role played by chromatin dynamics in the regulation of stochastic gene expression in higher eukaryotic cells [31]. For this purpose, our biological partner generated isogenic chicken-cell populations expressing a fluorescent reporter integrated in one copy per clone. Although the clones differed only in the genetic locus at which the reporter was inserted, they showed markedly different fluorescence distributions, revealing different levels of stochastic gene expression. Use of chromatin-modifying agents then showed that direct manipulation of chromatin dynamics had a marked effect on the extent of stochastic gene expression. We then fitted the experimental data to a two-state model describing the opening/closing process of the chromatin. The model showed that the differences between clones seemed to be due mainly to the duration of the closed state, and that the agents we used mainly seem to act on the opening probability. These results highlight the importance of chromatin dynamics in stochastic gene expression. They shed a new light on the mechanisms of gene expression in higher eukaryotic cells, and argues in favor of relatively slow dynamics with long (hours to days) periods of quiet state.

This work was part of Gaël Kaneko's PhD and results from our long-lasting collaboration with Olivier Grandrillon and his BM2A team in the CGphyMC (Centre de Génétique et de Physiologie Moléculaire et Cellulaire, Lyon).

6.2. The impact of anomalous diffusion on cell signaling

This year, we published two papers describing the impact of diffusion and clustering in membrane domains for ubiquitous biological pathways. In the first paper [18], we showed that clustering of receptors in membrane domains affect severely the activation of 'hit and run' type of pathways (eg. IRS1, G proteins). This is a pure diffusion result and it is obtained without modifying molecular affinity. This impairment is dramatically important when receptors are highly clustered such are the cases for insulin and adrenergic receptors. In the same direction, we studied the impact of modified diffusion on the other ubiquitous pathway: enzyme/substrate equilibrium [10]. In that case diffusion was modified either using subdiffusion - obstacles and Continuous Time Random Walk - or using space-based inhomogeneous diffusion. We showed that while impairing diffusion all three mechanisms behave differently in the stationary regime. Therefore it it not possible to assume simple space-dependent diffusion for subdiffusion at the equilibrium limit (as it is always assumed). Furthermore, we showed that in the case of space-dependent diffusion - the shape of the diffusion profile can drastically affect the equilibrium and modify the pathway. Since these three phenomenons are thought to occur either in the membrane or the cytosplasm, our results show they can have non trivial effect of all chemical reaction occurring within these medium.

This research will be carried on in the group in 2014 and expanded by the more mathematical approaches initiated in 2013 by collaborations with V Calvez (Numed Inria Lyon), T Lepoutre (Dracula, INIA Lyon) and S. Fedotov (Univ Manchester, UK).

6.3. Localization of protein aggregates in E. coli

Aggregates of misfolded proteins are a hallmark of many age-related diseases. Recently, they have been linked to aging of *Escherichia coli* (*E. coli*) where protein aggregates accumulate at the old pole region of the aging bacterium. Because of the potential of *E. coli* as a model organism, elucidating aging and protein aggregation in this bacterium may pave the way to significant advances in our global understanding of aging. A first obstacle

along this path is to decipher the mechanisms by which protein aggregates are targeted to specific intercellular locations. Here, using an integrated approach based on individual-based modeling, time-lapse fluorescence microscopy and automated image analysis, we show that the movement of aging-related protein aggregates in *E. coli* is purely diffusive (Brownian). Using single-particle tracking of protein aggregates in live *E. coli* cells, we estimated the average size and diffusion constant of the aggregates. Our results provide evidence that the aggregates passively diffuse within the cell, with diffusion constants that depend on their size in agreement with the Stokes-Einstein law. However, the aggregate displacements along the cell long axis are confined to a region that roughly corresponds to the nucleoid-free space in the cell pole, thus confirming the importance of increased macromolecular crowding in the nucleoids. We thus used 3D individual-based modeling to show that these three ingredients (diffusion, aggregation and diffusion hindrance in the nucleoids) are sufficient and necessary to reproduce the available experimental data on aggregate localization in the cells. Taken together, our results strongly support the hypothesis that the localization of aging-related protein aggregates in the poles of *E. coli* results from the coupling of passive diffusion-aggregation with spatially non-homogeneous macromolecular crowding. They further support the importance of "soft" intracellular structuring (based on macromolecular crowding) in diffusion-based protein localization in *E. coli*.

This work is a collaboration with the microbiology group led by A. Lindner (INSERM U1001, Cochin Med School, Paris). It has been published in [3] as part of A.S. Coquel's PhD (defended Nov 2012, co-supervision H. Berry-A. Lindner).

6.4. The molecular signaling basis of neuronal plasticity

Many of the cell-level properties of the neurons vary as a function of the signals from other neurons or past activity. These modifications are often maintained in the long term, giving rise to cell memory. We have developed models of how the implicated signaling networks self-organize to support a memory and how this leads to cell-level responses such as changes of the firing threshold [24] or the spike-timing dependence [34]. The latter, for instance, corresponds to the observation that the probability of transfer of an electrical signal (spike) between two connected neurons (the synaptic weight) adapts depending on the timing between previous consecutive presynaptic and postsynaptic spikes. Combining a model of the implicated signaling networks with experimental measurements, we have uncovered the molecular mechanisms supporting this memory.

This work is developed in collaboration with both with applied mathematicians (B. Cessac, Inria Neuromath-comp, Sophia-Antipolis) and experimental neurobiologists (L. Venance, Collège de France, Paris).

6.5. A model for adipocyte size based on size-dependent lipid fluxes

We proposed in a paper published this year [28] a novel model that explains some of the peculiarities in the fat tissue storage cells. Indeed, adipocytes, as they are called, come in various size – with up to one order of magnitude in amplitude – but do not possess any characteristic size. The cellularity, the cell size distribution, is bimodal. We showed that a simple model of size-dependent lipid fluxes (using data from Carmen Lab) can explain this bimodality and allow us to retrieve any target cell distribution. Our result also provides an elegant and testable hypothesis for the triggering of adipocytes proliferation. The amount of unstored free fatty acid is actually a marker that the population has reached its maximal volume. This amount could serve as an index to start the proliferation.

This was a joint work with experimentalists from the CARMEN Institute (INSERM UMR1060, Lyon), namely C. Soulage and A. Géloën, and was part of H. Julienne master's thesis.

6.6. Evolution of antibiotic resistance

The emergence of antibiotic resistant bacteria is a major threat to public health and there is a constant need for education to limit dangerous practices. Here, we propose to use alife software to develop training media for the public and the physicians. On the basis of the Aevol model we have been developing for more than six years, we built a game in which players fight bacterial infections using antibiotics. In this game the bacteria can evolve resistance traits, making the infection more and more difficult to cure. The game has been tested

with automatic treatment procedures, showing that it behaves correctly. It was demonstrated during the French "Nuit des Chercheurs" in October 2012 and was published in 2013 in the ECAL conference [2].

This is a joint work with Dominique Schneider from the Laboratoire Adaptation et Pathogénie des Microorganismes (LAPM, UMR CNRS 5163, Grenoble).

6.7. Spontaneous dynamics of genome size

Even though numerous genome sequences are now available, evolutionary mechanisms that determine genome size, notably their fraction of non-coding DNA, are still debated. In particular, although several mechanisms responsible for genome growth (proliferation of transposable elements, gene duplication and divergence, etc.) were clearly identified, mechanisms limiting the overall genome size remain unclear. By using a matrix population model, we showed that genome size can be simply limited by the spontaneous dynamics of duplications and large deletions, which tends to make genomes shrink even if the two types of rearrangements occur at the same rate. In the absence of Darwinian selection, we proved the existence of a stationary distribution of genome size even if duplications are twice as frequent as large deletions. To test whether selection can overcome this spontaneous dynamics, we also simulated our model numerically and chose a fitness function that directly favors genomes containing more genes, while keeping duplications twice as frequent as large deletions. In this scenario where, at first sight, everything seems to favor infinite genome growth, we showed that genome size remains nonetheless bounded. As a result, our study reveals a new pressure that could help limiting genome growth.

This work was part of Stephan Fischer's PhD thesis, which was defended in December 2013. A manuscript is currently under review. Stephan's PhD was co-supervised by Samuel Bernard (Inria Dracula team and Institut Camille Jordan, UMR CNRS 5208, Lyon).

6.8. Inference of evolutionary molecular events at different scales

We have progressed in the integration of several evolutionary events at different scales of genomes in a single model used for inference of ancient events from the observation of extant genomes. We handle nucleotide substitutions, gene duplications, losses, lateral transfers and rearrangements. We have tested the framework on 36 cyanobacteria species, reconstructing up to 80% of ancestral chromosomes in some clades [8]. The inference algorithm is still mainly sequential, in the sense that it first accounts for nucleotide substitutions, gene duplications, losses, lateral transfers [29], and then for rearrangements. But we also developed a way to provide a feedback of the result on rearrangements to the inference of substitutions by correcting gene trees [22], [38]. We have used these methods to reconstruct a nucleotide-scale sequence of the genome of the medieval black death agent [9],[44]. It includes a chromosome and three plasmids, and is different in structure from any extant strain. We follow the first ancient bacterial genome sequencing in 2011 and complete and order the genome with computational predictions. We then dispose of a complete view of the molecular evolution in the *Yersinia pestis* clade.

This work was part of Murray Patterson's post-doctoral fellowship. It also involved collaborations with L. Gueguen and V. Daubin from the Laboratoire de Biométrie et Biologie Evolutive (UMR CNRS 5558, Lyon), with Nadia El-Mabrouk from the Département d'Informatique et de Recherche Opérationnelle in Montréal (Canada), with Cédric Chauve from the Department of Mathematics of Simon Fraser University (Burnaby, Canada), and with G. Szollosi from the Biophysics Research Group in Budapest (Hungary).

BIPOP Project-Team

6. New Results

6.1. Multiple impacts modelling

Participants: Bernard Brogliato, Ngoc-Son Nguyen.

The work consists of studying two systems: the rocking block and tapered chains of balls, using the Darboux-Keller model of multiple impacts previously developed. The objectives are threefold: 1) show that the model predicts well the motion by careful comparisons with experimental data found in the literature, 2) study the system's dynamics and extract critical kinetic angles that allow the engineer to predict the system's gross motion, 3) develop numerical code inside the SICONOS platform that incorporates the model of multiple impact. The influence of the kinetic angles in the rocking block motion with friction is analysed as well, numerically. Extensive experimental works have been conducted by our colleague C. Liu at PKU on a disc-ball system. Results are in [30], [24] [64], and in the monograph [15]. Another work is dedicated to analysing the influence of bilateral holonomic constraints on the well-posedness of the complementarity problem obtained from the (frictionless) unilateral constraints. Gauss' principle extension to this case is also analysed [22].

6.2. Discrete-time sliding mode control

Participants: Vincent Acary, Bernard Brogliato, Olivier Huber, Bin Wang.

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. 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 [36]. This work is 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. In particular the high frequency bang-bang controllers which are observed with explicit discretizations, are completely suppressed.

6.3. Dissipativity preserving methods

Participants: Vincent Acary, Bernard Brogliato.

This work concerns the analysis of so-called theta-methods applied to linear complementarity systems that are dissipative (in the sense of Willems). Necessary and sufficient conditions for dissipativity preservation after the time-discretization are derived (preservation of the storage function, the supply rate and the dissipation function). The possible state jumps are also analyzed [57]. It is shown that excepted when the system is state lossless and theta = 0.5, the conditions for dissipativity preservation are very stringent. In this article we also provide (for the first time, to the best of our knowledge) a rigorous definition of numerical dissipation, which remained until now a vague notion in numerical analysis.

6.4. Lur'e set-valued dynamical systems

Participants: Bernard Brogliato, Aneel Tanwani, 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. This study consists in the mathematical analysis (existence and uniqueness of solutions) and the stability analysis (Lyapunov stability, invariance principle) of classes of set-valued Lur'e systems, with applications in complementarity dynamical systems, relay systems, mechanical systems with dry friction, electrical circuits, etc. Our works in this field started in [51]. The results in [53] extend those in [52] with an accurate characterization of the maximal monotonicity of the central operator of these systems, which consists of a projection-like operator. Concrete and verifiable criteria are provided for the above classes (complementarity, relay systems). Results on state observers for classes of Lur'e systems (namely: Moreau's sweeping process of first and second order, and with prox-regular sets) are proposed in [47], [39]. Therein the convexity is replaced by the far more general notion of prox-regularity, which destroys the monotonicity.

6.5. Analysis of Limit Cycles in Piecewise Linear Systems

Participants: Vincent Acary, Bernard Brogliato, Valentina Sessa.

Autonomous piecewise linear systems in the Lur'e form may exhibit periodic steady-state oscillations. For many practical systems belonging to this class the period and the shape of the oscillation is difficult to be predicted a priori. In this work the complementarity approach is used to tackle the issue. The complementarity formalism is used to represent the closed-loop system and a phase condition acting as an anchor equation for the periodic solution. By discretizing the dynamics a mixed complementarity problem is formulated. The corresponding solution provides an accurate prediction of the steady-state oscillation and its period. Numerical results show the effectiveness of the proposed technique for the computation of stable and sliding periodic solutions. The analysis of the steady-state solution of a Colpitts oscillator is considered as an illustration. This work has been presented at CDC 2013 in [37].

6.6. Simulation and stability of piecewise linear gene networks

Participants: Vincent Acary, Arnaud Tonnelier, Bernard Brogliato.

This work has been done in collaboration with the IBIS project team, it is reported in [45], [19]. Gene regulatory networks control the response of living cells to changes in their environment. A class of piecewise-linear (PWL) models, which capture the switch-like interactions between genes by means of step functions, has been found useful for describing the dynamics of gene regulatory networks. The step functions lead to discontinuities in the right-hand side of the differential equations. This has motivated extensions of the PWL models based on differential inclusions and Filippov solutions, whose analysis requires sophisticated numerical tools. We present a method for the numerical analysis of one proposed extension, called Aizerman-Pyatnitskii (AP)-extension, by reformulating the PWL models as a mixed complementarity system (MCS). This allows the application of powerful methods developed for this class of nonsmooth dynamical systems, in particular those implemented in the Siconos platform. We also show that under a set of reasonable biological assumptions, putting constraints on the right-hand side of the PWL models, AP-extensions and classical Filippov (F)-extensions are equivalent. This means that the proposed numerical method is valid for a range of different solution concepts. We illustrate the practical interest of our approach through the numerical analysis of three well-known networks developed in the field of synthetic biology.

In addition, we have investigated oscillatory regimes in repressilator-type models with piecewise linear dynamics [48]. We derived exact analytical conditions for oscillations and showed that the relative location between the dissociation constants of the Hill functions and the ratio of kinetic parameters determines the possibility of oscillatory activities. We also computed analytically the probability of oscillations. Results suggest that a switch-like coupling behaviour, a time-scale separation and a repressilator-type architecture with an even number of elements facilitate the emergence of sustained oscillations in biological systems.

6.7. Numerical analysis and simulation of mechanical systems with constraints

6.7.1. *Event-capturing schemes for nonsmooth mechanical systems*
Participant: Vincent Acary.

To perform the numerical time integration of nonsmooth mechanical systems, the family of event-capturing time–stepping schemes are the most robust and efficient tools. Nevertheless, they suffer from several drawbacks : a) a low–order accuracy (at best at order one), b) a drift phenomena when the unilateral constraints are treated at the velocity level and c) a poor "energetic" behavior in terms of stabilizing the high–frequency dynamics. We proposed self-adapting schemes by applying time–discontinuous Galerkin methods to the measure differential equation in [31]. In order to satisfy in discrete time, the impact law and the constraints at the position and the velocity level, an adaptation of the well-known Gear–Gupta–Leimkuhler approach has been developed in [18]. Finally, the energetic behavior of the standard Moreau–Jean scheme has been addressed in [26] by developing a Newmark–type scheme for nonsmooth dynamics.

6.7.2. *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 α-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. This work has been presented in [35], [40].

6.7.3. *Multibody systems woth contact, friction and clearances*

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 submitted to the ENOC 2014 conference. It is planned to extend these simulations to frictional cases and to mechanisms of circuit breakers.

6.8. Mechanical rods

6.8.1. *High-order models of mechanical rods*

Participants: Florence Bertails-Descoubes, Romain Casati.

Reduced-coordinate models for rods such as the articulated rigid body model or the super-helix model [50] are able to capture the bending and twisting deformations of thin elastic rods while strictly and robustly avoiding stretching deformations. In this work we are exploring new reduced-coordinate models based on a higher-order geometry. Typically, elements are defined by a polynomial curvature function of the arc length, of degree $d \geq 1$. The main difficulty compared to the super-helix model (where $d = 0$) is that the kinematics has no longer a closed form. Last year, in R. Casati's PhD's thesis, we extended this result to the full 3D case. The key idea was to integrate the rod's kinematics using power series expansion, and to design an accurate and efficient computational algorithm adapted to floating point arithmetics. Our method nicely propagates to the computation of the full dynamic of a linked chain of 3d clothoid. This year we thoroughly compared our methods against other rod models from the literature, in terms of both accuracy and computational efficiency. Our results demonstrate that our model is competitive compared to former models, and yields a better trade-off in the case of highly curly rods. All these results were published and presented this year at SIGGRAPH [25]. The source code is also freely distributed under a GPLv.3 license (see Section 5.3).

6.8.2. Inverse modeling of mechanical rods subject to frictional contact

Participants: Florence Bertails-Descoubes, Alexandre Derouet-Jourdan, Gilles Daviet.

Controlling the input shape of slender structures such as rods is desirable in many design applications (such as hairstyling, reverse engineering, etc.), but solving the corresponding inverse problem is not straightforward. In [54], [55] we noted that reduced-coordinates models such as the super-helix are well-suited for static inversion in presence of gravity.

We are facing two main difficulties: 1/ the geometrical fitting of a piecewise helix to an arbitrary input curve and 2/ the inversion a super-helix subject to gravity *and* contacting forces.

6.8.2.1. Geometrical fitting: from an arbitrary smooth curve to a C^1 piecewise helix
Participants: Florence Bertails-Descoubes, Alexandre Derouet-Jourdan.

In A. Derouet-Jourdan's PhD's thesis (co-supervised by Joëlle Thollot, EPI Maverick), we solved this problem by extending to 3d the floating tangents algorithm introduced in 2d in [54]. In this new method, only tangents are strictly interpolated while points are displaced in an optimal way so as to lie in a feasible configuration, *i.e.*, a configuration that is compatible with the interpolation by a helix. Our approach relies upon the co-helicity condition found by Ghosh [56], which was however only partially proved in [56]. To ensure the existence of the helix and prove its uniqueness in the general case, we complete the proof which serves as the basis for our reconstruction algorithm.

Our method proves to be efficient and robust as it can successfully handle large and complex datasets from real curve aquisitions, such as the capture of hair fibers or the magnetic field of a star. We also compared our method against a standard nonlinear least-squares methods. Unlike the optimization approach which often fail to converge in the case of frizzy input curves, our method remains extremely fast regardless the complexity of the input curves. The set of these results was published this year at Computer-Aided Geometric Design [28]. This work has been transferred to L'Oréal in December 2013. Some source code is also freely released for academics under the GPLv.3 license (see Section 5.3).

6.8.2.2. Inverse modeling of a super-helix assembly subject to frictional contact
Participants: Florence Bertails-Descoubes, Alexandre Derouet-Jourdan, Gilles Daviet.

In A. Derouet-Jourdan's PhD's thesis (co-supervised by Joëlle Thollot, EPI Maverick), we bring a first solution to the challenging problem consisting in identifying the intrinsic geometry of a fiber assembly under gravity and (unknown) frictional external and mutual contacts, from a single configuration geometry (a set of geometric curves). Taking an arbitrary fiber assembly geometry (such as hair) as input together with corresponding interacting meshes (such as the body mesh), we interpret the fiber assembly shape as a static equilibrium configuration of a fiber assembly simulator, in the presence of gravity as well as fiber-mesh and fiber-fiber frictional contacts. Assuming fibers parameters are homogeneous and lie in a plausible range of physical values, we show that this large, underdetermined inverse problem can be formulated as a well-posed constrained optimization problem (second-order cone quadratic program), which can be solved robustly and efficiently by leveraging the frictional contact solver of our direct simulator for fiber assemblies [8]. Our method was successfully applied to the animation of various hair geometries, ranging from synthetic hairstyles manually designed by an artist to the most recent human hair data reconstructed from capture. These results were published this year at SIGGRAPH Asia [27].

6.9. Threshold in neural models

Participant: Arnaud Tonnelier.

We studied the threshold for spike initiation in two-dimensionnal neural models. A threshold criterion that depends on both membrane voltage and recovery (or adaptation) variable is proposed. Our approach provides a simple and unified framework that can account for adapting threshold, threshold variability, dynamic threshold, inhibition-induced spike and postinhibitory facilitation. Implications on neural modeling and on neural dynamics are discussed.

6.10. Nonsmooth modes in chains of impact oscillators

Participants: Vincent Acary, Guillaume James, Franck Pérignon.

Chains of impact oscillators arise for example as finite-element models of thin oscillating mechanical structures (a string under tension or a clamped beam) contacting rigid obstacles. Nonlinear periodic waves are observed in experiments on such systems, but relatively little is known from a theoretical point of view on their existence and stability. In 2008, Gendelman and Manevitch have analyzed the existence and stability of nonlinear localized modes (breathers) for discrete linear chains with a single node undergoing rigid impacts. In this work, we introduce a numerical method allowing to compute branches of time-periodic solutions when an arbitrary number of nodes undergo rigid impacts without energy dissipation. For this purpose, we reformulate the search of periodic solutions as a boundary value problem incorporating unilateral constraints. We illustrate this numerical approach by computing different families of breathers and nonlinear normal modes. Our method is much more effective than a numerical continuation of periodic solutions based on compliant models, which requires to integrate stiff differential equations and lead to costly numerical continuation. These results have been submitted to the ENOC 2014 conference.

6.11. Traveling waves in spatially discrete excitable media

Participants: José Eduardo Morales, Arnaud Tonnelier, Guillaume James.

The propagation of traveling waves in excitable media is a widespread phenomenon, with applications ranging from forest fires to electrical signals propagating along nerve fibers. The case of spatially discrete excitable models is notoriously difficult to analyze. In particular, for the discrete FitzHugh-Nagumo reaction-diffusion system, the existence of pulses for a general class of bistable nonlinearities has been proved only recently (Hupkes and Sandstede, 2010). The existence of pulses under more general types of interactions (e.g. elastic instead of diffusive) remains an open question, as well as traveling wave propagation in higher-dimensional systems. These problems will be tackled in the PhD thesis of J.-E. Morales (advisors A. Tonnelier and G. James), which started on November 2013. J.-E. Morales has started to analyze pulse propagation in the excitable Burridge-Knopoff model, which finds applications in the context of nonlinear friction. This model includes elastic interactions between particles, and an additional difficulty linked with nonsmoothness of the (multivalued) Coulomb friction law.

6.12. Nonlinear waves in granular chains

Participants: Guillaume James, Bernard Brogliato, Ngoc-Son Nguyen.

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. When a large number of beads are present, their dynamics can be described by infinite-dimensional differential equations, which possess a limited smoothness when unilateral Hertzian contact interactions are considered. In this context, we have developed and analyzed new reduced-order models describing nonlinear wave propagation in such systems. In the work [49] (collaboration with D.Pelinovsky, McMaster Univ.), we analyze small amplitude slowly modulated compression waves in the limit when the exponent of the Hertz force is close to unity. From a multiple scale analysis, we derive a new type of Korteweg-de Vries equation with logarithmic nonlinearity allowing to approximate wave profiles, in particular solitary wave solutions.

In addition the LZB model introduced in [14] has been extensively used to numerically investigate wave phenomena in chains of aligned balls (tapered, monodisperse, anti-tapered, stepped chains). Thorough comparisons with experimental results reported in the Granular Matter literature have been made. The results are reported in the monograph [15].

6.13. Robotics

6.13.1. *Lexicographic Least-Squares solver*

Participants: Pierre-Brice Wieber, Dimitar Dimitrov.

We have been working on Multi-Objective Least-Squares problems with inequality constraints for the last few years, focusing especially on the Lexicographic case. A previous collaboration with LAAS-CNRS and CEA-LIST led to the development of a software, SOTH , based on Complete Orthogonal Decompositions, which has become a *de facto* reference in robotics when controlling robots (mobile, manipulator or humanoid) through constraints. The focus this year in the Bipop team has been to accelerate computations by reworking the inner matrix decomposition by combining QR and LU decompositions. The resulting solver, called LexLS, is approximately 5 times faster than the previous SOTH solver on most problems. But the main result has been to show both in theory and practice that it is faster to solve a Lexicographic problem than a Weighted problem, on the contrary to popular beliefs both in robotics and optimization theory. That leads to a reversal of popular approaches that prefer to solve weighted problems (thought to be faster to solve) as approximations to lexicographic problems (thought to be slower to solve).

6.13.2. Mobile manipulation by humanoid robots

Participants: Pierre-Brice Wieber, Dimitar Dimitrov, Alexander Sherikov, Jory Lafaye.

The realization of mobile manipulation by humanoid robots requires the handling of two simultaneous problems: taking care of the dynamic balance of the robot, what is usually done with Model Predictive Control (MPC) schemes, and redundant motion and force control of the whole body of the robot, what is usually done with a Quadratic Program, or a more advanced Lexicographic Least-Squares problem (see above). These two problems are usually solved in sequence: an MPC scheme first computes the necessary motion of the feet and Center of Mass (CoM) of the robot, then motion and force redundancy of the whole body of the robot is resolved. We have observed that this sequence corresponds to a lexicographic order between two objectives, feet and CoM motion first, the rest of the body after, which limits the possibility to tackle scenarios where we would like the motion of the CoM of the robot to be driven by the motion of the rest of the body of the robot, for example to catch an object with the hand. We have proposed therefore to reorganize the order between these different objectives, building on the LexLS solver presented above.

6.13.3. Reactive trajectory generation

Participants: Pierre-Brice Wieber, Dimitar Dimitrov, Saed Al Homsy, Matthieu Guilbert.

The goal of the ongoing collaboration with Adept Technologies is to generate near time optimal trajectories in the presence of moving obstacles in real time. Results are not public yet due to industrial constraints.

6.14. Optimization

6.14.1. 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.

We proposed a new family of semidefinite bounds for 0-1 quadratic problems with linear or quadratic constraints [61]. We have embedded the new bounds within branch-and-bound algorithms to solve 2 standard combinatorial optimization problems to optimality.

- *Max-cut.* We developed [60] an improved bounding procedure obtained by reducing two key parameters (the target level of accuracy and the stopping tolerance of the inner Quasi-Newton engine) to zero, and iteratively adding triangle inequality cuts. We also precisely analyzed its theoretical convergence properties. We show that our method outperform the state-of-the-art solver ([62]) on the large test-problems.

- *Heaviest k-subgraph problems.* Adapting the techniques we developed for the max-cut problem, we have proposed in [59] an big improvement of the first algorithm (up to 10 times faster). For the first time, we were 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).

We have also been working on a generic online semidefinite-based solver for binary quadratic problems using the generality of [61]. Finally, a first web interface for our solvers and our data sets are available online at http://lipn.univ-paris13.fr/BiqCrunch/.

6.14.2. On computing marginal prices in electricity production

Participants: Jérôme Malick, Sofia Zaourar.

Unit-commitment optimization problems in electricity production are large-scale, nonconvex and heterogeneous, but they are decomposable by Lagrangian duality. Realistic modeling of technical production constraints makes the dual objective function computed inexactly though. An inexact version of the bundle method has been dedicated to tackle this difficulty [58]. We have worked on two projects related to solving dual unit-commitment problem by inexact bundle methods.

- *Stabilization.* We observed that the computed optimal dual variables show a noisy and unstable behaviour, that could prevent their use as price indicator. We have proposed a simple and controllable way to stabilize the dual optimal solutions, by penalizing the total variation of the prices [63]. Our illustrations on the daily electricity production optimization of EDF show a strinking stabilization at a negligible cost.

- *Acceleration*. We have worked with Welington Oliveira (IMPA, Brazil) on the acceleration of inexact bundle methods by taking advantage of cheap-to-get inexact information on the objective function which comes without any tighness guarantee though. We came up with a new family of bundle methods incorporating this coarse inexact information, to get better iterates. We have studied the convergence of these method and we have conducted numerical experimentation on unit-commitment problems and on two-stage linear problems show a subtantial gain in the overall computing time. This research is about to be released in a preprint in HAL

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

6. New Results

6.1. Parameterized Construction of Program Representations for Sparse Dataflow Analysiss

Participants: André Tavares [UFMG, Belo Horizonte, Brazil], Benoit Boissinot [Ex-Compsys, Google Zurich], Fernando Magno Quintão Pereira [UFMG, Belo Horizonte, Brazil], Fabrice Rastello.

Data-flow analysis usually associates information with control flow regions. Informally, if these regions are too small like a point between two consecutive statements, we call the analysis dense. On the other hand, if these regions include many such points, then we call it sparse. This work presents a systematic method to build program representations that support sparse analyses. To pave the way to this framework, we clarify the literature about well-known intermediate program representations. We show that our approach, subsumes, up to parameter choices, many of these representations, such as the SSA, SSI, and e-SSA forms. In particular, our algorithms are faster, simpler and more frugal than the previous techniques used to construct SSI (static single information) form programs. We produce intermediate representations isomorphic to Choi *et al.*'s sparse evaluation graphs (SEG) for the family of data-flow problems that can be partitioned by variables. However, contrary to SEGs, we can handle - sparsely - problems that are not in this family. We have tested our ideas in the LLVM compiler, comparing different program representations in terms of size and construction time.

This work is part of the collaboration with UFMG (see Section 8.4) and has been accepted for presentation and publication at CC'14 (Compiler Construction Conference) [9].

6.2. A Framework for Enhancing Data Reuse via Associative Reordering

Participants: Kevin Stock [OSU, Columbus, USA], Louis-Noël Pouchet [UCLA, Los Angeles, USA], Fabrice Rastello, J. Ramanujam [LSU, Houston, USA], P. Sadayappan [OSU, Columbus, USA].

The freedom to reorder computations involving associative operators has been widely recognized and exploited in designing parallel algorithms and to a more limited extent in optimizing compilers. However, the use of associative reordering for enhancing data locality has not been previously explored to our knowledge.

In this work, we develop a novel framework for utilizing associativity of operations in regular loop computations to enhance register reuse. Stencils represent a particular class of important computations where our optimization framework can be applied to enhance performance. We use a multi-dimensional retiming formalism to characterize the space of valid transformations and to generate the transformed code. Experimental results demonstrate the effectiveness of the framework.

This work has been submitted to PLDI'14 and is part of the collaboration with P. Sadayappan from the University of Columbus (OSU) (see Section 8.4).

6.3. Function Cloning Revisited

Participants: Matheus Vilela [UFMG, Belo Horizonte, Brazil], Guilherme Balena [UFMG, Belo Horizonte, Brazil], Guilherme Marques [UFMG, Belo Horizonte, Brazil], Fernando Magno Quintão Pereira [UFMG, Belo Horizonte, Brazil], Fabrice Rastello.

Compilers rely on two main techniques to implement optimizations that depend on the calling context of functions: inlining and cloning. Historically, function inlining has seen more widespread use, as it tends to be more effective in practice. Yet, function cloning provides benefits that inline leaves behind. In particular, cloning gives the program developer a way to fight performance bugs, because it generates reusable code. Furthermore, it deals with recursion more naturally. Finally, it might lead to less code expansion, the inlining's nemesis.

In this work, we revisited function cloning under the light of these benefits. We discuss four independent code specialization techniques based on function cloning, which, although simple, find wide applicability, even in highly optimized benchmarks, such as SPEC CPU 2006. We claim that our optimizations are easy to implement and to deploy. We use Wu and Larus's well-known static profiling heuristic to measure the profitability of a clone. This metric gives us a concrete way to point out to program developers potential performance bugs, and gives us a metric to decide if we should keep a clone or not. By implementing our ideas in LLVM, we have been able to speed up some of the SPEC benchmarks by up to 6% on top of the -O2 optimization level.

This work is part of the collaboration with UFMG (see Section 8.4) and was also done in the context of the collaboration with Kalray and the ManycoreLabs project (see Section 7.2).

6.4. Register Allocation and Promotion through Combined Instruction Scheduling, Loop Splitting and Unrolling

Participants: P. Sadayappan [OSU, Columbus, USA], Fabrice Rastello, Lukasz Domanaga.

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, loop fission, 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, loop splitting, 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 a constrained 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 part of the collaboration with OSU (see Section 8.4) and was also done in the context of the collaboration with Kalray and the ManycoreLabs project (see Section 7.2). It contributes to the developments of the Tirex toolbox (see 5.17). It has also been submitted to PLDI'14.

6.5. Beyond Reuse Distance Analysis: Dynamic Analysis for Characterization of Data Locality Potential

Participants: Naznin Fauzia [OSU, Columbus, USA], Venmugil Elango [OSU, Columbus, USA], Mahesh Ravishankar [OSU, Columbus, USA], J. (ram) Ramanujam [LSU, Houston, USA], Fabrice Rastello, Atanas Rountev [OSU, Columbus, USA], Louis-Noël Pouchet [UCLA, Los Angeles, USA], P. Sadayappan [OSU, Columbus, USA].

Emerging computer architectures will feature drastically decreased flops/byte (ratio of peak processing rate to memory bandwidth) as highlighted by recent studies on Exascale architectural trends. Further, flops are getting cheaper while the energy cost of data movement is increasingly dominant. The understanding and characterization of data locality properties of computations is critical in order to guide efforts to enhance data locality.

Reuse distance analysis of memory address traces is a valuable tool to perform data locality characterization of programs. A single reuse distance analysis can be used to estimate the number of cache misses in a fully associative LRU cache of any size, thereby providing estimates on the minimum bandwidth requirements at different levels of the memory hierarchy to avoid being bandwidth bound. However, such an analysis only holds for the particular execution order that produced the trace. It cannot estimate potential improvement in data locality through dependence preserving transformations that change the execution schedule of the operations in the computation.

In this work, we develop a novel dynamic analysis approach to characterize the inherent locality properties of a computation and thereby assess the potential for data locality enhancement via dependence preserving transformations. The execution trace of a code is analyzed to extract a computational directed acyclic graph (CDAG) of the data dependences. The CDAG is then partitioned into convex subsets, and the convex partitioning is used to reorder the operations in the execution trace to enhance data locality. The approach enables us to go beyond reuse distance analysis of a single specific order of execution of the operations of a computation in characterization of its data locality properties. It can serve a valuable role in identifying promising code regions for manual transformation, as well as assessing the effectiveness of compiler transformations for data locality enhancement. We demonstrate the effectiveness of the approach using a number of benchmarks, including case studies where the potential shown by the analysis is exploited to achieve lower data movement costs and better performance.

This work is part of the collaboration with OSU (see Section 8.4) and has been accepted for publication at ACM TACO [2].

6.6. Characterizing the Inherent Data Movement Complexity of Computations via Lower Bounds

Participants: P. Sadayappan [OSU, Columbus, USA], Venmugil Elango [OSU, Columbus, USA], J. (ram) Ramanujam [LSU, Houston, USA], Louis-Noël Pouchet [UCLA, Los Angeles, USA], Fabrice Rastello.

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. Although the problem of characterizing data access complexity has been modeled previously using the formalism of Hong & Kung's red/blue pebble game [27], applicability of previously-developed approaches has been extremely limited. We improve on prior work in several ways: 1) we develop an approach to composing lower bounds from arbitrary decompositions of computational directed acyclic graphs, thereby eliminating a significant limitation of previous approaches that required homogeneity of analyzed computations, 2) we develop a complementary graph min-cut based strategy to Hong & Kung's S-partitioning approach, and 3) we develop an automated approach to generate concrete I/O lower bounds of arbitrary, possibly irregular computational directed acyclic graphs. We provide experimental results demonstrating the utility of the developed approach.

This work has been submitted to PLDI'14 and is part of an informal collaboration with P. Sadayappan from the University of Columbus (OSU) (see Section 8.4).

6.7. Enhancing the Compilation of Synchronous Dataflow Programs

Participants: Paul Feautrier, Abdoulaye Gamatié [LIRMM, Montpellier], Laure Gonnord.

In this work [12], which is an extension of [26], we propose an enhancement of the compilation of synchronous programs with a combined numerical-Boolean abstraction. While our approach applies to synchronous dataflow languages in general, here, we consider the SIGNAL language for illustration. In the new abstraction, every signal in a program is associated with a pair of the form (`clock`, `value`), where `clock` is a Boolean function and `value` is a Boolean or numeric function. Given the performance level reached by recent progress in satisfiability modulo theory (SMT), we use an SMT solver to reason on this abstraction. Through sample examples, we show how our solution is used to determine absence of reaction captured by empty clocks; mutual exclusion captured by two or more clocks whose associated signals never occur at the same time; or hierarchical control of component activations via clock inclusion. We also show that the analysis improves the quality of the code generated automatically by a compiler, e.g., a code with smaller footprint, or a code executed more efficiently thanks to optimizations enabled by the new abstraction. The implementation of the whole approach includes a translator of synchronous programs towards the standard input format of SMT solvers, and an ad hoc SMT solver that integrates advanced functionalities to cope with the issues of interest in this work. These results have been published in 2013 (but considered as published in 2012) in the CSI Journal of Computing [24].

6.8. Synthesis of Ranking Functions using Extremal Counter-Examples

Participants: David Monniaux [Verimag, Grenoble], Lucas Séguinot [Student at ENS Cachan Bretagne], Laure Gonnord.

In [14], we presented a new algorithm adapted from scheduling techniques to synthesize (multi-dimensional) affine functions from general flowcharts programs. But, as for other methods, our algorithm tried to solve linear constraints on each control point and each transition, which can lead to quasi-untractable linear programming instances.

In contrast to these approaches, we proposed a new algorithm based on the following observations:

- Searching for ranking functions for loop headers is sufficient to prove termination.
- Furthermore, there exist loops such that there is a linear lexicographic ranking function that decreases along each path inside the loop, from one loop iteration to the next, but such that there is no lexicographic linear ranking function that decreases at each step along these paths. For these reasons, it is tempting to treat each path inside a loop as a single transition.

Unfortunately the number of paths may be exponential in the size of the program, thus the constraint system may become very large, even though it features fewer variables. To face this theoretical complexity, even though the number of paths may be large, we argue that, in practice, few of them actually matter in the constraint system (we formalize this concept by giving a characterization as geometric extremal points). Our algorithm therefore builds the constraint system lazily, taking paths into account *on demand*.

We are currently testing our preliminary implementation and submitting a paper on these new results.

6.9. Data-Aware Process Networks

Participants: Christophe Alias, Alexandru Plesco.

The following results concern the applied research activities directly linked to the Zettice start-up (see Section 7.3), which aims at applying polyhedral techniques to high-level circuit synthesis (HLS). Following the guidelines of Inria DTI, as this research aims to be transferred, these results are not published before being "protected" or exploited. An Inria patent deposit is currently processed.

- **Data-aware process networks (DPN).** This is the intermediate representation of the HLS flow. DPN is a parallel execution model fitting the hardware constraints of circuit synthesis, in which the data transfer and the synchronizations are made explicit. We formally described the DPN model and a translation scheme from C programs, and we showed the consistency in the meaning where any terminating sequential program is translated to an equivalent DPN, guaranteed to be deadlock free.
- **Front-end analysis.** We designed many program analyses to produce a quality DPN from a C program:
  - *Throughput optimization.* A I/O scheme has been designed, with the corresponding compiler analysis, to minimize the I/O traffic with the external memory. This allows us to balance efficiently the spilling of temporary value to the memory, and the local buffer size. This scheme impacts the DPN structure itself.
  - *Communication vectorization.* The matrix structure of the memory allows us to load data by chunks. A polyhedral analysis has been designed to solve this issue.
  - *Synchronization scheme.* As parallel units need to communicate intermediate results, synchronizations must be ensured. Unlike KPN, DPN do not use FIFO, but buffers, which required an efficient synchronization mechanism.
- **Back-end analysis.** Once generated, a DPN must be mapped to an FPGA. This raises many interesting issues:
  - *Pipeline completion.* Data paths make an extensive use of pipelined operators, which delays the signal. An algorithm has been designed to enforce the time coherence of signals.
  - *Polyhedral units.* DPNs make an extensive use of piece-wise affine functions, which must be mapped properly to ensure the efficiency of the whole system. A preliminary algorithm has been designed to reach a correct trade-off between critical path size and LUT usage.

All these analyses have been fully implemented. The tool Dcc (DPN C Compiler) implements all the front-end analyses. The tool IceGEN implements the back-end analysis.

6.10. Program Equivalence Modulo A/C (Associativity/Commutativity)

Participants: Guillaume Iooss [PhD student], Christophe Alias, Sanjay Rajopadhye [Colorado State University].

Program equivalence is a well-known problem with a wide range of applications, such as algorithm recognition, program verification, and program optimization. This problem is also known to be undecidable if the class of programs is rich enough, in which case semi-algorithms are commonly used. We focus on programs represented as a system of affine recurrence equations (SARE), defined over parametric polyhedral domains, a well-known formalism for the *polyhedral model*, which includes as a proper subset, the class of affine control loop programs. Several semi-algorithms for program equivalence have already been proposed for this class. A few of them take into account algebraic properties such as associativity and commutativity. However, to the best of our knowledge, none of them is able to manage reductions, i.e., accumulations of a parametric number of sub-expressions using an associative and commutative operator.

Our contributions are:

- An equivalence checking algorithm able to manage associativity and commutativity properties. Our method subsumes the previous approaches and is, to the best of our knowledge, the first one able to manage these properties over a parametric number of expressions.
- A semi-algorithm to construct a perfect matching problem on a parametric bipartite graph. We partially solve this problem through a heuristic based on the augmenting path algorithm. This heuristic is able to find a set of non-interfering augmenting paths to improve a proposed maximum matching, as long as these augmenting paths do not have a parametric length.

A preliminary implementation is under development. This work has been submitted to ESOP'14.

6.11. Constant Aspect-Ratio Parametric Tiling

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

Parametric tiling is a well-known transformation that is widely used to improve locality, parallelism, and granularity. However, parametric tiling is also a non-linear transformation and this prevents polyhedral analysis or further polyhedral transformation after parametric tiling. It is therefore generally applied during the code generation phase.

This result consists on a method to stay polyhedral in a special case of parametric tiling, where all the dimensions are tiled and all the tile sizes are constant multiples of a single tile size parameter. We call this *Constant Aspect Ratio Tiling*. We show how to mathematically transform a polyhedron and an affine function into their tiled counterpart and show how to obtain good generated code.

This work has been accepted for publication at IMPACT'14 [8].

6.12. Parametric Tiling with Inter-Tile Data Reuse

Participants: Alain Darte, Alexandre Isoard.

Loop tiling is a loop transformation widely used to improve spatial and temporal data locality, increase computation granularity, and enable blocking algorithms, which are particularly useful when offloading kernels on platforms with small memories. When hardware caches are not available, data transfers must be software-managed: they can be reduced by exploiting data reuse between tiles and, this way, avoid some useless external communications. An important parameter of loop tiling is the sizes of the tiles, which impact the size of the necessary local memory. However, for most analyzes that involve 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. In this work, we showed that, actually, parametric tiling with inter-tile data reuse is nevertheless possible.

Our solution is the first parametric solution for generating the memory transfers needed when a kernel is offloaded to a distant accelerator, tile by tile after loop tiling, and when all intermediate results are stored locally on the accelerator. For such computations, there is a complete decoupling between loads and stores, and when a value has been defined in a previous tile, it has to be loaded from the local memory and not from the distant memory as this memory is not yet up-to-date. In other words, inter-tile reuse is mandatory. This also saves external communications. Our solution is parametric in the sense that we derive the set of loads and stores from and to the distant memory with the tile sizes as parameters. Although the direct formulation is quadratic, we can still solve it in an affine way by developing techniques that consider, in the analysis, all (unaligned) possible tiles obtained by translation and not just those that belong to a tiling (partitioning) of the iteration space. We were able to use a similar technique to also parameterize the computations of local memory sizes, thanks to parametric lifetime analysis and folding with modulos, even for pipeline schedules similar to double buffering. Our method is currently implemented with the `iscc` calculator of ISL, a library for the manipulation of integer sets defined with Presburger arithmetic.

Also, the whole analysis can handle approximations thanks to the introduction of the concept of pointwise functions, well suited to deal with unaligned tiles. We believe that this 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.

This work has been accepted for publication at IMPACT'14 [5].

6.13. Data Races in the Parallel Language X10

Participants: Tomofumi Yuki [Colorado State University and Inria/IRISA], Paul Feautrier, Sanjay Rajopadhye [Colorado State University], Vijay Saraswat [IBM Research].

Parallel programmers are now required to efficiently utilize the massive amount of parallelism provided by multi-core and many-core systems. Parallel programming is difficult, and the existing tools are mostly low-level extensions to sequential languages or libraries. As an effort to improve this situation, several groups have initiated the design of parallel programming languages, mostly based on the partitioned global address space (PGAS) paradigm. One of these languages is X10, which is developed at IBM Research by a team led by Vijay Saraswat.

While such languages hide the low-level details of parallel programming, they cannot guarantee that the object code will be correct by construction. Parallelism introduces two new types of bugs: non-determinism and deadlocks, and experience shows that it is possible to guarantee the absence of one type but not both. X10 programs are guaranteed deadlock-free but may have non-determinism. Non-determinism can be detected at runtime, but this approach cannot give absolute guarantees. However, it is possible, at least for a restricted class of X10 programs, to check for non-determinism at compile time.

The first step in this direction is to define the *polyhedral fragment* of X10, in which the only control constructs are `for` loops with affine bounds, and the only data structures are arrays with affine subscripts. X10 has many parallel constructs: as a first effort, we focused on `async`, which creates an activity (lightweight thread) and `finish`, which waits for termination of all impending activities. The execution order (or *happens-before relation*) of such a program is an incomplete lexicographic order, in which terms relating operations in different activities are removed. The dataflow analysis method of [23] has to be adapted to a partial execution order, which may have many extrema instead of a unique maximum. Multiple extrema denote data races, thus non-determinism. A detector along these lines has been implemented and presented at PPoPP'13 (Symposium on Principles and Practice of Parallel Programming) [10].

X10 other parallel programming primitive directives are *clocks* and `atomic`. The at construct allows downloading a computation to another *place*. Clocks are a dynamic version of barriers. Their analysis involves counting their instances. For polyhedral programs, this can be done using the Ehrhart and Barvinok theories; the results are polynomials. Checking whether clocks remove non-determinism involves finding integer

roots and hence is undecidable. However, modern SMT solvers are able to solve most of these problems. The resulting paper [13] has been submitted to the ECOOP conference.

6.14. Clock Removal in X10

Participants: Paul Feautrier, Eric Violard [Inria/Camus], Alain Ketterlin [Inria/Camus].

In the light of the previous work on the determinism of X10, a natural question is: are the parallel programming directives of X10 redundant? The answer is yes, at least for static control programs, i.e., programs in which the set of operations and their execution order do not depend on the input data. The basic idea is that the synchronization which occurs when several activities execute an advance is similar to the synchronization at the end of a finish. If one is able to count advances, one may construct a front by gathering all operations with the same advance count. Each front is executed inside one finish, and fronts are executed sequentially in order of increasing counts. For polyhedral programs, advance counting can be done at compile time. If the counts are affine functions, the restructuring can be done by classical polyhedral code generators like CLooG, and no overhead is incurred. For polynomial counts, one overall enclosing loop must be added, but the resulting program can usually be optimized by simple loop transformations, e.g., pushing guards into enclosing loop bounds. For arbitrary programs, the counts have to be computed dynamically; this is possible only if the program has static control.

This result does not contradict the previous undecidability proof (Section 6.13), as the translation of a polyhedral program is usually not polyhedral. Application of the method to a set of simple kernels has shown significant speedups. The interpretation of this result is that, at least in the present state of the X10 runtime, the implementation of the `async` primitive is more mature than the implementation of clocks. A paper on this topic has been accepted at CC'14 (Compiler Construction Conference) [7].

6.15. Static Analysis of OpenStream Programs

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

The objective of the collaboration between the Compsys and Parkas teams in the ManycoreLabs project (Section 7.2) is to evaluate the possibility of applying polyhedral techniques to the parallel language OpenStream, which is developed by Inria Parkas. When applicable, these techniques are invaluable for compile-time debugging and for improving the target code for a better adaptation to the target architecture.

OpenStream is a two-level language, in which a sequential control code directs the initialization of parallel task instances that communicate through *streams*. OpenStream programs are deterministic by construction, but may have deadlocks. If the control code is polyhedral, one may statically compute, for each task instance, its read and write indices for each stream. These indices may be polynomials of arbitrary degree. When linear, the full power of the polyhedral model may be brought to bear for dependence and dataflow analysis, scheduling and deadlock detection, and program transformations.

In the general case, one can think of two approaches: the first one consists in over-approximating dependences until problems become linear. In the second approach, one first leverages modern developments in SMT solvers, which allow them to solve polynomial problems, albeit with no guarantee of success. Furthermore, the task index functions have special properties that may be used to construct original analysis algorithms. Three preliminary results in this direction:

- the proof that deadlock detection is undecidable in general, thanks to an adaptation of the proof designed for X10 (Section 6.13),
- a characterization of deadlocks in terms of dependence graphs, which implies that streams can be safely bounded as soon as a schedule exists with such sizes,
- a preliminary analysis of some solvable cases.

A document is available as Deliverable 2.5.3 for the ManycoreLabs project.

6.16. Array Contraction in Parallel Programs

Participants: Alain Darte, Alexandre Isoard.

Array contraction is a technique to reuse array elements when they are dead, in a form of array folding. A standard technique for array contraction is to use affine remappings with modulos. When the modulo is equal to 1, this corresponds to the removal of the corresponding array dimension. Array contraction is well-known for sequential programs, after element-wise array liveness analysis. It has also been customized for parallel codes obtained through affine schedules by Lefebvre and Feautrier, and Quilleré-Rajopadhye, both frameworks being generalized by the lattice-based memory allocation framework of Darte, Schreiber, and Villard [17] and the construction of the set of conflicting array indices. We showed how the same framework can be used for a larger range of parallel programs, including programs with outer parallel loops, programs exhibiting pipelining, a subset of X10, etc. The optimality of the construction can be shown, despite a related (but actually non-contradictory here) NP-completeness result for worst-case of register pressure in the context of register allocation. A research report on this topic is in preparation.

<p style="text-align:center">**CONVECS Project-Team**</p>

6. New Results

6.1. New Formal Languages and their Implementations

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.

6.1.1. Translation from LNT to LOTOS

Participants: Hubert Garavel, Frédéric Lang, Wendelin Serwe.

The LNT2LOTOS, LNT.OPEN, and LPP tools convert LNT code to LOTOS, thus allowing the use of CADP to verify LNT descriptions. These tools have been used successfully for many different systems (see § 6.5 and § 9.1).

In 2013, in addition to 15 bug fixes, the following enhancements have been brought to these tools:

- The list of predefined functions that can be generated automatically for list and set types has been enriched, so as to include all operations commonly found in programming languages.

- A new "sorted set" type was added to LNT, where the automatically generated insertion function preserves the invariant that all elements in the set are sorted in ascending order and have a single occurrence.

- The use of range and predicate types has been facilitated, by translating explicit type annotations by type conversions if necessary.

- An implicit type conversion is applied by assignments to a variable; this helps the type-checker to solve overloaded definitions.

- The generated LOTOS and C code has been modified to avoid spurious warning messages from the LOTOS and C compilers.

- The demo examples demo_19, demo_30, and demo_35 have been enhanced with an LNT version.

6.1.2. 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 2013, in addition to fixing four bugs, the type checking algorithm of CAESAR and CAESAR.ADT was entirely revised to display less and better messages in case of typing errors, avoiding cascading error messages, e.g., when an undefined variable or constant is used, or when an overloaded function is improperly used in a context where a unique type is expected.

Also, the CAESAR compiler found a new usefulness as a means to easily produce large-size, realistic Petri nets that can be used as benchmarks by the Petri net community. To make this possible, a new option was added to CAESAR to export the BPN (*Basic Petri Net*) file generated from a LOTOS specification. The definition of the BPN format was made more precise by adding semantic constraints. The CAESAR.BDD tool of CADP was enhanced with two new options, one that checks whether a BPN file satisfies all semantic constraints, and another one that converts a BPN file into PNML (*Petri Net Markup Language*) format.

This work has been done in coordination with Fabrice Kordon and Lom-Messan Hillah (UPMC/LIP6, Paris, France) for the MCC (*Model Checking Contest*) workshop [5]. H. Garavel was in charge of redesigning the model forms used for this contest. One Petri net generated using CAESAR was selected as a benchmark for MCC'2013 and five Petri nets generated using CAESAR have been submitted to MCC'2014.

6.1.3. *Translation from an Applied Pi-Calculus to LNT*

Participants: Radu Mateescu, Gwen Salaün.

The π-calculus is a process algebra defined by Milner, Parrow, and Walker two decades ago for describing concurrent mobile processes. Despite a substantial body of theoretical work in this area, only a few verification tools have been designed for analysing π-calculus specifications automatically. Our objective is to provide analysis features for the π-calculus by reusing the verification technology available for process algebras without mobility. For this purpose, we extended the original polyadic π-calculus with the data types and functions of LNT. This yields a general-purpose applied π-calculus, which is suitable for specifying mobile value-passing concurrent systems belonging to various application domains. Our approach is based on a novel translation from the finite control fragment of π-calculus to LNT, making possible the analysis of applied π-calculus specifications using all verification tools of CADP. This translation is fully automated by the PIC2LNT translator (see § 5.3).

In 2013, we continued our work on the applied π-calculus and its translation to LNT. This resulted in a new version PIC2LNT 3.0 of the tool, which fixes several bugs and brings the following improvements:

- A bounded replication operator was added to the language, which expresses the parallel execution of a fixed number of π-calculus agents. This operator is translated into LNT by instantiating the appropriate number of corresponding processes.

- A type Chan representing channel names was implemented, which can be freely combined with ordinary data types. This increases the versatility of the language by allowing, e.g., the definition of agents parametrized by sets of channel names.

- Several options were added to the tool for enhancing its ergonomy and tuning the state space generation (specify the set of private channels that can be created, generate the state space of a particular agent).

A paper describing this work has been published in an international conference [16].

6.1.4. *Translation from EB3 to LNT*

Participants: Frédéric Lang, Radu Mateescu.

In collaboration with Dimitris Vekris (University Paris-Est Créteil), we considered a translation from the EB3 language [39] for information systems to LNT. EB3 has a process algebraic flavor, but has the particularity to contain so-called *attribute functions*, whose semantics depend on the history of events. We have proposed a formal translation scheme, which ensures the strong equivalence between the LTSs corresponding to an EB3 specification and to the LNT code generated. A prototype translator has been developed at University Paris-Est Créteil, which enables EB3 specifications to be formally verified using CADP.

In 2013, a paper has been published in an international conference [19].

6.1.5. *Coverage Analysis for LNT*

Participants: Gwen Salaün, Lina Ye.

In the classic verification setting, the designer has a specification of a system in a value-passing process algebra, a set of temporal properties to be verified on the corresponding LTS model, and a data set of examples (test cases) for validation purposes. At this stage, building the set of validation examples and debugging the specification is a complicated task, in particular for non-experts.

---

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

In 2013, we proposed a new framework for debugging value-passing process algebraic specifications by means of coverage analysis and we illustrated our approach with LNT. We define several coverage notions before showing how to instrument the specification without affecting its original behavior. Our approach helps the specifier to find dead code, ill-formed conditional structures, and other errors in the specification, but also to improve the quality of a data set of examples used for validation purposes. We have implemented a prototype tool, named CAL, for automating the verification of coverage analysis, and we applied it to several real-world case studies in different application areas. A paper has been submitted to an international conference.

6.1.6. *Other Compiler Developments*

Participants: Soraya Arias, Hubert Garavel, Frédéric Lang, Wendelin Serwe.

- In co-operation with Jérôme Hugues (ISAE, Toulouse), we investigated the translation of AADL (*Architecture Analysis and Design Language*) into LNT. An AADL example was manually tackled, leading to the conclusion that LNT could be a suitable target language for translating a large fragment of AADL.

  In co-operation with Holger Hermanns (Saarland University, Germany) and Joost-Pieter Katoen (RWTH Aachen, Germany), we prepared a contribution for the AADL standardization committee to detail semantics issues of the GSPN (*Generalized Stochastic Petri Nets*) model.

- We continued our work on the FLAC tool, which translates the Fiacre intermediate language into LOTOS to enable verification using CADP. In 2013, we eliminated spurious compilation warnings, we removed the definitions of integer operations *div* and *mod*, which have been added to a standard LOTOS library, and we improved the encoding of integer numbers. These changes have led to revisions 76 to 79 of the FLAC code, which is available on the development forge dedicated to Fiacre compilers [6].

- In co-operation with Holger Hermanns, we started studying the PseuCo language that is being defined and implemented at Saarland University. Developed from an educational perspective as a means to teach concurrency theory to bachelor students, PseuCo combines features from Java and Go, the language promoted by Google for concurrent programming. PseuCo supports both message-passing and shared-memory concurrency in a way that is easy to use and that can readily be transferred to Java, Go, or other mainstream languages. PseuCo has been awarded with the 2013 German national "*Preis des Fakultätentages Informatik*" for its innovative role in undergraduate education.

  In 2013, we undertook the manual translation of various PseuCo sample programs into LNT and started enhancing LNT with features that would enable automated PseuCo-to-LNT translation. We also reviewed a PseuCo-to-CCS translator recently developed at Saarland University and wrote an evaluation report for this software.

6.2. Parallel and Distributed Verification

6.2.1. *Manipulation of Partitioned LTSs*

Participants: Hubert Garavel, Radu Mateescu, Wendelin Serwe.

For distributed verification, CADP provides the PBG format, which implements the theoretical concept of *Partitioned LTS* [45] and provides a unified access to an LTS distributed over a set of remote machines.

In 2013, we continued the development of the prototype tool PBG_OPEN, which is an OPEN/CAESAR-compliant compiler for the PBG format, enabling the use of all CADP on-the-fly verification tools on a partitioned LTS. The main advantage of PBG_OPEN is that it can use the memory of several machines to store the transition relation of a partitioned LTS. Therefore, PBG_OPEN can explore on-the-fly large partitioned LTSs that could not be explored using other tool combinations. To reduce the amount of communications, PBG_OPEN can use a cache to store already encountered states, together with their outgoing transitions.

---

[6]http://gforge.enseeiht.fr/projects/fiacre-compil

We also developed another prototype tool, named PBG_INVERT, which changes the storage of the transitions of a partitioned LTS, transforming a partitioned LTS where each fragment stores the transitions leading to the states of the fragment (as generated by DISTRIBUTOR) into a partitioned LTS where each fragment stores the transitions going out from the states of the fragment. Adding this transformation step yields a reduction of up to 25% of the overall execution time, when verifying the partitioned LTS with PBG_OPEN. We experimented all these tools on the Grid'5000 computing infrastructure [31] using up to 512 distributed processes. These experiments confirmed the good scalability of our distributed LTS manipulation approach. A paper describing this work has been published in an international conference [13].

6.2.2. *Distributed Code Generation for LNT*

Participants: Hugues Evrard, Frédéric Lang.

Rigorous development and prototyping of a distributed verification algorithm in 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.

In 2013, we formally modelled and verified several existing synchronization protocols. This revealed an error in one of them, which led to a publication in an international conference [12]. Based on this study, we selected a suitable protocol and adapted it to the LNT synchronization operators.

Using this protocol, we developed a prototype distributed code generator, taking as input the model of a distributed system, described as a set of LNT processes and their parallel composition written in EXP. The LNT.OPEN and CAESAR tools are used to obtain the sequential implementation of each LNT process, and the EXP.OPEN tool is used to compute the possible interactions between processes. Then, our prototype generates the corresponding implementation of the distributed synchronization protocol and all necessary glue code between processes and the protocol. Our prototype automatically performs all these steps, such that a complete and runnable distributed implementation can easily be obtained from the original model.

So far, our prototype manages synchronizations with no data or data of enumerated types only, in which case the implementation checks that data values and types match before allowing a synchronization.

6.3. Timed, Probabilistic, and Stochastic Extensions

Participants: Hubert Garavel, Frédéric Lang, Radu Mateescu.

Process calculi provide a suitable formal framework for describing and analyzing concurrent systems, but need to be extended to model refined aspects of these systems. For instance, it may be necessary to represent probabilistic choices (in addition to deterministic and nondeterministic choices) as well as delays and latencies governed by probability laws. Many such extensions have been proposed in the literature, some of which have been implemented in software tools and applied to nontrivial problems. In particular, two of these extensions (namely, *Interactive Markov Chains* and *Interactive Probabilistic Chains*) are implemented in CADP. Despite these achievements, the state of the art is not satisfactory as the extended languages primarily focus on the probabilistic and stochastic aspects, leaving away the expressive and user-friendly features that process calculi provide for describing conventional concurrent systems.

In 2013, we did the following steps to progress our agenda of bridging the gap between functional verification and quantitative evaluation:

- We equipped CADP with a new tool named BCG_CMP, which enables to compare quantitative models modulo probabilistic and stochastic variants of strong bisimulation and branching bisimulation. Such comparison relations were not available in the BISIMULATOR tool that already existed in CADP.

- We investigated the feasibility of creating interconnections between mainstream verification tools for probabilistic and stochastic systems. In a first step, we focused on the DTMC (*Discrete-Time Markov Chain*) model and on three mainstream tools: CADP (Grenoble), MRMC (Aachen), and PRISM (Birmingham-Oxford).

  We developed translation tools to perform conversions between the various formats of these tools (".aut" and ".bcg" for CADP, ".tra/.sta/.lab" for MRMC, ".pm" and ".tra/.sta/.lab" for PRISM). So doing, we reported one bug in MRMC and five minor issues in PRISM. By discussing with Dave Parker (University of Birmingham), we contributed to the introduction in PRISM 4.1 of two new options "-importmodel" and "-exportmodel" that greatly simplify exchanges of models between PRISM and other tools.

  We developed a generator of random DTMCs in CADP, MRMC, and PRISM formats, and undertook the construction of a collection of DTMCs, which we used to compare the performance and scalability of CADP and PRISM.

- We started to investigate the evaluation of temporal logic properties on extended DTMCs, in which transitions are labeled with probabilities and optional actions. For this purpose, we developed a new prototype XTL library (consisting of XTL and C code) encoding the PCTL (*Probabilistic CTL*) temporal logic [50]. This new PCTL library enables the specifier to combine data-based, discrete-time, and probabilistic properties of DTMCs in a uniform way.

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

6.4.1. *Compositional Model Checking*

Participants: Frédéric Lang, Radu Mateescu.

We have continued our work on partial model checking following the approach proposed in [26]. Given a temporal logic formula φ to be evaluated on a set S of concurrent processes, partial model checking consists in transforming φ into another equivalent formula φ' to be evaluated on a subset of S. Formula φ' is constructed incrementally by choosing one process P in S and incorporating into φ the behavioral information corresponding to P — an operation called quotienting. Simplifications must be applied at each step, so as to maintain formulas at a tractable size.

In 2013, we extended the approach to handle fairness operators of alternation depth two, and we conducted new experiments. This resulted in a new version of the PMC prototype tool (see § 5.4) supporting all features of the input language of EXP.OPEN 2.1. An article has been published in an international journal [5].

6.4.2. *On-the-Fly Test Generation*

Participants: Radu Mateescu, Wendelin Serwe.

In the context of the collaboration with STMicroelectronics (see § 6.5.1 and § 7.1), we studied techniques for testing if an implementation is conform to a formal model written in LNT. Our approach is inspired by the theory of conformance testing [68], as implemented for instance in TGV [53] and JTorX [30].

We developed two prototype tools. The first tool implements a dedicated OPEN/CAESAR-compliant compiler for the particular asymmetric synchronous product of the model and the test purpose, and uses slightly extended generic components for graph manipulation (τ-compression, τ-confluence reduction, determinization) and resolution of Boolean equation systems. The second tool generates the complete test graph, which can be used to extract concrete test cases or to drive the test of the implementation. The principal advantage of our approach compared to existing tools is the use of LNT for test purposes, facilitating the manipulation of data values.

In 2013, we continued the development of these tools, with a focus on reducing execution time. We also implemented a prototype tool to extract from a complete test graph one or all test cases of minimal depth. We experimented with these tools on two case-studies, namely the ACE coherence protocol (see § 6.5.1) and the EnergyBus (see § 6.5.5).

6.4.3. Equivalence Checking

Participant: Frédéric Lang.

Equivalence relations can be used for verification in two complementary ways: for the minimization of an LTS and the comparison of two LTSs.

In 2013, we worked along the following lines:

- We added observational equivalence (following a request from LAAS-CNRS) as well as divergence-sensitive branching bisimulation (together with its stochastic and probabilistic variants) in BCG_MIN.
- We improved the speed of BCG_MIN in the case of branching reduction applied to a graph with a high branching factor and many internal transitions, by correcting a function that has a quadratic complexity instead of a linear one.
- We added the new tool BCG_CMP, which takes as input two BCG graphs and checks whether they are equivalent modulo a relation chosen among strong and branching bisimulation (and their stochastic and probabilistic variants), divergence-sensitive branching bisimulation, or observational equivalence. BCG_CMP checks equivalence using the partition-refinement algorithm of BCG_MIN. We compared BCG_CMP and BISIMULATOR on the VLTS benchmark suite [7], showing that BCG_CMP is generally slightly less efficient than BISIMULATOR for comparisons yielding a FALSE result, but much more efficient than BISIMULATOR for comparisons yielding a TRUE result.
- The new tool BCG_CMP as well as the new equivalence relations added to BCG_MIN have been added to the EUCALYPTUS graphical user interface and to the SVL scripting language.

6.4.4. Other Software Developments

The OPEN/CAESAR environment was enhanced with a new generic library (named CAESAR_CACHE_1) for manipulating hierarchical caches, with 15 built-in replacement strategies and the possibility to define new ones.

We also maintained the CADP toolbox, taking into account the feedback received from numerous users in the world. In addition to fixing 41 bugs, we evolved CADP to support the latest versions of Windows, Cygwin, Mac OS X, and their corresponding C compilers. The documentation for installing CADP has been updated and shortened. Finally, support for Sparc, Itanium, and PowerPC processors was dropped at the end of 2013 based on the observation that these architectures are almost no longer used among the CADP user community.

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 [25]. In a first step, we developed a formal LNT model (about 3200 lines of LNT) of a system consisting of an ACE-compliant cache coherent interconnect, processors, and a main memory. The model is parametric and can be instantiated with different configurations (number of processors, number of cache lines, number of memory lines) and different sets of supported elementary ACE operations (currently, a representative subset of 15 operations), including an abstract operation that represents any other ACE operation. We handled the global requirements of the ACE specification using a constraint oriented programming style, i.e., by representing each global requirement as a dedicated process observing the global behaviour and inhibiting incorrect executions.

---

[7]http://cadp.inria.fr/resources/vlts

In a second step, we generated for several configurations the corresponding LTS (up to 100 million states and 350 million transitions). We wrote two liveness properties in MCL expressing that each read (respectively write) transaction is executed until its termination. We also wrote two properties expressing cache coherence and data integrity. This required to transform state-based properties into action-based properties, by adding information about the cache state to actions executed by the cache. For all considered configurations, we checked these properties using parametric SVL scripts (about 100 lines) and EVALUATOR. For some scenarios without the processes representing the global requirements, EVALUATOR generated counterexamples for the cache coherence and data integrity. We are currently using these counterexamples to derive test cases for the architecture under design at STMicroelectronics.

This work led to publications [21], [15].

6.5.2. *Choreography-based Communicating Systems*

Participants: Radu Mateescu, Gwen Salaün, Lina Ye, Kaoutar Hafdi.

Choreographies are contracts specifying interactions among a set of services from a global point of view. These contracts serve as reference for the further development steps of the distributed system. Therefore, their specification and analysis is crucial to avoid issues (e.g., deadlocks) that may induce delays and additional costs if identified lately in the design and development process.

In 2013, we have obtained the following results:

- In collaboration with Meriem Ouederni (University of Toulouse) and Tevfik Bultan (University of California at Santa Barbara), we have proposed a branching definition of the synchronizability property, which identifies systems whose interaction behavior remains the same when asynchronous communication is replaced with synchronous communication. We have also shown how these results can be used for checking the compatibility of a set of asynchronously communicating components [17].

- In collaboration with Matthias Güdemann (Systerel), we have defined sufficient conditions for checking the repairability property, which indicates whether realizability can be enforced for choreography-based communicating systems using distributed controllers. A paper has been submitted to an international conference.

- We have proposed an approach for computing the degree of parallelism of BPMN processes using model checking techniques. A paper has been submitted to an international conference.

- In collaboration with Pascal Poizat (University of Paris Ouest Nanterre), we have been working on the development of the VerChor platform, which aims at assembling all the verification techniques and tools automating the analysis of choreography specifications [14].

6.5.3. *Deployment and Reconfiguration Protocols for Cloud Applications*

Participants: Rim Abid, Gwen Salaün.

We collaborated with Noël de Palma and Fabienne Boyer (University Joseph Fourier), Xavier Etchevers and Thierry Coupaye (Orange Labs, Meylan, France) in the field of cloud computing applications, which are complex distributed applications composed of interconnected software components running on distinct virtual machines. Setting up, (re)configuring, and monitoring these applications involves intricate management protocols, which fully automate these tasks while preserving application consistency as well as some key architectural invariants.

In 2013, we focused on the reliability of the self-configuration protocol [23]. This protocol always succeeds in deploying a cloud application, even when facing a finite number of virtual machine or network failures. Designing such highly parallel management protocols is difficult, therefore formal modelling techniques and verification tools were used for validation purposes. These results were accepted for publication in an international conference [11]. Also, an experience export on the verification tasks for such (re)configuration protocols has been published in an international journal [8].

We have also worked on the design and verification of a reconfiguration protocol, where virtual machines interact altogether using a publish-subscribe messaging system. The verification of this protocol with CADP helped to refine several parts of the protocol and correct subtle bugs. These results have been published in an international conference [10]. In collaboration with Francisco Durán (University of Málaga), we have also worked on the design of a variant of this reconfiguration protocol, where the virtual machines interact via FIFO buffers. A paper has been submitted to an international conference.

6.5.4. *Networks of Programmable Logic Controllers*

Participants: Hubert Garavel, 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 a pivot language that will enable the connection to testing and verification tools covering the synchronous and asynchronous aspects. Our work focuses on the translation from the pivot language to LNT, which will provide a direct connection to all verification functionalities of CADP, in particular model checking and equivalence checking.

In 2013, we studied the existing approaches and languages that address formal modeling and verification of GALS systems. We concluded that the current landscape lacks general-purpose, flexible, and formal representation of GALS systems suitable for efficient verification. To fulfill this requirement, we have designed GRL (*GALS Representation Language*), a language with user-friendly syntax and formal semantics, to efficiently model GALS systems for the purpose of formal verification. GRL targets GALS systems consisting of networks of synchronous systems interacting with their environments and communicating via asynchronous media. GRL draws mainly from two foundations. Regarding asynchronous concurrency, GRL builds upon process calculi (in particular LNT). Thereby, it leverages process calculi expressiveness, versatility, and verification efficiency. Regarding synchronous features, GRL holds a dataflow-oriented model based on the dataflow diagram model (also called block-diagram model). The GRL synchronous model inherits from the simplicity and modularity of the block-diagram model.

We defined the lexical and the abstract syntax of GRL (about 80 grammar rules), its static semantics (about 150 binding, typing, and initialization rules), and its dynamic semantics (about 20 structured operational semantics rules). Using the SYNTAX and LOTOS NT compiler construction technology, we started the development of a prototype translator GRL2LNT (about 8000 lines). The tool currently performs the lexical and syntactic analysis of GRL programs, together with some static semantic checks. A database containing about 30 examples of GRL programs has been constructed and used for non-regression testing of GRL2LNT. A reference manual for GRL (130 pages up to now) containing the definition of the language and its translation to LNT has been written. A paper presenting the GRL language has been submitted to an international conference.

Regarding the analysis of PLC networks by equivalence checking, we defined variants of classic equivalence relations (strong, τ^*.a, and branching) for comparing the Mealy machine corresponding to a PLC network with the Moore machine corresponding to its external behaviour. We reformulated the verification problem as the resolution of a Boolean equation system, and we developed a prototype tool, based on the CAE-SAR_SOLVE_1 library, for the on-the-fly comparison of a Mealy and a Moore machine modulo the strong or the τ^*.a equivalences.

6.5.5. *EnergyBus Standard for Connecting Electric Components*

Participants: Hubert Garavel, Wendelin Serwe.

The EnergyBus [8] 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[9] 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 [49], with the active collaboration of CONVECS.

In 2013, CONVECS provided help in modelling using the LNT language and the TGV tool, and enhanced the CADP toolbox to address a number of issues reported by Saarland University. At present, this LNT specification (1670 lines) is used for generating test suites using the TGV tool [53]. The formal modelling prompted for modifications in the EnergyBus standard and the generated test suites revealed three unknown bugs in an industrial CANopen implementation.

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 co-operation with Gaëlle Calvary, Eric Ceret, 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 [35] designed to provide human operators with an overview of a running nuclear 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 status information and alarms resulting from faults, disturbances, or unexpected events in the plant. 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.

In 2013, CONVECS contributed to the following results. Based upon the available information published by EDF, a formal specification in LNT of this new-generation interface was developed (2600 lines). This specification not only encompasses 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, a few desirable properties of the interface have been expressed in the MCL language of CADP and verified on the LNT model.

So doing, three main difficulties have been faced. The description of the prototype available in the published literature is not exhaustive, which required us to provide those missing details needed to obtain a realistic model. Quite often, we faced a combinatorial explosion in the number of states of the model, which forced us to restrict the complexity of operator behaviour and fault models. Finally, this case study revealed several LNT-specific issues, which triggered enhancements in the LNT language and tools.

---

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

DANTE Team

6. New Results

6.1. Probabilistic resource management

Participants: Paulo Gonçalves [correspondant], Thomas Begin, Shubhabrata Roy, Thibaud Trolliet.

This contribution is part of the PhD work of S. Roy (Dec. 2010 – March 2014) on probabilistic resource management in the context of highly volatile workloads. We proposed a Markovian model that can reproduce the workload volatility occurring in real-life VoD systems, such as Video On Demand (VoD). We derived an original MCMC based identification procedure to calibrate model on real data. We assess the accuracy of the proposed procedure in terms of bias and variance through several numerical experiments, and we compared its outcome with a former ad-hoc method that we had designed. We also compared the performance of our approach to that of other existing models examining the goodness-of-fit of the steady state distribution and of the autocorrelation function of real workload traces. Results show that the combination of out model and its MCMC based calibration clearly outperforms the existing state-of-the art. (See [17], [18])

peta

6.2. Semi-supervised machine learning

Participant: Paulo Gonçalves [correspondant].

This contribution is part of the PhD work of M. Sokol (EPI MAESTRO, Oct. 2009 – May 2014), co-supervised with K. Avrachenkov and Ph. Nain, on the classification of content and users in peer-to-peer networks using graph-based semi-supervised learning methods. Semi-supervised learning methods constitute a category of machine learning methods which use labelled points together with unlabelled data to tune the classifier. The main idea of the semi-supervised methods is based on an assumption that the classification function should change smoothly over a similarity graph, which represents relations among data points. This idea can be expressed using kernels on graphs such as graph Laplacian. Different semi-supervised learning methods have different kernels which reflect how the underlying similarity graph influences the classification results. In a recent work, we analysed a general family of semi-supervised methods, provided insights about the differences among the methods and gave recommendations for the choice of the kernel parameters and labelled points. In particular, it appeared that it was preferable to choose a kernel based on the properties of the labelled points. We illustrated our general theoretical conclusions with an analytically tractable characteristic example, clustered preferential attachment model and classification of content in P2P networks. (See [8])

6.3. Analysis of heart beat rate variability

Participant: Paulo Gonçalves [correspondant].

Intrapartum fetal heart rate monitoring constitutes an important stake aiming at early acidosis detection. Measuring heart rate variability is often considered a powerful tool to assess the intrapartum health status of fetus and has has been envisaged using various techniques. In the present contribution, the power of scale invariance parameters, such as the Hurst exponent and the global regular- ity exponent, estimated from wavelet coefficients of intrapartum fetal heart rate time series, to evaluate the health status of fetuses is quantified from a case study database, constituted at a French Academic Hospital in Lyon. Notably, the ability of such parameters to discriminate subjects incorrectly classified according to FIGO rules as abnormal will be discussed. Also, the impact of the occurrence of decelerations identified as complicated by obstetricians on the values taken by Hurst parameter is investigated in detail. (See [7])

6.4. Hierarchical Modeling of IEEE 802.11 Multi-hop Wireless Networks

Participants: Thiago Wanderley Matos de Abreu, Thomas Begin, Isabelle Guérin Lassous.

IEEE 802.11 is implemented in many wireless networks, including multi-hop networks where communications between nodes are conveyed along a chain. We present a modelling framework to evaluate the performance of flows conveyed through such a chain. Our framework is based on a hierarchical modelling composed of two levels. The lower level is dedicated to the modelling of each node, while the upper level matches the actual topology of the chain. Our approach can handle different topologies, takes into account Bit Error Rate and can be applied to multi-hop flows with rates ranging from light to heavy workloads. We assess the ability of our model to evaluate loss rate, throughput, and end-to-end delay experienced by flows on a simple scenario, where the number of nodes is limited to three. Numerical results show that our model accurately approximates the performance of flows with a relative error typically less than 10%. (See [6])

6.5. Available Bandwidth Estimation in GPSR for VANETs

Participant: Isabelle Guérin Lassous.

We propose an adaptation of the collision probability used in the measure of the available bandwidth designed for Mobile Ad hoc Networks (MANETs) and which is used in ABE. Instead, we propose a new ABE+ that includes a new function to estimate the probability of losses. This new function has been specially designed for Vehicular Ad hoc Networks, to be suited to the high mobility and variable density in vehicular environments. In this new solution, we do not only consider the packet size, but also other metrics, such as, density and speed of the nodes. We include the ABE+ algorithm in the forwarding decisions of the GBSR-B protocol, which is an improvement of the well-known GPSR protocol. Finally through simulations, we compare the performance of our new ABE+ compared to the original ABE. These results show that ABE+ coupled with GBSR-B achieves a good trade-off in terms of packet losses and packet end-to-end delay. (See [19])

6.6. Reduced complexity in $M/Ph/c/N$ queues

Participant: Thomas Begin [correspondant].

This contribution stems from a long-existing collaboration with Pr. Brandwajn (UCSC), which is devoted to innovative numerical solution of classical queueing systems. Many real-life systems can be modelled using the classical $M/G/c/N$ queue. A frequently-used approach is to replace the general service time distribution by a phase-type distribution since the $M/Ph/c/N$ queue can be described by familiar balance equations. The downside of this approach is that the size of the resulting state space suffers from the "dimensionality curse", *i.e.*, exhibits combinatorial growth as the number of servers and/or phases increases. To circumvent this complexity issue, we propose to use a reduced state description in which the state of only one server is represented explicitly, while the other servers are accounted for through their rate of completions. The accuracy of the resulting approximation is generally good and, moreover, tends to improve as the number of servers in the system increases. Its computational complexity in terms of the number of states grows only linearly in the number of servers and phases. (See [9])

6.7. Throughput maximisation in multi-radio wireless networks

Participants: Isabelle Guérin Lassous, Busson Anthony.

Wireless mesh network offers a simple and costless solution to deploy wireless based infrastructure network. They are particularly suitable when the network is deployed temporarily, such as substitution networks (studied in the ANR RESCUE project). In order to ensure an important capacity, the mesh nodes may be equipped with several 802.11 network interfaces. The classical approach to assign 802.11 channels to these interfaces aim to minimise global interference, *i.e.* minimise the conflict graph. Our proposition is two folds. We define a new benefit function that describes the network capacity rather than interference/conflicts. Also, we derive an efficient algorithm that maximises this function. Simulation results show that the proposed function is very close to the measured end-to-end throughputs, empirically proving that it is the good function to optimise. Moreover, the channel assignation algorithm based on this optimisation presents an important throughput increase compared to the classical approaches.

6.8. Aggregation of temporal contact series into graph series

Participants: Christophe Crespelle, Eric Fleury, Yannick Léo.

We consider the problem of aggregating a temporal contact series into a series of graph. This consists in slicing time into time-windows of equal length and forming for each window the graph of the contacts occurred within it. The length chosen for the windows has a great impact on the properties of the graph series obtained. Then the key question that arises is: how one should choose the length of aggregation windows? In the master internship of Yannick Léo (spring 2013), we designed a method to do so, by using the occupation rate of paths in the graph series. We have applied this method on several real-world data and obtained very good results. The method has also greatly beneficiated of a new notion of shortest dynamic paths that we developed during the master internship of Pierre-Alain Scribot (spring 2013).

6.9. Dynamic Contact Network Analysis in Hospital Wards

Participants: Christophe Crespelle, Eric Fleury, Lucie Martinet.

We analysed a huge and very precise trace of contact data collected during 6 months on the entire population of a rehabilitation hospital. We investigated the graph structure of the average daily contact network, and we unveiled striking properties of this structure in the considered hospital, as a very strong introversion of services, the key role of the contacts between patients and staff in connecting those introverted services all together, and very different pattern of contacts during one day between patients and staffs. The methodology we designed to lead these analysis is very general and can be applied for analysing any dynamic complex network where nodes are classified into groups. Those results are part of Lucie Martinet's PhD thesis.

6.10. A Linear-Time Algorithm for Computing the Prime Decomposition of a Directed Graph with Regard to the Cartesian Product

Participant: Christophe Crespelle.

We design the first linear-time algorithm for computing the prime decomposition of a digraph G with regard to the cartesian product. A remarkable feature of our solution is that it computes the decomposition of G from the decomposition of its underlying undirected graph, for which there exists a linear-time algorithm. First, this allows our algorithm to remain conceptually very simple and in addition, it provides new insight into the connexions between the directed and undirected versions of cartesian product of graphs [11]

6.11. Linear-time Constant-ratio Approximation Algorithm and Tight Bounds for the Contiguity of Co-graphs

Participant: Christophe Crespelle.

We consider a graph parameter called *contiguity* which aims at encoding a graph by a linear ordering of its vertices. The purpose is to obtain very compact encoding of a graph which still answers in optimal time to neighbourhood queries on the graph (*i.e.* list the neighbours of a given vertex). This allows to deal with very large instances of graphs by loading them entirely into the memory, without penalising the running time of algorithms treating those instances. We designed a linear time algorithm for computing a constant-ratio approximation of the contiguity of an arbitrary co-graph. Our algorithm does not only give an approximation of the parameter, but also provides an encoding of the co-graph realising this value [10]

6.12. Model for Time-Varying Graphs.

Participant: Éric Fleury.

We propose a novel model for representing finite discrete Time-Varying Graphs (TVGs). The major application of such a model is for the modelling and representation of dynamic networks. In our proposed model, an edge is able to connect a node u at a given time instant t_a to any other node v (u possibly equal to v) at any other time instant t_b (t_a possibly equal to t_b), leading to the concept that such an edge can be represented by an ordered quadruple of the form (u, t_a, v, t_b). Building upon this basic concept, our proposed model defines a TVG as an object $H = (V, E, T)$, where V is the set of nodes, $E \subseteq V \times T \times V \times T$ is the set of edges, and T is the finite set of time instants on which the TVG is defined. We show how key concepts, such as degree, path, and connectivity, are handled in our model. We also analyse the data structures used for the representation of dynamic networks built following our proposed model and demonstrate that, for most practical cases, the asymptotic memory complexity of our TVG representation model is determined by the cardinality of the set of edges. (See [20])

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

6. New Results

6.1. Economy of the world data flows

We have attempted to measure data flows in the world to estimate the concentration of the data industry. It is well known that the main plateforms of the Web, Google, Facebook, Amazon, etc. are concentrated in a few countries, mostly in the USA. Some countries, mostly asian, such as China, Russia, Korea or Japan have succesfully developed their own Web 2.0 industry, while others, such as European countries, hava failed to do so. We have explored in [6], the strategy of China, which has the largest Web industry behind the US and has made a priority of keeping its data at home, with systems in all activity sectors developed in general only one or two years after their main americain counterparts. The innovation strategy of China aims in all fields to achieve technological independence, with at most 30% of foreign IP as we have shown in [2].

A tentative measure of the flows of personal data from different regions is proposed in [3], based on the traffic on the largest platform at the international level. We show in particular that personal data captured in Europe on Web platforms mostly go to the US industry. In [4], we investigate the invisible part of the Web constituted by the trackers that are hidden on Web pages and transfer data to third parties, and show that the domination of the US is even stronger on trackers than it is on the visible Web.

6.2. Flow systems

We are currently working with Bull SA, Manuel Selva (PhD) and Lionel Morel from the Socrates team to build a monitoring framework for dynamic data-flow system in many-core architectures. Data-flow computing models computation as a pipeline of computation units absorbing a continuous stream of data. This computing model suits application development for embedded devices such as MPEG-4 video encoders. The incoming data flow is sliced into small size token (e.g. video frames). Each time, all computational units take some tokens from their inputs and produce some tokens on their outputs. We focus [7], [8] on a management layer for handling dynamic dataflow programs in many-core architectures, where computation units may be relocated at runtime from one core to another. The questions raised by Twitter Storm, Google Millwheel or Yahoo S4, are in essence very similar. Can our current architectures hold the information dataflow produced by users in terms of computing power and memory usage? We are currently extending these embedded results to study dataflow architectures with ATOS on flow computing inside Web browsers.

François Goichon will defend his PhD on resource access equity into best-effort operating systems such as Linux. Linux is built over a layered architecture, where each layer owns a local policy that may lead to a global policy being far from best-effort. With Guillaume Salagnac from Socrates team, we show [5], [9] that we can develop malware user space applications exploiting embedded linux firmware and device drivers differential policy that can block other concurrent applications from accessing CPU time. When this kind of applications are installed in multi-tenant architectures as found in cloud shared space, it can slowdown the entire system. These results are interesting for Dice when considering access time in web browser. Current in-browser applications are developed in Javascript, which imposes a single threaded executed model to the developer, yet operated on a multi-core architecture. Best-effort operating systems are not the best approaches to handle flow based applications that become the norm, and we think that some small, low-level shifts, should be considered.

DRACULA Project-Team

6. New Results

6.1. Mathematical modeling in chronobiology

Circadian clocks are autonomous oscillators entrained by external Zeitgebers such as light-dark and temperature cycles. On the cellular level, rhythms are generated by negative transcriptional feedback loops. In mammals, the suprachiasmatic necleus (SCN) in the anterior part of the hypothalamus plays the role of the central circadian pacemaker. Coupling between individual neurons in the SCN leads to specify self-sustained oscillations even in the absence of external signals. These neuronal rhythms orchestrate the phasing of circadian oscillations in peripheral organs. Altogether, the mammalian circadian system can be regarded as a network of coupled oscillators. In order to understand the dynamic complexity of these rhythms, mathematical models successfully complement experimental investigations. In [19], we discuss basic ideas of modeling on three different levels: (i) rhythm generation in single cells by delayed negative feedbacks, (ii) synchronization of cells via external stimuli or cell-cell coupling, and (iii) optimization of chronotherapy.

6.2. Hybrid Models of Cell Population

The paper [20] is devoted to hybrid discrete-continuous models of cell populations dynamics. Cells are considered as individual objects which can divide, die by apoptosis, differentiate and move under external forces. Intra-cellular regulatory networks are described by ordinary differential equations while extracellular species by partial differential equations. We illustrate the application of this approach to some model examples and to the problem of tumor growth. Hybrid models of cell populations present an interesting nonlinear dynamics which is not observed for the conventional continuous models.

6.3. Multiscale Model in Biology

Biological processes span several scales in space, from the single molecules to organisms and ecosystems. Multiscale modelling approaches in biology are useful to take into account the complex interactions between different organisation levels in those systems. We present in [6] several single- and multiscale models, from the most simple to the complex ones, and discuss their properties from a multiscale point of view. Approaches based on master equations for stochastic processes, individual-based models, hybrid continuous-discrete models and structured PDE models are presented.

6.4. Model of hematopoiesis

In [2], a model of blood cell production in the bone marrow (hematopoiesis), has been investigated. It describes both the evolution of primitive hematopoietic stem cells and the maturation of these cells as they differentiate to form the three kinds of progenitors and mature blood cells (red blood cells, white cells and platelets). The three types of progenitors and mature cells are coupled to each other via their common origin in primitive hematopoietic stem cells compartment. The resulting system is composed by eleven age-structured partial differential equations. To analyze this model, we don't take into account cell age-dependence of coefficients, that prevents a usual reduction of the structured system to an unstructured delay differential system. We study the existence of stationary solutions: trivial, axial and positive steady states. Then we give conditions for the local asymptotic stability of the trivial steady state and by using a Lyapunov function, we obtain a sufficient condition for its global asymptotic stability. In some particular cases, we analyze the local asymptotic stability of the positive steady state by using the characteristic equation. Finally, by numerical simulations, we illustrate our results and we show that a change in the duration of cell cycle can cause oscillations.

6.5. Analysis of radiocarbon to facilitate identification of unknown decedents

The characterization of unidentified bodies or suspected human remains is a frequent and important task for forensic investigators. However, any identification method requires clues to the person's identity to allow for comparisons with missing persons. If such clues are lacking, information about the year of birth, sex and geographic origin of the victim, is particularly helpful to aid in the identification casework and limit the search for possible matches. We present in [4] results of stable isotope analysis of $(13)C$ and $(18)O$, and bomb-pulse $(14)C$ analyses that can help in the casework. The $(14)C$ analysis of enamel provided information of the year of birth with an average absolute error of 1.8 ± 1.3 years. We also found that analysis of enamel and root from the same tooth can be used to determine if the $(14)C$ values match the rising or falling part of the bomb-curve. Enamel laydown times can be used to estimate the date of birth of individuals, but here we show that this detour is unnecessary when using a large set of crude $(14)C$ data of tooth enamel as a reference. The levels of $(13)C$ in tooth enamel were higher in North America than in teeth from Europe and Asia, and Mexican teeth showed even higher levels than those from USA. DNA analysis was performed on 28 teeth, and provided individual-specific profiles in most cases and sex determination in all cases. In conclusion, these analyses can dramatically limit the number of possible matches and hence facilitate person identification work.

<div align="center">

E-MOTION Project-Team

</div>

5. New Results

5.1. Perception and Situation Awareness in Dynamic Environments

5.1.1. *Sensor Fusion for state parameters identification*

Participants: Agostino Martinelli, Chiara Troiani.

5.1.1.1. *General theoretical results*

We continued to investigate the visual-inertial structure from motion problem by further addressing the two important issues of observability and resolvability in closed form. Regarding the first issue, we extended our previous results published last year on the journal of Transaction on Robotics [44] by investigating the case when the visual sensor is not extrinsically calibrated. In order to deal with this case, we must augment the state to be estimated by including all the parameters that characterize the extrinsic camera calibration, i.e., the six parameters that describe the relative transformation between the frame attached to the camera and the frame attached to the Inertial Measurement Unit (IMU). On the other hand, because of the larger size of the resulting state, it became prohibitive a direct application of the method that we introduced two years ago (see [43]) in order to discover the observability properties for this new state. For this reason, our first novel contribution during this year was the introduction of new methodologies able to significantly reduce the computational burden demanded by the implementation of the method in [43]. These methodologies have been published in [22] and a deeper description of their use is currently under revision on the journal Foundations and Trends in Robotics. The new results obtained by using these methodologies basically state that also the new six parameters that describe the camera extrinsic calibration are observable. Finally, we started a new research that in the literature is known as the Unknown Input Observability (UIO) and it is investigated by the automatic control community. We started this new research since we investigated the observability properties of the visual inertial structure from motion as the number of inertial sensors is reduced. Specifically, instead of considering the standard formulation, which assumes a monocular camera, three orthogonal accelerometers and three orthogonal gyroscopes, the considered sensor suit only consists of a monocular camera and one or two accelerometers. This analysis has never been provided before. A preliminary investigation seems to prove that the observability properties of visual inertial structure from motion do not change by removing all the three gyroscopes and one accelerometer. By removing a further accelerometer, if the camera is not extrinsically calibrated, the system loses part of its observability properties. On the other hand, as the camera is extrinsically calibrated, the system maintains the same observability properties as in the standard case. This contribution clearly shows that the information provided by a monocular camera, three accelerometers and three gyroscopes is redundant. Additionally, it provides a new perspective in the framework of neuroscience to the process of vestibular and visual integration for depth perception and self motion perception. Indeed, the vestibular system, which provides balance in most mammals, consists of two organs (the *utricle* and the *saccule*) able to sense the acceleration only along two independent axes (and not three). In order to analyze these systems with a reduced number of inertial sensors, we had to consider control systems where some of the inputs are unknown. Indeed, the visual-inertial structure from motion problem can be characterized by a control system where the inputs are known thanks to the inertial sensors. Hence, to deal with the visual-inertial structure from motion as the number of inertial sensors is reduced, we had to introduce a new method able to address the more general UIO problem. We believe that our solution to the UIO is general and this is the reason because we started this new research domain in control theory. Preliminary results are currently under revision on the journal Foundations and Trends in Robotics and we also plan to present them at the next ICRA conference. Regarding the second issue, i.e., the problem resolvability in closed form, a new simple closed form solution to visual-inertial structure from motion has been derived. This solution expresses the structure of the scene and the motion only in terms of the visual and inertial measurements collected during a short time interval. This allowed us to introduce deterministic algorithms able to simultaneously determine the structure

of the scene together with the motion without the need for any initialization or prior knowledge. Additionally, the closed-form solution allowed us to identify the conditions under which the visual-inertial structure from motion has a finite number of solutions. Specifically, it is shown that the problem can have a unique solution, two distinct solutions or infinite solutions depending on the trajectory, on the number of point-features and on their arrangement in the 3D space and on the number of camera images. All the results have been published on the international journal of Computer Vision [15].

5.1.1.2. Applications with a Micro Aerial Vehicle

We introduced a new method to localize a micro aerial vehicle (MAV) in GPS denied environments and without the usage of any known pattern [26]. The method exploits the planar ground assumption and only uses the data provided by a monocular camera and an inertial measurement unit. It is based on a closed solution which provides the vehicle pose from a single camera image, once the roll and the pitch angles are obtained by the inertial measurements. Specifically, the vehicle position and attitude can uniquely be determined by having two point features. However, the precision is significantly improved by using three point features. The closed form solution makes the method very simple in terms of computational cost and therefore very suitable for real time implementation. Additionally, because of this closed solution, the method does not need any initialization. Results of experimentation show the effectiveness of the proposed approach.

We proposed a novel method to estimate the relative motion between two consecutive camera views, which only requires the observation of a single feature in the scene and the knowledge of the angular rates from an inertial measurement unit, under the assumption that the local camera motion lies in a plane perpendicular to the gravity vector [27]. Using this 1-point motion parametrization, we provide two very efficient algorithms to remove the outliers of the feature-matching process. Thanks to their inherent efficiency, the proposed algorithms are very suitable for computationally-limited robots. We test the proposed approaches on both synthetic and real data, using video footage from a small flying quadrotor. We show that our methods outperform standard RANSAC-based implementations by up to two orders of magnitude in speed, while being able to identify the majority of the inliers.

5.1.2. A new formulation of the Bayesian Occupancy Filter : an hybrid sampling based framework

Participants: Lukas Rummelhard, Amaury Nègre.

The Bayesian Occupancy Filter (BOF) is a discretized grid structure based bayesian algorithm, in which the environment is subdivised in cells to which random variables are linked. These random variables represent the state of occupancy and the motion field of the scene, without any notion of object detection and tracking, making the updating part of the filter an evaluation of the distribution of these variables, according to the new data acquisition. In the classic representation of the BOF, the motion field of each cell is represented as a neighborhood grid, the probability of the cell moving from the current one to another of the neighborhood being stocked in an histogram. If this representation is convenient for the update, since the potential antecedents of any cell is exactly determined by the structure, and so the propagation model is easily parallelizable, it also raises determinant issues :

- the structure requires the process rate to be constant, and a priori known.
- in the case of a moving grid, such as an application of car perception, many aliasing problems can appear, not only in the occupation grid, but in the motion fields of cells. A linear interpolation in 4-dimension field to fill each value of the histograms can quickly become unreasonable.
- to be able to match the slowest moves in the scene and the tiniest objects, the resolution of the grid and the motion histogram must be the high. On the other hand, since the system must be able to evaluate the speed of highly dynamic objects (typically, a moving car), the maximum encoded speed is to be high as well. This results in a necessary huge resolution grid, which prevent the system from being used with satisfying results on an embedded device. This huge grid is also mostly empty (most of the motion field histogram for a occupied cell will be empty). On top of that, the perception system being used to represent the direct environment of a moving car, the encoded velocity is a relative velocity, which implies, if we consider the maximal speed of a car to be V_max, to maintain

a motion field able to represent speeds from $-2 * V_{max}$ to $2 * V_{max}$. The necessity of such a sized structure is a huge limitation of practical use of the method.

Considering those limitations, a new way to represent the motion field has been developped. To do so, a new formulation of the BOF has been elaborated. This new version allow to introduce in the filter itself a distinction between static and dynamic parts, and so adapt the computation power. The main idea of this new representation is to mix two forms of sampling : a uniform one, represented as a grid, for the static objects and the empty areas, and a non uniform one, based on particles drawn from dynamic regions. The motion field in a cell will be represented as a set of samples from the distribution for values which are not null, and a weight given to the static hypothesis. The use of a set of samples to represent the motion field leads to a important decrease of the needed memory space, as well as the classification between dynamic objects and static objects or free areas. In the updating process, the antecedent of a cell can be either from the static configuration or from the dynamic configuration, which are both way easier to project in the new reference frame of the moving grid. The first results are stimulating, in term of occupancy evaluation and mostly in term of velocity prediction, being way more accurate and responsive than the older version. Those improvements will soon be presented in detail in upcoming papers, one being currently in the process of redaction.

5.1.3. DATMO

Participants: Dung Vu, Mathias Perrollaz, Amaury Nègre.

In the current work, we have been developing a general framework for tracking multiple targets from lidar data.

In the past decades, multiple target tracking has been an active research topic. When object observations are known, object tracking becomes a data association (DA) problem. Among popular DA methods, multiple hypothesis tracking (MHT) is widely used. MHT is a multi-frame tracking method that is capable of handling ambiguities in data association by propagating hypotheses until they can be solved when enough observations are collected. The main disadvantage of MHT is its computational complexity since the number of hypotheses grows exponentially over time. The joint probabilistic data association (JPDA) filter is more efficient but prone to make erroneous decision since only single frame is considered and the association made in the past is not reversible. Other sequential approaches using particle filters share the same weakness that they cannot reverse time back when ambiguities exist. All DA approaches mentioned above requires a strong assumption of one-to-one mapping between targets and observations which is usually violated in real environments. For instance, a single object can be seen by several observations due to occlusion, or multiple moving objects can be merged into a single observation when moving closely.

In the research, we propose a new data association approach that deals with split/merge nature of object observations. In addition, our approach also tackles ambiguities by taking into account a sequence of observations in a sliding window of frames. To avoid the high computational complexity, a very efficient Markov Chain Monte Carlo (MCMC) technique is proposed to sample and search for the optimum solution in the spatio-temporal solution space. Moreover, various aspects including prior information, object model, motion model and measurement model are explicitly integrated in a theoretically sound framework.

5.1.4. *Visual recognition for intelligent vehicles*

Participants: Alexandros Makris, Mathias Perrollaz, Christian Laugier.

We have developed an object class recognition method. The method uses local image features and follows the part-based detection approach. It fuses intensity and depth information in a probabilistic framework. The depth of each local feature is used to weigh the probability of finding the object at a given distance. To train the system for an object class, only a database of images annotated with bounding boxes is required, thus automatizing the extension of the system to different object classes. We apply our method to the problem of detecting vehicles from a moving platform. The experiments with a data set of stereo images in an urban environment show a significant improvement in performance when using both information modalities.
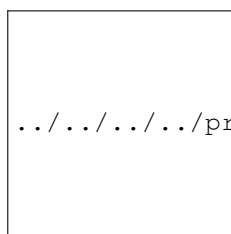
In 2013, the method has been published in IEEE Transactions on Intelligent Transportation Systems [14].

../../../../projets/e-motion/IMG/pbog0_cam.jpg ../projets/e-motion/IMG/pbog1_cam.jpg

(a) Camera image, two-way street   (b) Camera image, highway

../../../../projets/e-motion/IMG/pbog0_occ_speed.jpg ../projets/e-motion/IMG/pbog

(c) Occupancy and velocity result, two-way street   (d) Occupancy and velocity result, highway
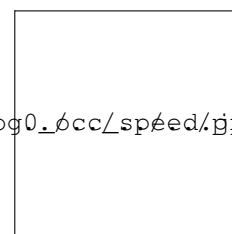
Figure 1. Results of the new algorithm : the free cells are black, the occupied ones are white. Unknown areas are grey. When a cell is seen as dynamic, a vector representing the average of the associated motion field is drawn in red.

5.1.5. Experimental platform for road perception

5.1.5.1. Experimental platform material description

Our experimental platform for road perception is shown in Figure 2 . This platform is a commercial Lexus car LS600h equipped with a variety of sensor, including two IBEO Lux lidars placed toward the edges of the front bumper, a TYZX stereo camera plus a high resolution color camera situated behind the windshield, and an Xsens MTi-G inertial sensor with GPS. To do online data computation and data acquisition, a standard computer is located on the back of the car.

This platform allows us to conduct experimentation and data acquisition in various road environments (country roads, downtown and highway), at different time of the day, with various driving situations (light traffic, dense traffic, traffic jams).

../../../../projets/e-motion/IMG/plateforme_lexus.png

Figure 2. Lexus LS600h car equipped with two IBEO Lux lidars, a stereo plus a monocular camera, and an Xsens MTi-G inertial sensor with GPS.

5.1.5.2. Software architecture

The perception and situation awareness software architecture is integrated in the ROS framework. ROS (http://www.ros.org) is an open source robotics middleware designed to be distributed and modular. For the Lexus platform, we developed a set of ROS module for each sensor and for each perception component. Each

perception module can be dynamically connected with the required drivers or other perception modules. The main architecture of the perception components is illustrated on Figure 3 .

../../../../projets/e-motion/IMG/architecture_lexus.png

Figure 3. Architecture of the main perception components in the Lexus platform.

5.1.6. Software and Hardware Integration for Embedded Bayesian Perception

Participants: Mathias Perrollaz, Christian Laugier, Qadeer Baig, Dizan Vasquez, Lukas Rummelhard, Amaury Nègre.

The objective of this recently started research work is to re-design in a highly parallel fashion our Bayesian Perception approach for dynamic environments (based on the BOF concept), in order to deeply integrate the software components into new multi-processor hardware boards. The goal is to miniaturize the software/hardware

perception system (i.e., to reduce the size, the load, the energy consumption and the cost, while increasing the efficiency of the system).

For supported this research, we began to work in the "Perfect" project. This project, included in the IRT-Nano program, involves the CEA-LETI DACLE lab and ST-Microelectronics. Perfect is focusing onto the second integration objectives (6 years) and the development of integrated open platforms in the domain of transportation (vehicle and infrastructure) and in a second step in the domain of health sector (mobility of elderly and handicapped people, monitoring of elderly people at home...). The objective of e-Motion in this project is to transfer and port its main Bayesian perception modules from traditional computing system to embedded low-power multi-processors board. The targeted board is a STHorm from ST Microelectronics which is has a many-core architecture with a very low consumption. In 2013 we worked with the CEA to obtain a first implementation of the Bayesian occupancy grid filter on STHorm. Those preliminary results demonstrated the feasibility of the concepts but highlighted some key points to improve such as the memory footprint we need to reduce to obtain real-time accurate results.

../../../../projets/e-motion/IMG/BOF_Integration.png

Figure 4. First objective for software/hardware of Bayesian perception : developing and using multiple processor board from mobile technologies. The approach will be validated with real demonstrators..

5.2. Dynamic Change Prediction and Situation Awareness

5.2.1. *Vision-based Lane Tracker*

Participants: Mathias Perrollaz, Amaury Nègre.

In the field of vehicle risk assessment system, vehicle to road localization is an essential information to predict drivers behaviors as well as collision risk evaluation. To achieve this task, we have developed a vision based lane tracker to estimate the geometry of the lane using the line markers. Previous development was based on a Monte-Carlo particle filter to estimate simultaneously the road plane orientation, the lane curvature and the camera position. To perform the particle evaluation, the algorithm first process a ridge extraction of the camera image and then projected the left and right marker line represented by the particle on the ridge image.

The first improvement of the tracker consists on dynamically adapting the scale of the ridge filter to improve the efficiency and the precision of the particle evaluation. The second improvement is capability the perform multi-lane tracking for example in highway environment. To solve this problem, the position of the lane is added to the state and the particle evaluation consider the total number of line marker (a-priori known). Figure 5 shows the results of the lane tracker program on a highway environment.

5.2.2. *Vision-based Lane Change Prediction*

Participants: Suryansh Kumar, Dizan Vasquez, Mathias Perrollaz, Stephanie Lefevre, Amaury Nègre, Maiwen Gault.

For both Advanced Driving Assistance Systems and Autonomous Vehicles, it is very important to have the capability of predicting and understanding the driver's behavior. This work addresses this subject in a bottom-up fashion by first detecting low-level "atomic" maneuvers which can be used as a building block for more complex behavior. Concretely, we have developed a learning-based approach that uses lane tracking data to predict lane changes.

Most works in the literature address this as a classification problem, and often use some version of Support Vector Machines (SVM) to solve it. The problem with this approach is that it is sensitive to noise and can yield high-frequency oscillations in the obtained predictions, moreover, they do not provide any information concerning the Time To Change (TTC). Other approaches use a filtering approaches, using Hidden Markov Models (HMM), for example. Although they produce smoother predictions and, in some cases, even a TTC estimate, some studies [33] suggest that HMM-based approaches are less accurate than those based on SVMs.

Our work combines the advantages of both approaches in a hierarchical fashion. First, lane tracking data (i.e. lane-relative yaw, lane-relative lateral position and their fist-order derivatives) are used as an input of a multi-class SVM. Then, the Bradley-Terry model is used to translate the SVM output into a probability which is used as the observation model of a Bayesian filter (Fig. 6).

This work has been published in the Intelligent Vehicles conference [18]. Since then, we have worked on an improved HMM-only approach which addresses the shortcomings of similar approaches by using a continuous observation model. In our preliminary experiments, this approach leads to improved predictions over the hybrid one.

5.2.3. *Feature-based human behavior modeling*

Participants: Suryansh Kumar, Yufeng Yu, Dizan Vasquez.

When it comes to modeling and learning complex human behavior, the preferred approach in the literature is to try to learn the typical motion patterns that people or vehicles get engaged into in a given environment. This has, however, a major drawback: the learned patterns only apply to the environment where they have been learned. This means that, for a new environment, previous knowledge cannot be used and patterns should be learned from scratch.

This situation has been recently addressed [49] by a family of approaches which rely on two complementary hypotheses:

- *Agents behave like planners*. The idea is that when people or other intelligent agents move around, they plan in order to minimize a cost function. Thus, if this function is known, it fully determines the agent's behavior.

../../../../projets/e-motion/IMG/lane_tracker_camera.jpg

(a) Camera image

../../../../projets/e-motion/IMG/lane_tracker_particles.jpg ../../projets/e-motion/IMG/lane_tracker_

(b) Particles cloud   (c) Lane tracker result

Figure 5. Visual Particle based lane tracking.

../../../../projets/e-motion/IMG/lane_change-hybrid.png

Figure 6. Overview of our hybrid lane change prediction approach.

- *The cost is a function of local features*. This extends the previous hypothesis by assuming that the cost function does not only depends on the agent's state but also in a number of external features (e.g. local people density, nearby traffic signs). Since the cost function depends on the features, it is possible to compute it even for previously unseen environments, as long as they contain the same kind of features.

Under these assumptions, the problem becomes that of learning the unknown cost function by observing how people move. This is often called *apprenticeship learning* and, when the underlying planning model is a Markov Decision Process, *inverse reinforcement learning*.

We have been working on a software library and evaluation testbed for different features and cost-function learning algorithms. We have conducted, in collaboration with the University of Freiburg, a first round of experiments concerning people moving in crowds. The results will be the subject of a submission to the IROS 2014 conference.

../../../../projets/e-motion/IMG/torcs.png

Figure 7. The TORCS racing simulator.

We have also been working on an application for intelligent vehicles and ADAS. As a first step, we have developed a ROS interface for the TORCS racing simulator, as well as a road simulation using the same platform (Fig. 7). This work has been the product of a collaboration with the Beijing University and IIIT Hyderabad.

5.2.4. *Safety applications at road intersections for connected vehicles*
Participants: Stéphanie Lefèvre, Christian Laugier.

From a safety perspective, road intersections are the most dangerous areas in the road network. They are also the most complex. Because of the extended situational awareness that they provide, wireless vehicular communications (or Vehicle-to-X communications, V2X) could greatly reduce the rate of intersection accidents. However, numerous research challenges remain before the full use of this technology can be achieved. A PhD was started on this topic in 2009 in collaboration with Renault, and was successfully defended in 2012 [42]. The purpose was to formulate and develop a probabilistic reasoning framework which would allow combining the information shared by the vehicles to estimate the situation and the associated risk as a vehicle negotiates an intersection. The first contribution of the PhD was to model the motion of vehicles using a Dynamic Bayesian Network where the maneuvers of different vehicles influence each other via an "expected maneuver". This "expected maneuver" represents what a driver is expected to do given the state of the other vehicles in the area and the traffic rules which apply at the intersection. Thanks to the use of a probabilistic framework, uncertainties related to sensor errors and interpretation ambiguities are handled. The second contribution was a novel approach to risk estimation based on the comparison between what drivers intend to do and what they are expected to do. The reasoning is carried out by performing inference on the Dynamic Bayesian Network introduced earlier, using a particle filter. The approach was validated with field trials using Renault passenger vehicles equipped with vehicle-to-vehicle wireless communication modems [41], and in simulation [40]. The results show that the algorithm is able to detect dangerous situations early and complies with real-time constraints. We also developed a theoretical extension of the model to generalize it to arbitrary traffic situations [29]. This work is still ongoing thanks to an Inria@SiliconValley fellowship granted to S. Lefevre at the end of her PhD. Since January 2013 she is working in the Teleimmersion group at the University of California Berkeley, as a postdoctoral researcher. The research conducted there lead to two new developments on the topic of "Safety applications at road intersections for connected vehicle".

The first development concerns probabilistic decision making for Collision Avoidance (CA) systems. In the processing chain of a CA system, the "Decision making" module follows the "Risk assessment" module. The research done during the PhD stopped at the "Risk assessment" module, and we now address the challenges present in the "Decision making" module. We identified two main challenges:

1. The first one is that the decision making module has to make decisions based on uncertain knowledge. Sensors provide noisy measurements, digital maps contain errors, and interpreting a vehicle's motion in terms of driver intention is uncertain. These uncertainties propagate to the risk assessment module and to the decision making module, but the latter is still required to make a decision from that uncertain knowledge.

2. The second one is that the timing of interventions is critical. If an intervention is triggered at a time when the uncertainty about the occurrence of a collision is too large, there is a chance that it will end up being a false alarm. High false alarm rates are detrimental to the driver acceptance of safety systems and can lead to the user losing trust in the system. On the other hand, if the system waits until the last moment (certainty about the occurrence of a collision) to trigger an intervention, it might be too late to avoid the accident.

In recent work [20] we proposed to introduce the possibility for a CA system to postpone making a decision. Our objective is to implement the fact that in some situations the new observations obtained by waiting will reduce the uncertainty about the occurrence of a collision, therefore the decision will be more reliable if it is made later using this additional information. The important question to solve is whether the potential gain brought by the additional information outweighs the cost of waiting. In order to answer this question, our decision making approach runs a *preposterior analysis* to determine the expected value and cost of the additional information. The value of the additional information can be quantified by means of the Expected Value of Sample Information (EVSI). It corresponds to the additional expected payoff possible through knowledge of the additional information and is computed by subtracting the expected costs of deciding with and without additional information. The cost of the additional information is quantified by means of the Expected Cost of Waiting (ECW). It is computed as the difference between the probability that the CA system will be able to avoid the potential collision if it intervenes now and if it intervenes at time t+1. Our decision making strategy is to postpone the decision making process to time t+1 if and only if the EVSI is positive and the ECW is null. The algorithm was tested in simulation at a two-way stop intersection for collision scenarios

and no-collision scenarios involving two vehicles. A comparative evaluation with a decision making strategy which does not allow postponing decisions showed that our approach generates fewer false alarms and avoids as many collisions. These results were published at the conference IEEE IROS'13 [20] and a patent application was filed with UC Berkeley [32].

The second development addresses the relationship between privacy strategies for V2X and safety applications which rely on V2X. User privacy is a requirement for wireless vehicular communications, and a number of privacy protection strategies have already been developed and standardized. In particular, methods relying on the use of temporary pseudonyms and silent periods have proved their ability to confuse attackers who would attempt to track vehicles. However, these privacy protection schemes are not without consequences for safety applications. Such applications make decisions (e.g. warning drivers of an upcoming danger) based on their current estimation of the state of the real world, and this representation is created from the information contained in beacons received from other vehicles. Therefore, interruptions in the transmission of information will impact the decision-making process. If a silent period is scheduled to start at a safety-critical moment, it could result in safety systems not intervening when they should have, namely a "missed intervention". From a user and safety perspective, this is not acceptable. In this work we address this issue and evaluate the impact of pseudonym change strategies on V2X-based Intersection Collision Avoidance (ICA) system. We use the ICA system developed during the PhD and simulated 3 different privacy protection strategies:

1. The "*Fixed ID*" strategy assigns a fixed pseudonym to a vehicle for the entire duration of a trip (i.e. a new pseudonym is assigned to the vehicle every time it starts). Testing this case gives us a reference for how well the collision avoidance system performs when there is no pseudonym change and no silent period during a trip, which is what was assumed in our PhD work.

2. The "*Baseline*" strategy follows the recommendations of the SAE J2735 standard for V2X communications. Pseudonyms are changed every 120 seconds and are followed by a silent period of random duration.

3. The "*Adaptive*" strategy is a modified version of the *Baseline* strategy where the risk of the situation is taken into account to decide whether or not a vehicle should be allowed to change pseudonym at time t. It relies on the estimation of the current ability of the collision avoidance system to keep the vehicle on a collision-free trajectory. The idea here is to authorize a pseudonym change and silent period only if it will not affect the performance of the safety application.

Simulations were conducted using the same simulator and the same scenarios as the tests run during the PhD. The performance of the three privacy strategies was evaluated both in terms of privacy and in terms of successful interventions of the ICA system. The results show that the ICA application requires silent periods to be shorter than two seconds in order to operate correctly in conjunction with the SAE J2735 standard. They also indicate that the addition of simple rules which authorize or not a pseudonym change depending on the context leads to major safety improvements compared to the SAE J2735 standard alone (see Figure 8). These results, which were published at the conference IEEE VNC'13 [21], highlight the necessity of a joint design. That is, the requirements of safety applications should be taken into account when designing privacy strategies, and pseudonym change schemes should be accounted for when designing safety applications which rely on V2X communications. This collaboration is necessary in order to ensure that vehicular communications and safety applications do not neutralize each other, but instead, work together toward safer roads.

5.2.5. *Guidance for Uncertain shooting domain*

Participant: Emmanuel Mazer.

This study is made in collabortion with MBDA (Monsieur Le Menec) and Probayes (Monsieur Laurent Saroul) under the ITP framework financed by the british MOD and the french DGA.

From N°51 of Vector Magazine (copyright MBDA) *It's an enduring question facing those in military conflict: 'when do you pull the trigger?' However, for pilots in air combat there is an added question: 'Once you've pulled the trigger, when do you break the link with your missile?' A new answer to that problem was a highlight of the MCM-ITP conference at Lille in May (see Vector 50). Entitled 'Guidance in Uncertain Shooting Domains', this joint project between MBDA and French company Probayes has produced a set of algorithms*

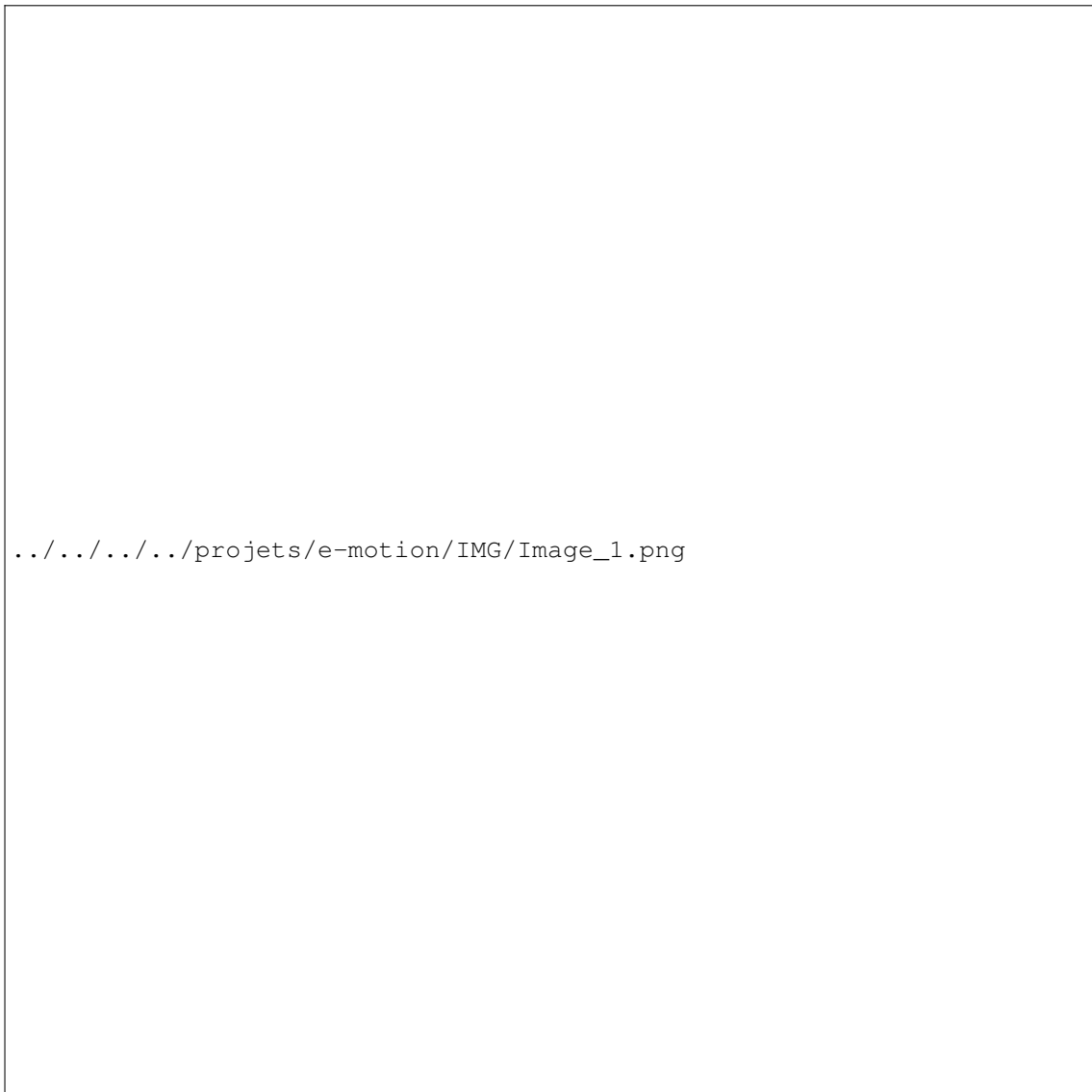../../../../projets/e-motion/IMG/Image_1.png

Figure 8. Percentage of missed interventions, avoided collisions, and failed interventions as a function of the duration of the silent period for the Baseline strategy (left) and the Adaptive strategy (right).

to help a pilot decide when to break a telemetry link. It's a decision fraught with danger. As pilots approach a target, the longer they keep the link, the greater the chance of their missile finding its target. However, closing in on the target usually means entering the enemy's own kill zone. So, a calculation of the best trade-off between mission success and pilot safety needs to be made.

"At the moment, when a pilot is engaging an enemy aircraft, he's obliged to do sums in his head," explained Graham Wallis, MBDA UK's Chief Technologist. "What we're trying to do is to take that away, and hide it behind a probability display." The problem arises because a seeker's scanning range is often far shorter than the maximum travelling distance, or kinematic range, of a missile. Though less of an issue at short range, it's a problem for medium-range weapons – where the target is likely to manoeuvre beyond the initial seeker scan area; hence the need for guidance from the pilot, who may have the target in radar sight.

SUCCESS PROBABILITY Stepping back, there are two sides to the firing equation. As customers require an air-to-air missile that will find its target almost 100 of the time, MBDA provides distance data (known as Launch Success Zone tables) to achieve that. They are understandably conservative. Clearly, though, the final decision to break the datalink can only be taken by the pilot, who is also making mental calculations to minimise his or her exposure to deadly risk. GUSD effectively offers a 'probability meter' to help reconcile those two imperatives.

Physically, GUSD could be a display with four bars and the circle of a pie-chart. Each bar represents the probability in percentage terms of an enemy pilot adopting one of four typical behaviours during air combat: flying head-on to attack; turning tail and heading for home, and either turning left or right – along with the probabilities that the enemy has just launched its own missile. The pie chart gives a single percentage – of the MBDA missile's chances of hitting its target. "The figures displayed are effectively our computer trying to read the mind of the enemy pilot," said Graham Wallis, whose team also drew on the experience of MBDA's former air force pilots.

QUICK CALCULATIONS Not surprisingly, the computations behind GUSD are hugely complex. The main input is the realtime radar tracking data of the enemy aircraft, although other elements such as seeker acquisition data, missile dynamics and the missile's inertial navigation errors are also included. A set of algorithms (and their associated mathematical methods, see box) then process this input – with the three key algorithms covering target behaviour and identification (Hidden Markov Models); the generation of bundles of trajectories (Markov Chains and Monte Carlo techniques) and trajectory collision checking (R-Tree).

Currently at Technology Readiness Level 3, GUSD's future depends on being incorporated into a programme. The generic data used to date would be replaced by real missile and seeker data, requiring GUSD to move out of the open forum of MCM-ITP and into the area covered by defence secrecy. "Implementation of research and technology is particularly difficult when it comes to jet fighters," Graham Wallis declared. "But I'm optimistic it will get into a future upgrade of current aircraft, and could even replace the Launch Success Zone tables as a firing cue for pulling that trigger."

MATHEMATICAL MODELS Named after 19th century Russian mathematician Andrey Markov, the Markov Chain uses a set of rules to predict what will happen next in a situation, when all the variables are known. In a game of bridge, if all the hands were known to all players and they had no discretion over which card to play next, every card laid down could be predicted as a Markov Chain. With a Hidden Markov Model, a player doesn't know what an opponent holds but can infer that, for example, they no longer have any trump cards because of how they have played their last hand. The Model takes multiple states e.g. 'Opponent Has Trumps' or 'Opponent Doesn't Have Trumps' and establishes a percentage weighting of probability to each – which can then be used for a decision over the best card to play. Though it is only an analogy, GUSD's algorithms use the mass of input data to infer the probability of an enemy pilot's tactical moves. R-trees are a way of handling multi-dimensional information - in GUSD's case by using geometric models that drastically cut the computation load.

OTHER APPLICATIONS Even before its own future is decided, GUSD could already lead to a successor MCM-ITP project with Probayes in the area of mission planning for long-range missiles. Called Rapid Mission

Planning and Rehearsal, it could automate much of what is currently a labour-intensive process and to provide the customer with different options for mission routes – each with its own probability rating for success.

../../../../projets/e-motion/IMG/Guidance.jpg

Figure 9. artistic view of GUSD (copyright MBDA)

5.3. Human Centered Navigation in the physical world

5.3.1. Social Mapping

Participants: Panagiotis Papadakis, Anne Spalanzani, Christian Laugier.

With robots technology shifting towards entering human populated environments, the need for augmented perceptual robotic skills emerges that complement to human presence. In this integration, perception and adaptation to the implicit human social conventions plays a fundamental role. Toward this goal, we introduce a novel methodology to detect and analyse complex spatial interactions of multiple people and encode them in the form of a social map, whose structure is obtained by computing a latent space representation of human proxemic behaviour. We accomplish this by appointing to humans distinct, skew-normal density functions that quantify social sensitivity and by using them in the sequel to induce a training set for regressing a collective density function of social sensitivity (see fig. 10). Finally, we extract level-sets of constant social sensitivity levels within the social map by which we can effectively and efficiently analyse individual as well as shared interaction zones of varying shape and size. Extensive experiments on human interaction scenarios demonstrate the feasibility and utility of the proposed approach in diverse conditions and promote its application to social mapping of human-populated environments. This work was published at IROS [23] and submitted to RAS journal.

../../../../projets/e-motion/IMG/SocialMap.png

Figure 10. A representative example of mapping human social interactions using the proposed methodology. The regressed global sociality density along with isocontours of constant social comfort are superimposed on the corresponding scene.

5.3.2. *Goal oriented risk based navigation in dynamic uncertain environment*

Participants: Anne Spalanzani, Jorge Rios-Martinez, Arturo Escobedo-Cabello, Procopio Silveira-Stein, Gregoire Vignon, Alejandro Dizan Vasquez Govea, Christian Laugier.

Navigation in large dynamic spaces has been adressed often using deterministic representations, fast updating and reactive avoidance strategies. However, probabilistic representations are much more informative and their use in mapping and prediction methods improves the quality of obtained results. Since 2008 we have proposed a new concept to integrate a probabilistic collision risk function linking planning and navigation methods with the perception and the prediction of the dynamic environments [36]. Moving obstacles are supposed to move along typical motion patterns represented by Gaussian Processes or Growing HMM. The likelihood of the obstacles' future trajectory and the probability of occupation are used to compute the risk of collision. The proposed planning algorithm, call RiskRRT (see Figure 11 for an illustration), is a sampling-based partial planner guided by the risk of collision. Results concerning this work were published in [37] [38] [39]. [47] and [48]. In 2013, Jorge Rios defended his phD on this topic. We obtained an Inria ADT to optimize and share the RiskRRT algorithm.

../../../../projets/e-motion/IMG/grow2.png

Figure 11. Social navigation example. RiskRRT selected a plan (red line) to the goal (red arrow).

5.3.3. *Navigation Taking Advantage of Moving Agents*

Participants: Procopio Silveira-Stein, Anne Spalanzani, Christian Laugier.

In this work, we proposes a different form of robotic navigation in dynamic environments, where the robot takes advantage of the motion of pedestrians, in order to improve its own navigation capabilities. The main idea is that, instead of treating persons as dynamic obstacles that should be avoided, they should be treated as special agents with an expert knowledge of navigating in dynamic scenarios. To benefit from the motion of pedestrians, this work proposes that the robot select and follow them, so it can move along optimal paths, deviate from undetected obstacles, improve navigation in densely populated areas and increase its acceptance by other humans. To accomplish this proposition, novel approaches are developed in the area of leader selection, where two methods are explored. The first uses motion prediction approaches while the second uses a machine learning method, to evaluate the leader quality of subjects, which is trained with real examples. Finally, the leader selection methods are integrated with motion planning algorithms and experiments are conducted in order to validate the proposed techniques. One of the most relevant application is navigation among crowds. Figure 12 illustrates the concept.

../../../../projets/e-motion/IMG/LeaderFollow.png

Figure 12. Navigation among crowds. The robot (rectangle) needs to reach the right side of the corridor. 2 groups of people (one in yellow, the other in blue) are crossing in this corridor. A classical motion planning would not find a path. The robot choses a leader (represented by a green circle around a person) and follows him.

The work is published in [25] [24] and Procopio Stein defended his phD the 11th of december 2013 at the Aveiro University (phD co-directed by Anne Spalanzani and Vitor Santos).

5.3.4. *Autonomous Wheelchair for Elders Assistance*

Participants: Arturo Escobedo-Cabello, Gregoire Vignon, Anne Spalanzani, Christian Laugier.

The aging of world's population is bringing the need to provide robotic platforms capable to assist elder people to move [46]. It is necessary that such transportation is reliable, safe and comfortable. People with motor disabilities and elders are expected to benefit from new developments in the field of autonomous navigation robotics.

Autonomously driven wheelchairs are a real need for those patients who lack the strength or skills to drive a normal electric wheelchair. The services provided by this kind of robots can also be used to provide a service of comfort, assisting the user to perform difficult tasks as traversing a door, driving in a narrow corridor etc.

Simple improvements of the classical powered wheelchair can often diminish several difficulties while driving. This idea of comfort has emerged as a design goal in autonomous navigation systems, designers are becoming more aware of the importance of the user when scheming solution algorithms. This is particularly important when designing services or devices intended to assist people with some disability. In order for the robot to have a correct understanding of the intention of the user (when moving around) it is necessary to create a model of the user that takes into account his habits, type of disability and environmental information. The ongoing research project is centered in the understanding of the intentions of the user while driving an autonomous wheelchair, so that we can use this information to make this task easier.

During the last two years the work was centered in the improvement of the usability of the system. A review of the state of the art in user's intention estimation algorithms was made and a new model to infer the intentions of the user in a known environment was presented [34], [35].

The algorithm models the intention of the user as 2D topological goals in the environment. Those places are selected according to how frequently they are visited by the user (user habits). The system was designed so that the user can give orders to the wheelchair by using any type of interface, as long as he can show the direction of the intended movement (joystick, head tracking, brain control, etc). As shown in figure13 , the chosen

approach uses a Dynamic Bayesian Network to model and infer the intentions. The main contribution of this work is to model the intention of the user as topological goals instead of normal trajectory-based methods, therefore the model is simpler to deal with.
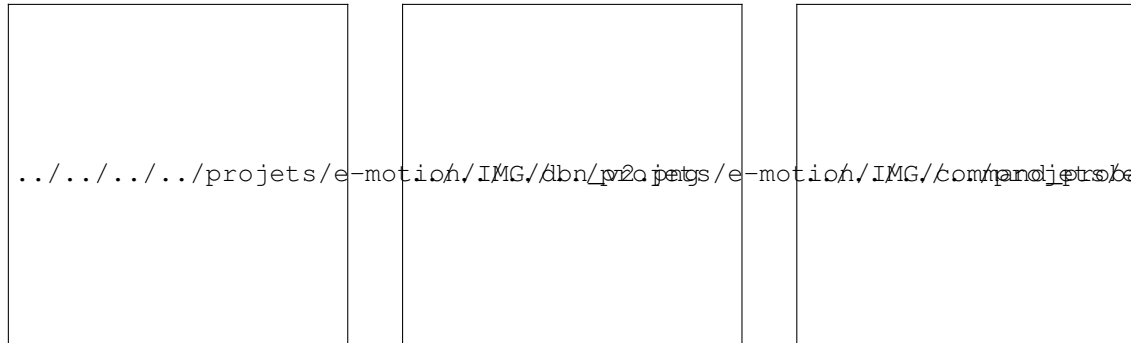
../../../../projets/e-motion/IMG/dbn_v2.png../../projets/e-motion/IMG/command_probability.png../../projets/e-motion/IMG/de

Figure 13. User's Intention Estimation Algorithm Left:User's intention model. The Bayesian network used to estimate the current user's intended; Center: The probability value for a given command C_t (big arrow) is proportional to the angle a_i formed respect to each goal g_i in the environment. Right: The user is looking to the left (in the direction of his desired goal). Once that the user's intention estimation system computes the goal with the highest probability, the autonomous navigation module plans the path and controls the movement of the wheelchair to take the user to the destination.

In 2013 the results of the user intention destination method were published in the IROS conference [16]. The presented work covered the following aspects:

User Intention Estimation: Two different methods to drive the wheelchair were compared, a semi-autonomous and a manual mode. In semi-autonomous mode the user's intention is estimated from the position of the face and the wheelchair takes care of all the planning to arrive there while avoiding obstacles. In manual mode the wheelchair is driven using the face without assistance from the robotic controller.

The Bayesian estimator shown in Fig.13 was used to do the inference of the desired destination of the user in semi-autonomous mode. The user's intention was modeled as a set of destinations commonly visited by the user and the task consisted in finding the destination targeted by the user.

Interfaces: People with motor disabilities and elders often have problems using joysticks and other standard control devices. Under this consideration our experimental platform was equipped with different types of user-interfaces to provide a multi-modal functionality as described in [35]. A face pose interface allows to control the wheelchair's motion by changing the face direction, while voice recognition interface is used to guarantee an adequate control of the wheelchair for those commands that otherwise would be difficult to give by only using the face (Stop, start, etc). This exploitation of more natural and easy-to-use human machine interfaces was one of the main contributions of the work presented in [16].

Experimental evaluation: Experiments were done in the hall of the Inria Rhône-Alpes laboratory. People in the scene were tracked to detect the most visited destinations in this setting (red circles). Those typical destinations were then placed in the map used by the robotic wheelchair. Each destination has a probability value related to the number of times that it is visited by people. The extracted typical destinations and related probabilities are used as prior knowledge when inferring the user's desired destination. In Fig.14 (Right) the spheres represent the typical destinations placed in the internal map o the wheelchair and the size of the sphere represents the probability of being the desired destination of the user in the wheelchair given it's position in the map and direction of the face (blue arrow).

To evaluate the performance of the method different persons were asked to drive the wheelchair in both "manual" and "semi-autonomous" mode. The trajectories of followed by the wheelchair were recorded and evaluated in [16], In Fig.15 one sample of those trajectories is presented where it can be appreciated how those trajectories executed with assistance of the robot are considerably softer than those obtained in manual mode as explained in [16].

Human aware navigation: Current work is being done in the construction of a social cost-map that is able to work with different open source path planning algorithms. This plug-in was developed using the method presented by Rios in [45].
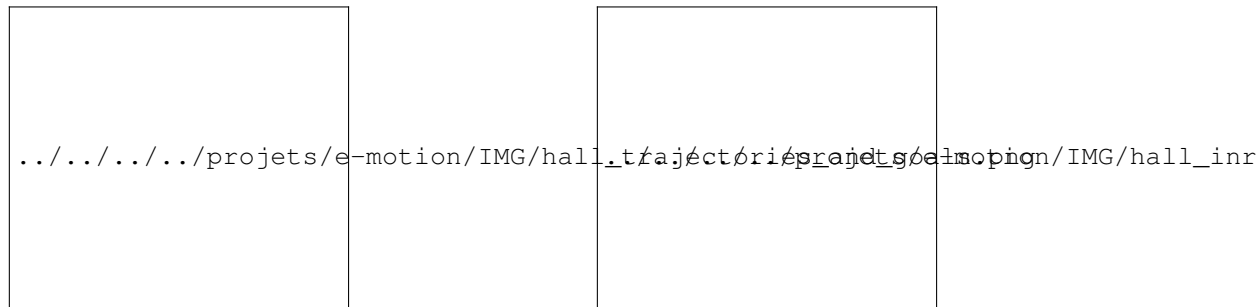

Figure 14. Left: The users of a normal environment move between typical destinations that can be learned. Right: The typical destinations marked in the map used by the wheelchair, The probability for each destination given the position of the wheelchair and direction of the command is proportional to the size of the sphere.

5.3.5. Bayesian modelling to implement and compare different theories of speech communication

Participants: Raphael Laurent, Pierre Bessière, Julien Diard, Jean-Luc Schwartz.

A central issue in speech science concerns the nature of representations and processes involved in communication. The search for phoneme or syllable specific invariants led to three major sets of approaches: motor, auditory and perceptuo-motor theories. They have been widely argued for and against, but the theoretical debate appears to be stagnating. It is our belief that computationnal models designed within a rigorous mathematical framework may allow to put forward new arguments to support either theory, and new ideas for experiments to be carried out on human subjects.

We have designed an integrative Bayesian model which allows to study auditory, motor and perceptuo-motor aspects of speech production and perception. This model was tested on perception tasks on evaluation corpora with more and more variability compared to the learning corpus. This showed a really high robustness of the purely motor model, which contained more information that it is the case in practise, due to unrealistic learning methods. The work was then focused on more realistic learning algorithms, where speech motor gestures are unsupervisedly learned through imitation, by generating motor gestures trying to reach auditory targets, and memorising the acoustics corresponding to these motor commands. This work was published in 2013 [19]. Raphael Laurent defended his phD in 2013.

5.3.6. Bayesian computing

Participants: Emmanuel Mazer, Pierre Bessière.

A book and the compagnon software on bayesian programming have been released this year :

The book "Bayesian Programming" is available at http://www.crcpress.com/product/isbn/9781439880326 and the associated software at http://www.probayes.com/fr/Bayesian-Programming-Book/

../../../../projets/e-motion/IMG/trajectories.png

Figure 15. Experimental evaluation of the user's intention method. Some samples of the resulting trajectories are presented. (a) and (c) show the results when using the assistance of the user's intention estimation system. (b) and (d) were achieved by driving the wheelchair using the face without any assistance. Here we can it can be appreciated how those trajectories executed with assistance of the robot are considerably softer than those obtained in manual mode.

Features

Presents a new modeling methodology and inference algorithms for Bayesian programming Explains how to build efficient Bayesian models Addresses controversies, historical notes, epistemological debates, and tricky technical questions in a dedicated chapter separate from the main text Encourages further research on new programming languages and specialized hardware for computing large-scale Bayesian inference problems Offers an online Python package for running and modifying the Python program examples in the book

Summary

Probability as an Alternative to Boolean Logic While logic is the mathematical foundation of rational reasoning and the fundamental principle of computing, it is restricted to problems where information is both complete and certain. However, many real-world problems, from financial investments to email filtering, are incomplete or uncertain in nature. Probability theory and Bayesian computing together provide an alternative framework to deal with incomplete and uncertain data.

Decision-Making Tools and Methods for Incomplete and Uncertain Data

Emphasizing probability as an alternative to Boolean logic, Bayesian Programming covers new methods to build probabilistic programs for real-world applications. Written by the team who designed and implemented an efficient probabilistic inference engine to interpret Bayesian programs, the book offers many Python examples that are also available on a supplementary website together with an interpreter that allows readers to experiment with this new approach to programming.

Principles and Modeling

Only requiring a basic foundation in mathematics, the first two parts of the book present a new methodology for building subjective probabilistic models. The authors introduce the principles of Bayesian programming and discuss good practices for probabilistic modeling. Numerous simple examples highlight the application of Bayesian modeling in different fields.

Formalism and Algorithms

The third part synthesizes existing work on Bayesian inference algorithms since an efficient Bayesian inference engine is needed to automate the probabilistic calculus in Bayesian programs. Many bibliographic references are included for readers who would like more details on the formalism of Bayesian programming, the main probabilistic models, general purpose algorithms for Bayesian inference, and learning problems.

FAQs Along with a glossary, the fourth part contains answers to frequently asked questions. The authors compare Bayesian programming and possibility theories, discuss the computational complexity of Bayesian inference, cover the irreducibility of incompleteness, and address the subjectivist versus objectivist epistemology of probability.

The First Steps toward a Bayesian Computer A new modeling methodology, new inference algorithms, new programming languages, and new hardware are all needed to create a complete Bayesian computing framework. Focusing on the methodology and algorithms, this book describes the first steps toward reaching that goal. It encourages readers to explore emerging areas, such as bio-inspired computing, and develop new programming languages and hardware architectures.
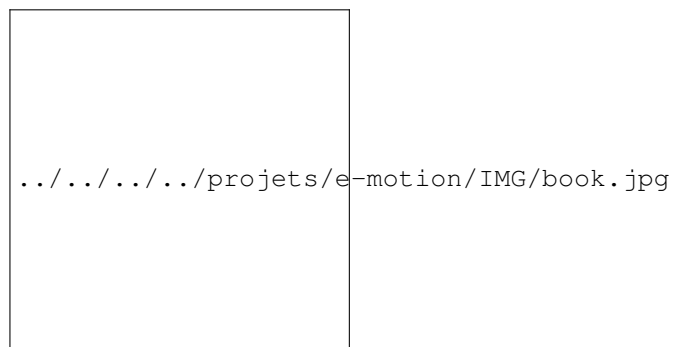
../../../../projets/e-motion/IMG/book.jpg

Figure 16. Bayesian Programming book covert

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

6. New Results

6.1. Ontology matching and alignments

We pursue our work on ontology matching and alignment support [5], [12] with contributions to evaluation and alignment semantics.

6.1.1. *Evaluation*

Participant: Jérôme Euzenat.

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 [2].

This year, we ran the OAEI 2013 evaluation campaign [7]. It offered 8 different test sets (7 of which under the SEALS platform). This issue brought the following results:

- Once again, more participants than ever (23);
- Most ontology matchers running on the SEALS platform (20);
- Increased performances in terms of precision and recall;
- Matchers are now faster and more scalable. There are also more matchers using networked resources.

We used again the our generator for generating new version of benchmarks [4]. The Alignment API was used for manipulating alignments and evaluating results.

A novelty of this year was the evaluation of interactive systems, included in the SEALS client. It brings interesting insight on the performances of such systems and should certainly be continued.

The participating systems and evaluation results were presented in the 8th Ontology Matching workshop, that was held in Sydney, Australia [13]. More information on OAEI can be found at http://oaei.ontologymatching.org/.

6.1.2. *Algebras of relations in alignments*

Participants: Armen Inants [Correspondent], Jérôme Euzenat.

We had previously shown that algebras of relations between concepts can be used for expressing relations in alignments. We have worked this year as extending them in two ways.

We increased the expressiveness of relations between concepts, not restricting the algebra to necessarily non empty concepts. This describes all taxonomical (as opposed to mereological) relation algebras, i.e., all those relations that have been used by matchers so far.

We also dealt with relations among different kinds of entities – individuals or concepts. For this, relation algebra structures are considered in an arbitrary one- or many-sorted logical theory. We established a sufficient condition for a set of dyadic formulas in a first-order theory to generate a relation algebra. This result is extended to many-sorted theories by means of Schröder categories.

This work is part of the PhD of Armen Inants.

6.2. 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 critical 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.

6.2.1. Interlinking cross-lingual RDF data sets

Participants: Tatiana Lesnikova [Correspondent], Jérôme David, Jérôme Euzenat.

Data interlinking is a difficult task in a cross-lingual environment like the Web. Even systems based on graph structure, ultimately rely on anchors based on language fragments. If languages are different, fragments have to be compared by more sophisticated techniques. In that context, we are developing an approach which represents RDF entities as (virtual) text documents and compare them using different strategies [9], [10]. We investigate two directions: (1) a translation-based approach where the virtual documents are automatically translated; (2) a language-independent approach where important terms found in documents are mapped to a terminological resource like Wordnet to compute document similarity.

This work is part of the PhD of Tatiana Lesnikova developed in the LINDICLE project (see §7.1.2).

6.2.2. Data interlinking from expressive alignments

Participants: Zhengjie Fan [Correspondent], Jérôme Euzenat.

In the context of the DATALIFT project, we are further developing the data interlinking module. We have developed an algorithm able to determine potential attribute correspondences of two classes depending on their features. For that purpose, we use k-means or k-medoids clustering. These correspondences are then used to construct a SILK script which generates an initial link set. Some of the links are presented to the user who assesses their validity. We then use an improvement of the disjunctive version space supervised learning method to learn a better script from the assessed links. Such a technique can be iterated until satisfactory links are found.

This work is part of the PhD of Zhengjie Fan, co-supervised with François Scharffe (LIRMM), and developed in the DATALIFT project (see §7.1.1).

6.2.3. Key and pseudo-key detection for web data set interlinking

Participants: Jérôme David [Correspondent], Manuel Atencia Arcas, Anthony Delaby, Jérôme Euzenat.

Keys are sets of properties which uniquely identify individuals (instances of a class). We have refined the notion of database keys in a way which is more adapted to the context of description logics and the openness of the semantic web. We have also refined the weaker notion of a linkkey introduced in [12]. Then we have shown how such keys, together with ontology alignments, and linkkeys may be used for deducing equality statements (links) between individuals across data sources in the web of data.

However, ontologies do not necessarily come with key descriptions, and never with linkkey assertions (which would hold across ontologies). But, these can be extracted from data by assuming that keys holding for specific data sets, may hold universally. We have extended these classical key extraction techniques for extracting linkkeys.

This work is developed partly in the LINDICLE and DATALIFT projects. A proof of concept implementation is available at http://rdfpkeys.inrialpes.fr/.

6.3. Ontology networks

Dealing with the semantic web, we are interested in ontology networks, i.e., sets of distributed ontologies that have to work together. One way for these systems to interact consists of exchanging queries and answers. For that reason, we pay particular attention to query systems.

6.3.1. Path queries and μ-calculus

Participant: Jérôme Euzenat.

Querying the semantic web is mainly done through the SPARQL language or its extensions through paths and entailment regimes [14]. Query containment is the problem of deciding if the answers to a query are included in those of another query for any queried data sources. This problem is very important for query optimisation purposes. In the SPARQL context, it can be equally useful for distributing federated queries or for implementing schema-based access control. In order to experimentally assess implementation strengths and limitations, we provided a first SPARQL containment test benchmark. We studied the query demographics on DBPEDIA logs to design benchmarks for relevant query containment solvers. We tested available solvers on their domain of applicability on three different benchmark suites [6] and found that (i) tested solutions are overall functionally correct, (ii) in spite of its complexity, SPARQL query containment is practicable for acyclic queries, (iii) state-of-the-art solvers are at an early stage both in terms of capabilities and implementation.

This work has been developed in collaboration with the TYREX team and within the PhD thesis of Melisachew Wudage Chekol now in the ORPAILLEUR team. The benchmarks, results and software are available at http://sparql-qc-bench.inrialpes.fr.

<p style="text-align:center">**IBIS Project-Team**</p>

5. New Results

5.1. Analysis of gene regulatory networks by means of piecewise-linear (PL) models

GENETIC NETWORK ANALYZER (GNA) is a tool for the qualitative modeling and simulation of the dynamics of gene regulatory networks by means of PLDE models, as described in Section 4.1 . GNA has been integrated with the other bioinformatics tools distributed by Genostar (http://www.genostar.com/). Version 8.5 of GNA was released by IBIS and Genostar this year. This version is an update of version 8.0, deposited at the Agence pour la Protection des Programmes (APP). Some bugs have been corrected in the new version and the program has been adapted to the latest versions of Java and the software platform of Genostar. Version 8.5 supports the SBML standard and is also capable of exporting its models to the newly-developed standard for qualitative models, SBML Qual. This standard has been elaborated by the community of developers of logical and related modeling tools (CoLoMoTo), in which the GNA developers participate, and has been described in detail in a paper published in *BMC Systems Biology* [6].

The predictions obtained with the help of GNA are purely qualitative, describing the dynamics of the network by means of a state transition graph. While a qualitative analysis is appropriate for certain problems, the absence of precise quantitative predictions may not be desirable in others, such as the analysis of a limit cycle or the design of a controller for a synthetic network. The quantitative study of PLDE models of gene regulatory networks is hindered by the fact that the step functions describing the logic of regulatory interactions lead to discontinuities in the right-hand side of the PLDE models (Section 3.1). This has motivated extensions of the PLDE models based on differential inclusions and Filippov solutions. As of now, no numerical simulation tool for the simulation of these Filippov extensions is available.

In collaboration with Vincent Acary and Bernard Brogliato of the BIPOP project-team, we have shown how tools developed for the simulation of nonsmooth mechanical, electrical and control systems can be adapted for this purpose, in a paper published in *Physica D* [2]. We have presented a method for the numerical analysis of one proposed extension, called Aizerman–Pyatnitskii (AP)-extension, by reformulating the PLDE models as mixed complementarity systems (MCSs). This allows the application of powerful methods developed for this class of nonsmooth dynamical systems, in particular those implemented in the SICONOS PLATFORM developed by BIPOP. We have also shown that under a set of reasonable biological assumptions, putting constraints on the right-hand side of the PLDE models, AP-extensions and classical Filippov extensions are equivalent. This means that the proposed numerical method is valid for a range of different solution concepts. We have illustrated the practical interest of our approach through the numerical analysis of three well-known networks developed in the field of synthetic biology.

5.2. 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 4.2), to allow biologists to make the most out of the information contained in reporter gene expression data. A web-based version of WELLREADER is currently in preparation. Valentin Zulkower has analyzed the measurement models underlying WELLREADER, work was presented at the *Journées Ouvertes Biologie, Informatique et Mathématiques (JOBIM'13)* [13] and submitted for publication.

The above tools have been used in a series of studies directed at the experimental mapping of gene regulatory networks in *E. coli*. A first example, which was carried out in the framework of the PhD thesis of former IBIS member Guillaume Baptist, concerns the development of a new screening methodology for identifying all genes that control the expression of a target gene through genetic or metabolic interactions. The screen combines mutant libraries with luciferase reporter constructs. Instead of a static picture of gene expression, this method allows dynamical monitoring in different environmental conditions. Mutants with interesting phenotypes can thus be selected based on multiple criteria, and the expression dynamics of the target gene can be extensively characterized. The method has been applied to the identification of the direct and indirect regulators of the gene *acs* in *Escherichia coli*. We confirmed known genetic regulators of the gene and identified new regulatory influences, many of which involve metabolic intermediates or metabolic sensing. An analysis of mutants involved in glycolysis and glucose transport demonstrates that the classical model of catabolite repression in *E. coli* needs to be amended. A paper describing the above work was published in *Nucleic Acids Research* this year [3].

A second example is a study, led by Stéphan Lacour in collaboration with Akira Ishihama and Hiroshi Ogasawara in Japan, on the lifestyle adaptation of *E. coli*. The study concerns the switch between swimming motility and biofilm formation in response to changes in environmental growth conditions. The stationary phase sigma factor RpoS is an important regulator of this switch since it stimulates adhesion and represses flagellar biosynthesis. By measuring the dynamics of gene expression, we show that RpoS inhibits the transcription of the flagellar sigma factor, FliA, in exponential growth phase. RpoS also partially controls the expression of CsgD and CpxR, two transcription factors important for bacterial adhesion. We have demonstrated that these two regulators repress the transcription of *fliA*, *flgM* and *tar*, and that this regulation is dependent on the growth medium. CsgD binds to the flgM and fliA promoters around their -10 promoter element, strongly suggesting direct repression. The results show that CsgD and CpxR also affect the expression of other known modulators of cell motility. An updated structure of the regulatory network controlling the choice between adhesion and motility was proposed in the paper based on this work, published in the *Journal of Bacteriology* [7].

A third study, published in *Research in Microbiology* [8], also focuses on the alternative sigma factor RpoS. The small protein Crl increases the interaction between RpoS and RNA polymerase and thereby activates certain RpoS-dependent promoters. However, the growth-phase dependence of the interaction of Crl with different forms of polymerase remains unknown. We have used 41 GFP transcriptional fusions to study the dynamics of gene regulation by RpoS and Crl during growth transition from exponential to stationary phase in *Escherichia coli*. This has confirmed that RpoS can regulate gene expression in exponential phase, both positively and negatively. Crl slightly stimulates transcription by RpoS in exponential phase and controls a subset of RpoS-dependent genes in stationary phase. Growth temperature strongly affects induction of specific promoters by RpoS, whereas its impact on gene regulation by Crl is much less significant. In addition, we have identified five new genes regulated by Crl (*ada*, *cbpA*, *glgS*, *sodC* and *flgM*), and demonstrated that Crl improves promoter binding and opening by RpoS-containing RNA polymerase at the *hdeA* promoter. The study also shows that Crl is a cognate enhancer of RpoS activity under different growth conditions, since its deletion has no effect on genes transcribed by other sigma factors.

In the framework of the PhD thesis of Diana Stefan, a network inference method developed by Eugenio Cinquemani and colleagues, first published in *Bioinformatics* in 2010, has been applied to reporter gene data from the network regulating motility of *E. coli*, described above. The results are currently being prepared for publication.

5.3. Models of carbon metabolism in bacteria

Kinetic models capture the dynamics of the large and complex networks of biochemical reactions that endow bacteria with the capacity to adapt their functioning to changes in the environment. In collaboration with Matteo Brilli and Daniel Kahn (Inra and Université Claude Bernard in Lyon), we previously developed an approximate model of central metabolism of *E. coli*, using linlog kinetics, and estimated the parameter values from metabolomics, transcriptome, proteomics data sets, as described in an article published in *Bioinformatics*

in 2011. The results of this study revealed the fundamental role played by the identifiability of the model parameters, an issue often overlooked in systems biology. This prompted us for a thorough investigation of the concepts of structural identifiability (in presence of perfect, idealized data), practical identifiability (in presence of noisy and limited amounts of data), and the relations between the two. In addition, we looked into the implications of this analysis for the reduction of nonidentifable to identifiable models. While having a solid mathematical basis, the study was tailored to the actual experimental practice, and resulted in a practical model reduction method that improves upon our previous approach in case of large measurement noise. This study, and the results from its application to both *in-silico* case studies and state-of-the-art datasets, were reported in a paper that appeared in the *Journal of Mathematical Biology* this year [4]. Although the theoretical development has focused on linlog models and related classes of approximate kinetic models, it is important to note that the results also bear on more general classes of nonlinear models of metabolism.

A second line of work is based on the use of classical kinetic models that are, in comparison with the above-mentioned linlog models, much reduced in scope (the focus is on the metabolic and genetic regulation of the glycolysis pathway) and granularity (individual reactions are lumped together). The models, developed by Delphine Ropers, are being calibrated using experimental data from IBIS group and the group of Jean-Charles Portais at Inra/INSA in Toulouse, and will be used to understand some key mechanisms in the adaptation of *E. coli* to the exhaustion of glucose. The PhD thesis of Manon Morin, in the framework of a collaboration supported by a Contrat Jeune Scientifique Inra-Inria, will further develop these research directions. In the framework of their PhD theses, Stéphane Pinhal and Valentin Zulkower also study specific aspects of carbon metabolism, using both models and experimental data. In parallel, we collaborate with Myriam Ferro at CEA in Grenoble to investigate how state-of-the-art measurements of the absolute concentrations of enzymes in *E. coli* can be integrated with other high-throughput data sets and kinetic models. The first results of this collaboration were accepted for publication in *Molecular and Cellular Proteomics* early 2014.

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

At the single-cell level, the processes that govern gene expression are often 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 control of cell populations and even of single cells. General modeling paradigms, such as the Chemical Master Equation, exist for the description of stochastic dynamics at the single-cell level. However, due to the complexity of the interactions, current studies have often preferred to focus on specific cases of interest by *ad-hoc* modeling and analysis. In addition, theoretical and practical challenges inherent in the inference of stochastic models from biological experimental data have limited the development of general identification approaches.

Work in IBIS on the probabilistic modeling of gene expression and interaction dynamics at the level of individual cells is centered around two main challenges. On the one hand, we address identification from microscopy data and analysis of the arabinose uptake dynamics in *E. coli* upon glucose exhaustion. Starting from a reduced arabinose uptake model, Eugenio Cinquemani and Michel Page are working on methods for the estimation of unknown stochastic model parameters from statistical population snapshot data collected by fluorescence microscopy experiments. Analysis of the model focuses on the problem of model-based real-time single-cell state estimation, with feedback control applications in mind, in collaboration with Alfonso Carta (BIOCORE). Based on a stochastic model reflecting switch-like dynamics in the form of sigmoidal reaction rates, taking a Chemical Master Equation model with cell-dependent parameters as a gold standard desdcription of the system, a Chemical Langevin Equation approximation is proposed as a convenient approximation of the model for observer design purposes. On top of this model approximation, a so-called Square-Root Unscented Kalman filter (SRUKF) is designed. Based on simulations of a realistically tuned model, SRUKF is found to perform as good as much heavier particle filters based on the gold-standard model.

Results were presented at the *European Control Conference (ECC)* in 2013 [11], where we also showed that including extrinsic noise effects explicitly in the estimation process allows one to improve the knowledge of the hidden states.

On the other hand, we investigate the use mixed effects-modelling and identification techniques to characterize single-cell profiles 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. In collaboration with Gregory Batt (CONTRAINTES) and Giancarlo Ferrari-Trecate (University of Pavia, Italy), we are adapting and applying existing procedures from pharmacokinetics to the context of microfluidic data, with focus on the budding yeast response to osmolarity shocks. The first results of the work were presented at the *European Control Conference (ECC)* this year [12]. Rigorous model identification and validation steps are performed on data from real-time control experiments performed in Pascal Hersen's lab at Université Paris Descartes, for both mixed-effects modelling and for the competing method of moment-based identification. Results show the tendency of mixed-effects modelling to avoid overfitting for this system, trading fitting performance for validation performance and hence predictive capabilities. The work is being further developed and the collaboration tightened by the ongoing visit of Andres Gonzalez, PhD candidate at the University of Pavia, to CONTRAINTES and IBIS. A first journal publication is in preparation, which will be followed by extensions and refinements of the method.

In parallel, work concerning the study of noise propagation in gene regulatory networks is carried out in collaboration with Irina Mihalcescu (Université Joseph Fourier). Finally, collaboration of Eugenio Cinquemani with Marianna Rapsomaniki, PhD student affiliated with Zoi Lygerou (University of Patras, Greece) and John Lygeros (ETH Zürich, Switzerland), has been devoted to the analysis of data from Fluorescence Recovery After Photobleaching (FRAP) experiments. It has given rise to a novel method for reconstructing physical diffusion and immobilization parameters at the level of single cells. The method has been applied to nuclear species of mammalian cells and results are part of a journal paper under revision.

5.5. Shared control of gene expression by global physiological effects and specific regulators

Gene expression is controlled by the joint effect of (1) the global physiological state of the cell, in particular the activity of the gene expression machinery, and (2) DNA-binding transcription factors and other specific regulators. While many studies have focused on networks of transcription factors, the analysis of the relative contributions of both transcription factors and global effects of the physiological state has received relatively little attention thus far.

In the framework of the PhD thesis of former IBIS member Sara Berthoumieux, we have developed a model-based approach to distinguish between these two effects using time-resolved measurements of promoter activities. We have demonstrated the strength of the approach by analyzing a circuit involved in the regulation of carbon metabolism in *E. coli*, consisting of two pleiotropic regulators of the cell (Crp and Fis), the gene *acs* encoding the enzyme acetyl-CoA synthetase (Acs), and the signaling metabolite cyclic AMP (cAMP) which activates Crp. *acs* is strongly expressed in the absence of glucose and is thus an excellent indicator of the transcriptional response of carbon metabolism to a growth-phase transition.

Our results show that the transcriptional response of the network is controlled by the physiological state of the cell and the signalling metabolite cAMP. The (surprising) absence of a strong regulatory effect of transcription factors suggests that they are not the main coordinators of gene expression changes during growth transitions, but rather that they complement the effect of global physiological control mechanisms. This change of perspective has important consequences for the interpretation of transcriptome data and the design of biological networks in biotechnology and synthetic biology. An article presenting the above results was published in *Molecular Systems Biology* this year [5] and selected as an Editor's choice in *Science* (http://ibis.inrialpes.fr/article1040.html).

In the above-mentioned work, the activity of the gene expression machinery was indirectly measured, by monitoring the activity of a constitutive gene, that is, a gene whose expression does not depend on any specific regulators but only on the activity of the gene expression machinery. There exists a huge literature on the molecular mechanisms coupling the activity of the gene expression machinery to changes in the nutritional quality of the environment, but a quantitative and dynamic picture of this very complicated regulatory system is still missing. Delphine Ropers and Edith Grac as well as Nils Giordano are developing models to achieve this, from bottom-up and top-down perspectives, respectively.

5.6. Control of regulatory networks in bacteria

A bacterial cell adapts its growth rate to the environment, notably to the availability of nutrients providing the molecular building blocks and the energy required for growth. Upon a change in the environment, the global physiology of the cell is adjusted in parallel with the adaptation of the growth rate. In the context of the PhD thesis of former IBIS member Jérôme Izard, we have studied the relation between the gene expression machinery, the global physiology of the cell, and the growth rate from a different perspective. Our aim was to change the mechanisms regulating the activity of the gene expression machinery in such a way so as to be able to externally control the growth rate of the cell.

More precisely, we have engineered an *E. coli* strain in which the transcription of an essential component of the global gene expression machinery, RNA polymerase, is under the tight control of an inducible promoter. By adjusting the inducer concentration in the medium we can adjust the RNA polymerase concentration and thereby reversibly tune the growth rate of the bacterium between zero and the maximal growth rate. The growth arrest is completely reversed when RNA polymerase is provided again. The analysis of the transcriptome at growth rates restricted by the concentration of RNA polymerase confirms that the concentration of RNA polymerase is the major determinant of changes in gene expression patterns. Our modified *E. coli* strain provides a novel way of setting growth rate in a tunable, reversible, modular, and medium-independent way. The strain, described in a paper submitted for publication, opens new perspectives for studying the mechanisms of growth control as well as for developing biotechnological applications, the subject of the post-doctoral fellowship of Cindy Gomez Balderas-Barillot. We have submitted a patent proposing such applications, which underlies the technology transfer activities undertaken in the recently-started Reset project (Section 7.2).

<p style="text-align:center">**IMAGINE Project-Team**</p>

5. New Results

5.1. Introduction

We are developing user-centred, knowledge-based models in three main domains: shape, motion and narrative design, leading us to three research axes. The fourth one is the combination of these models with intuitive interaction tools, in order to set up interactive creative environments dedicated to specific categories of content. The following sections describe our activities in 2013 for each axis.

5.2. High level model for shapes

- **Scientist in charge**: Stefanie Hahmann
- **Other permanent researchers**: Marie-Paule Cani, Jean-Claude Léon.

5.2.1. Implicit modeling

Participants: Antoine Bégault, Adrien Bernhardt, Marie-Paule Cani, Mohamed-Galal Koraa, Cédric Zanni.

Implicit surfaces are an appealing representation for free-form, volumetric shapes. In addition to being able to represent shapes of arbitrary topological genius, they have the ability to be constructed by successively blending different components, which eases interactive modeling.

In collaboration with Loic Barthe in Toulouse, we contributed to a new binary blending operator, called Gradient Blending [7], which enables us to blend implicit shapes not only in function of their field values but also of their gradients. This solves a number long standing problems in implicit modeling: we can generate bulge-free blending, ensure that the topological genius of the blended shape remains the one of the union of the input one, and avoid the blur of small details.

Figure 4. Example of surface generated using our SCALIS approach.

Within Cédric Zanni's PhD [2] we introduced closed-form solutions for convolution surfaces along helical skeletons and extended Gabor-noise texturing to enable the creation of repetitive geometric details along implicit surfaces. We also developed a novel extension to convolution surfaces, so-called SCALe-invariant Integral Surfaces (SCALIS) [15], see Figure 4 . Thanks to their scale invariant blending properties, these surfaces have three main advantages: the radius of the surface around a skeleton can be explicitly controlled, shapes are self-similar regardless of the scale of the model, and thin components are not smoothed-out when blended into larger ones. This is done while preserving the main benefits of integral surfaces, namely n-ary blending with a simple plus, and shape preservation whatever the way the skeletons is spit into smaller primitives. We are currently extending this work to enable the topology of the implicit surface to always reflect the one of the skeleton

5.2.2. *Analysis of CAD models*

Participants: Francois Faure, Stefanie Hahmann, Jean-Claude Léon, Olivier Palombi, Flavien Boussuge, Ahmad Shahwan.

CAD models, as part of assemblies defining manufactured products, are often shaped in accordance with their physical counterpart. However, one can observe that the shape of some components, as modeled in CAD, may differ from that of their physical instance. In addition, assemblies representing products are most often reduced to a collection of CAD models representing each component and the designation of each component is neither a reliable information nor a faithful connection with one or more functions of a component. As a result, geometric interfaces between components are unknown and they cannot be reduced to contacts. Interferences may exist that are also relevant for several applications. Determining precisely, the geometric interfaces between components is a first requirement to enrich geometric models with functional information because a subset of functions derive from interfaces between components.

As an example, this is particularly useful for structural mechanics to be able to generate rapidly a Finite Element model of assemblies and it is especially critical when assemblies get very complex. [9] addresses the problem to generate automatically a class of geometric interfaces for very complex assemblies (see fig. 5). GPU-based algorithms have proved suitable to obtain reliable results on CAD models.

Using these geometric interfaces as well as the newly introduced concept of conventional interfaces, [6], [4] and [11], [12], [19] have proposed an approach using qualitative reasoning, ontology reasoning to connect CAD components, their geometric interfaces, to functions and functional designations of components: an intrinsic identifier of a component in an assembly that connects it to its function. As a result, it is shown how geometric models of components need to be restructured, which extends the concept of annotation presently reduced to a elementary link between geometric models and symbolic information.

At the level of assembly components, shape analysis [30] is particularly useful to generate dimensionally reduced models needed for structural mechanics. [3] shows that analyzing a B-Rep CAD model to derive a construction graph, i.e. a set of construction trees, can be a robust basis to generate dimensionally reduced models [18], [32].

5.2.3. *Knowledge-based shape transfert*

Participants: Marie-Paule Cani, Ali Dicko, Francois Faure, Olivier Palombi.

Characters with precise internal anatomy are important in film and visual effects, as well as in medical applications. We have proposed the first semi-automatic method for creating anatomical structures, such as bones, muscles, viscera and fat tissues [5], as illustrated in 6 . This is done by transferring a reference anatomical model from an input template to an arbitrary target character, only defined by its boundary representation (skin). The fat distribution of the target character needs to be specified. We can either infer this information from MRI data, or allow the users to express their creative intent through a new editing tool. The rest of our method runs automatically: it first transfers the bones to the target character, while maintaining their structure as much as possible. The bone layer, along with the target skin eroded using the fat thickness information, are then used to define a volume where we map the internal anatomy of the source model using harmonic (Laplacian) deformation. This way, we are able to quickly generate anatomical models for a large range of target characters, while maintaining anatomical constraints.

../../../../projets/imagine/IMG/Hahmann-Raweb2013.png

Figure 5. Computation of assembly interfaces in NURBS representation.

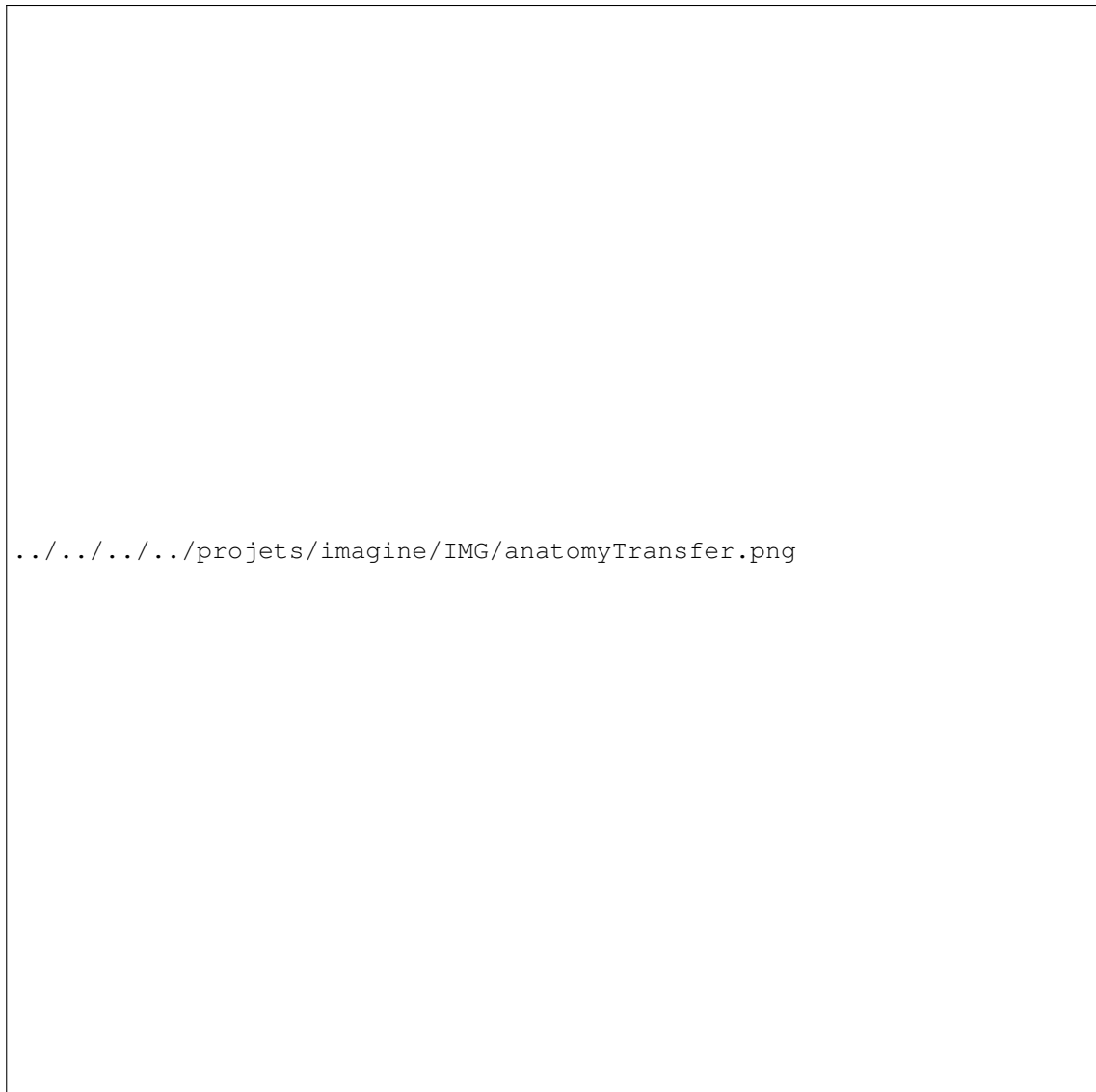../../../../projets/imagine/IMG/anatomyTransfer.png

Figure 6. A reference anatomy (left) is automatically transferred to arbitrary humanoid characters. This is achieved by combining interpolated skin correspondences with anatomical rules.

5.3. Models for motion and animation

- **Scientist in charge**: François Faure

- **Other permanent researchers**: Marie-Paule Cani, Damien Rohmer, Rémi Ronfard.

5.3.1. Physical models
Participants: Marie-Paule Cani, François Faure, Pierre-Luc Manteaux.

Frame-based deformable solids Our frame-based deformable model was published as a book chapter [31]. It combines the realism of physically based continuum mechanics models and the usability of frame-based skinning methods, allowing the interactive simulation of objects with heterogeneous material properties and complex geometries. The degrees of freedom are coordinate frames. In contrast with traditional skinning, frame positions are not scripted but move in reaction to internal body forces. The deformation gradient and its derivatives are computed at each sample point of a deformed object and used in the equations of Lagrangian mechanics to achieve physical realism. We introduce novel material-aware shape functions in place of the traditional radial basis functions used in meshless frameworks, allowing coarse deformation functions to efficiently resolve non-uniform stiffnesses. Complex models can thus be simulated at high frame rates using a small number of control nodes.

../../../../projets/imagine/IMG/ReposSPHClassique.jpgts/imagine/IMG/ReposSPHARP

Figure 7. A dam break simulation with 5000 particles simulated with WCSPH (on the left) and with our adaptive method (on the right). On the right image, blue corresponds to full-dynamics particles, green to transition particles and red to restrained particles.

Adaptive particle simulation In collaboration with the NANO-D Inria Team, we have explored the use of Adaptively Restrained (AR) particles for graphics simulations [25]. Contrary to previous methods, Adaptively Restrained Particle Simulations (ARPS) do not adapt time or space sampling, but rather switch the positional degrees of freedom of particles on and off, while letting their momenta evolve. Therefore, inter-particles forces do not have to be updated at each time step, in contrast with traditional methods that spend a lot of time there. We first adapted ARPS to particle-based fluid simulations, as illustrated in 7 and proposed an efficient incremental algorithm to update forces and scalar fields. We then introduced a new implicit integration scheme enabling to use ARPS for cloth simulation as well. Our experiments showed that this new, simple strategy for adaptive simulations can provide significant speedups more easily than traditional adaptive models.

5.3.2. Skinning virtual characters
Participants: Marie-Paule Cani, Damien Rohmer.

Skinning is a widely used technique to deform articulated virtual characters. It can be computed fastly and therefore can deliver real-time feedback at the opposite of physically based simulation. Still standard skinning approaches cannot handle well large deformations and may require manual corrections.

In collaboration with Loic Barthe and Rodolphe Vaillant from University of Toulouse, and collaborators from Victoria University, Inria Bordeaux and University of Bath, we develop a new automatic correction for skinning deformation that has been published in SIGGRAPH [14]. Based on the volumetric implicit representation paradigm, it adjust the mesh vertices and improves the visual appearance of the deformed surface. Moreover, it seamlessly handle skin contact ensuring that no self collision can occurs as seen in fig. 8 . Finally, the method can mimic muscular bulges controled by the implicit blending operators described in the work [7].

../../../../projets/imagine/IMG/skinning.png

Figure 8. Example of large skinning deformation obtained in [14].

5.3.3. Animating crowds

Participants: Marie-Paule Cani, Quentin Galvane, Kevin Jordao, Kim Lim.

Crowd animation is an interesting case, since it can be either computed by developing artificial intelligence methods, by using physically-based simulation of some extended particle systems, or by applying a kinematic

texturing methodology, made possible by the repetitive nature of crowd animations. We launched this new topic in the group in 2013, enabling us to explore the two last crowd animation methods:

Firstly, in collaboration with the University SAINTS, Malaysia, we extended particle-based crowd simulation to the case when 4 different populations, with different goals and behaviors, are interacting within the same environment [24]. This as illustrated by a cultural heritage application, with the reconstruction of past life in a harbor in Malaysia in the 19th century: see Figure 9 .

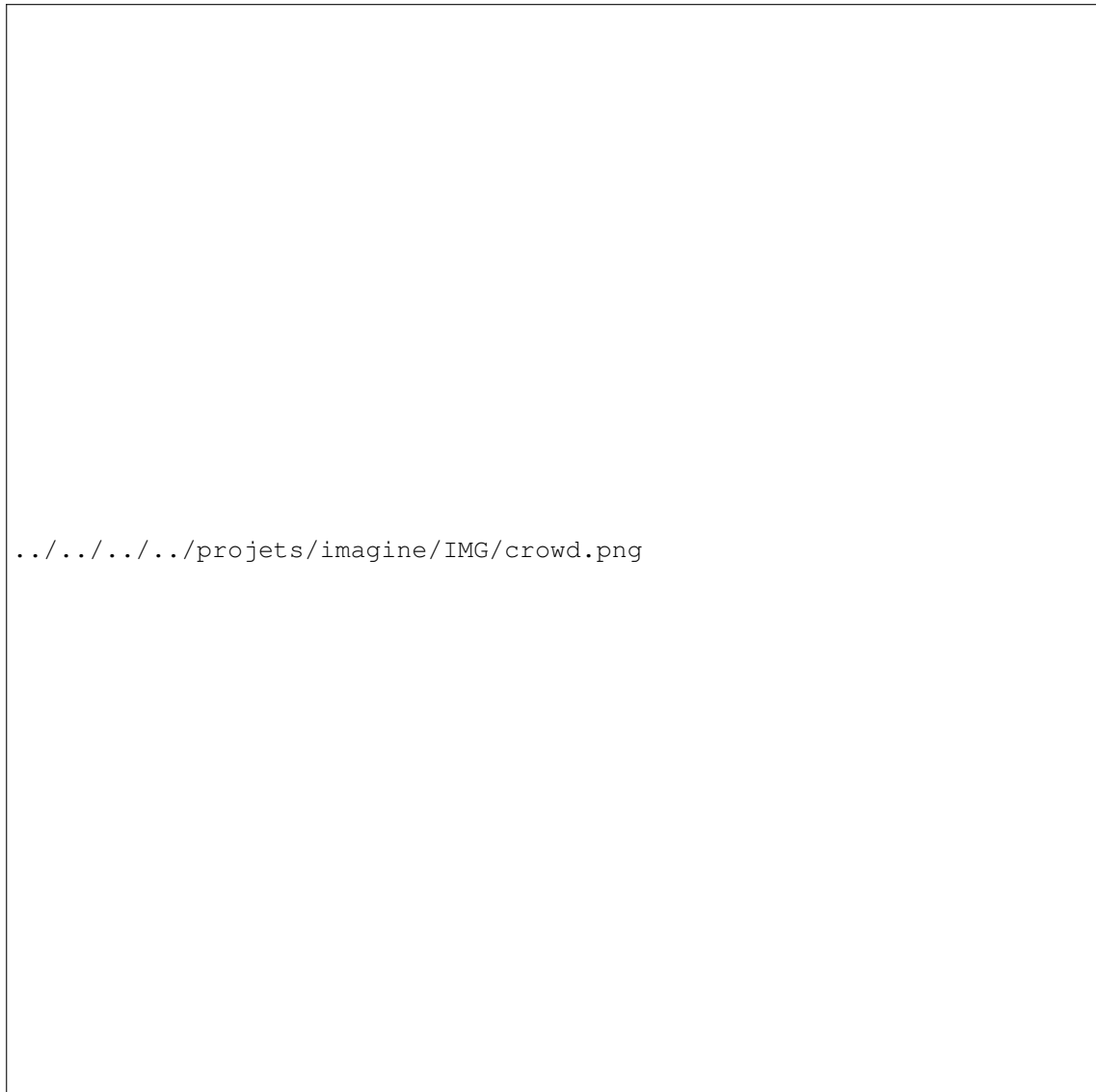../../../../projets/imagine/IMG/crowd.png

Figure 9. Crowd simulation with 4 different populations from [24].

Secondly, within the ANR project CHROME with Inria Rennes, we adopted the crowd-patches technique, i.e. the idea of combining patches carrying pre-computed crowd trajectories, for quickly populating very large

environments [23]. We are currently developing novel methods for enabling the interactive space-time editing of these animations (a paper will be published at the next Eurographics conference).

5.4. Knowledge-based models for narrative design

- **Scientist in charge**: Rémi Ronfard
- **Other permanent researchers**: Marie-Paule Cani, François Faure, Jean-Claude Léon, Olivier Palombi

5.4.1. Cinematographic virtual camera control

Participants: Marie-Paule Cani, Quentin Galvane, Vineet Gandhi, Chen Kim Lim, Rémi Ronfard.

Steering Behaviors for Autonomous Cameras [21] : We proposed a new method for automatically filming crowd simulations with autonomous cameras, using specialized camera steering behaviors and forces. Experimental results show that the method provides a good coverage of events in moderately complex crowds simulations, with consistently correct image composition and event visibility.

The prose storyboard language [26] : We presented a formal language for describing movies shot by shot, where each shot is described with a unique sentence. The language uses a simple syntax and limited vocabulary borrowed from working practices in traditional movie-making, and is intended to be readable both by machines and humans. The language is designed to serve as a high-level user interface for intelligent cinematography and editing systems.

5.4.2. Virtual actors

Participants: Adela Barbulescu, Rémi Ronfard.

Audio-Visual Speaker Conversion using Prosody Features [17] : We presented a new approach towards speaker identity conversion using speech signals and 3D facial expressions. Audio prosodic features are extracted from time alignment information for a better conversion of speaking styles. A subjective evaluation was performed to illustrate that the converted sequences are perceived as belonging to the target speakers. We are working to extend that approach to visual prosody features and to apply it to the situation where a director controls the expressions of a virtual actor, while maintaining its personality traits.

5.4.3. Narrative analysis of video

Participants: Vineet Gandhi, Rémi Ronfard.

Naming and detecting actors in movies [22]: We proposed a generative model for localizing and naming actors in long video sequences. More specifically, the actor's head and shoulders are each represented as a constellation of optional color regions. Detection can proceed despite changes in view-point and partial occlusions. This work is being extended to the case of theatre actors during performances and rehearsals. It also opens the way to future work in automatic analysis of cinematographic and editing styles in real movie scenes. This was also presented as a poster at the International Conference on Computational Photography (ICCP).

Recording theatre rehearsals [29] : We presented a contribution to the International Federation for Theatre Research describing our ongoing collaboration with the Theatre des Celestins in Lyon, emphasising that high quality vídeo recordings make it possible to study the genetic evolution of a theatre performance, and make it an object of scientific study as well as an object of aesthestic appreciation.

5.5. Creating and interacting with virtual prototypes

- **Scientist in charge**: Jean-Claude Léon
- **Other permanent researchers**: Marie-Paule Cani, Olivier Palombi, Damien Rohmer, Rémi Ronfard.

5.5.1. Sketch-based modeling

Participants: Marie-Paule Cani, Martin Guay, Rémi Ronfard.

The Line of Action [8] : The line of action is a conceptual tool often used by cartoonists and illustrators to help make their figures more consistent and more dramatic. In this paper, we proposed a mathematical definition of the line of action (LOA), which allows us to automatically align a 3D virtual character to a user-specified LOA by solving an optimization problem. This woek is now being extended to the more challenging case of creating complete animations from storyboard-like hand-drawn sketches (see fig. 10).

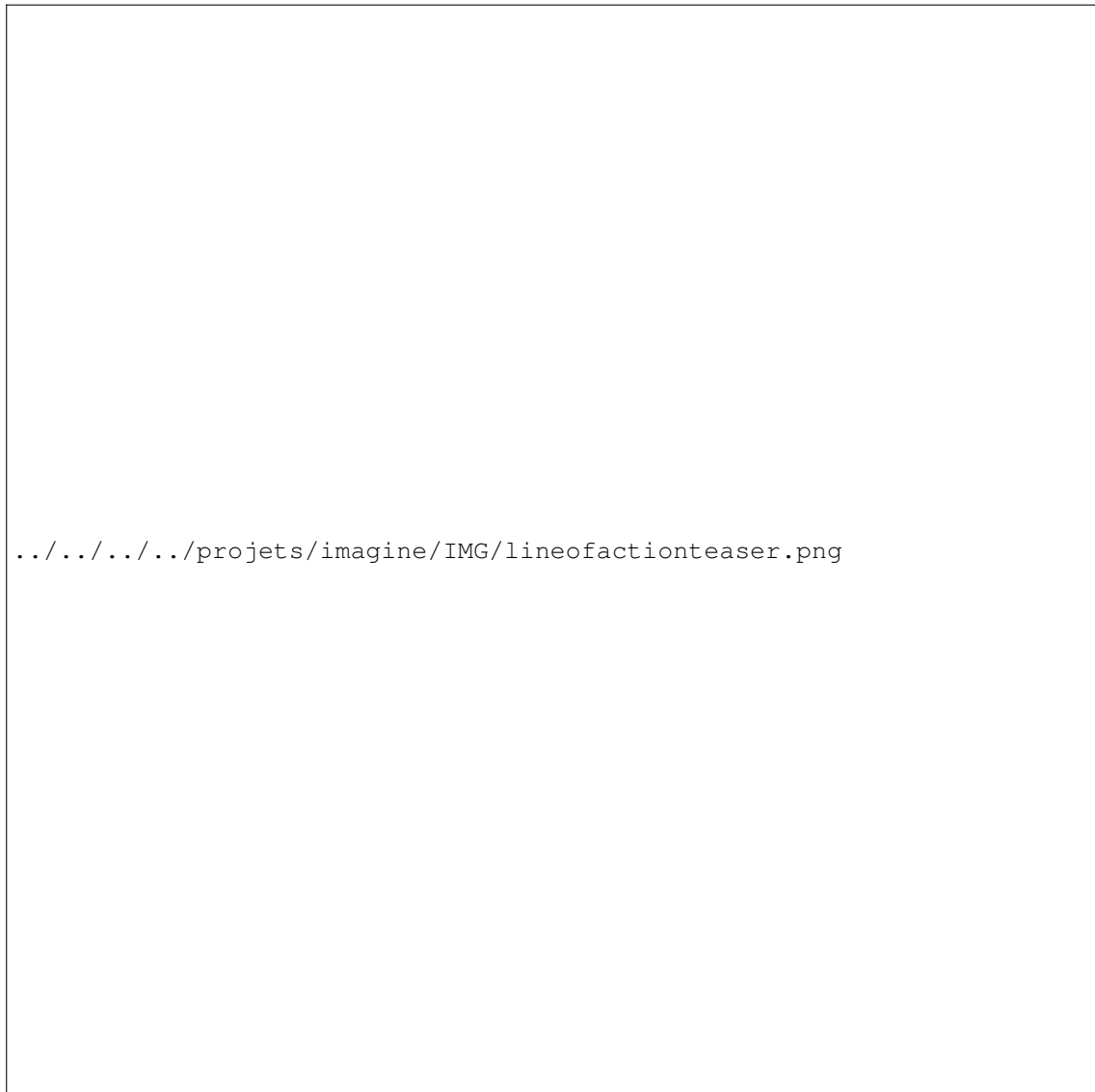../../../../projets/imagine/IMG/lineofactionteaser.png

Figure 10. The line of action defines the pose of the virtual character.

5.5.2. *Sculpting methods*

Participants: Marie-Paule Cani, Stefanie Hahmann, Damien Rohmer, Lucian Stanculescu.

Sculpting methods is a very powerful approach to design virtual models from an existing model. In the work of Lucian Stanculescu [13] we extend the standard sculpting paradigm of surfaces, to multi-dimensional nested structures. In this method, lower dimensional structures such as points and curves can be defined on the surface to defined a nested structure. Each part can follow a specific deformation behaviors. We therefore categorize the geometrical and topological behavior of the structure (such as rigidity or mutability) to develop a wider range of possible deformation. This method facilitate the persistence of sharp features that automatically split or merge with variable rigidity, even when the shape chages genus. This approach enable to deform a surface exhibiting typical behavior of both CAD model with sharp edges, and CG model with smooth surfaces as seen in fig. 11 .

../../../../projets/imagine/IMG/lucian.png

Figure 11. Deformed surface with nested structure from [13].

We also extends the sculpting approach to handle detailed surfaces. During sculpting deformation such as local stretching, the surface details should not extend as the global shape, but rather duplicates to ensure that the surface keeps his detailed appearance. We studied this question under two different approaches.

The first one, in collaboration with Max-Planck Institute focussed on the deformation of 1D-like parametric structure such as castle walls of centripede characters. The method enable to freely extend, compress, split and merges parts of the structures. The deformed structure is generated by an assembly of basic parts whose behaviors are encoded using a discrete shape grammar. During deformation, the system finds the most suitable

collection of parts to assemble and ensure that the global shapes is coherent with the input rules. This work as been published in Eurographics 2014 [10].

The second approach consists in extending sculpting to continuous freeform deformation of a 2D surface with details. During the deformation gesture of stretching or compression, details on the surfaces should seamlessly appear or desappear continuously. In this work, we studied the simpler case of a planar surface with high field details and presented our result in GTMG [27]. We now work on the more general extention to this work as a collaboration with Max-Plack Institute and University of College London.

5.5.3. Interactive control of procedural models: terrains and waterfalls

Participants: Adrien Bernhardt, Marie-Paule Cani, Arnaud Emilien, Ulysse Vimont.

Procedural models, used for easily modeling large, natural environments, pose a specific challenge in terms of user control: how can these automatic methods, useful for quickly generating a huge number of self-similar details, be adapted to allow the coarse to fine level of control needed by the users?

This topic was first explored within Adrien Bernhard's PhD thesis [1], where we introduced a real-time terrain modeling tool using a fast GPU-based terrain solver with a lightweight CPU-based data structure. This tool was recently extended in collaboration with Cambridge University, to enable first-person sketch-based editing of terrains models.

Secondly, we have been working on interactive procedural modeling of plausible waterfalls, in collaboration with Montreal University. Offering interactive user control for this application is particularly challenging, since the shape taken by a fall heavily depends on the underlying terrain. Our solution, based on vectorial user-control, on a flow solver, and on procedural adaptation of the underlying terrain, enable users to quickly create plausible flowing water, while controlling which fall segments are in contact with the terrain (vs. in free fall), the topology of the network, and how much the flow should adapt to the current terrain, vs. the terrain to the user-designed trajectories (fig. 12). A paper is under review and a presentation has been made at the AFIG conference [16].

../../../../projets/imagine/IMG/trou_2.jpg ../../projets/imagine/IMG/trou.jpg

Figure 12. Waterfalls modeling using our approach developped in [16]

5.5.4. Interaction methods

Participants: Rémi Brouet, Marie-Paule Cani.

We are currently exploring the use of multi-touch tables for the interactive design and editing of 3D scenes, in collaboration with the human-computer interaction group of LIG laboratory. The main challenge here is to find out how to use a 2D interaction media for editing 3D content, hence how to intuitively control the third dimension (depth, non-planar rotations, 3D deformations, etc).

Our first work consisted in an user study where we analyzed all possible hand interactions on table-tops and explored the ways users would intuitively try to manipulate 3D environments, either for changing the camera position or for moving objects around [20]. We extracted a general interaction pattern from this study. Our implementation enables both seamless navigation and docking in 3D scenes, without the need for any menu or button to change mode. We are currently extending this work to object editing scenarios, where shapes are to be bent of twisted in 3D using 2D interaction.

<p style="text-align:center;color:red;font-weight:bold;">LEAR Project-Team</p>

6. New Results

6.1. Visual recognition in images

6.1.1. Label-Embedding for Attribute-Based Classification

Participants: Zeynep Akata, Florent Perronnin, Zaid Harchaoui, Cordelia Schmid.

Attributes are an intermediate representation, which enables parameter sharing between classes, a must when training data is scarce. We propose in [13] to view attribute-based image classification as a label-embedding problem: each class is embedded in the space of attribute vectors. We introduce a function which measures the compatibility between an image and a label embedding, as shown in Figure 1 . The parameters of this function are learned on a training set of labeled samples to ensure that, given an image, the correct classes rank higher than the incorrect ones. Results on the Animals With Attributes and Caltech-UCSD-Birds datasets show that the proposed framework outperforms the standard Direct Attribute Prediction baseline in a zero-shot learning scenario. The label embedding framework offers other advantages such as the ability to leverage alternative sources of information in addition to attributes (e.g. class hierarchies) or to transition smoothly from zero-shot learning to learning with large quantities of data.

6.1.2. Good Practice in Large-Scale Learning for Image Classification

Participants: Zeynep Akata, Florent Perronnin, Zaid Harchaoui, Cordelia Schmid.

In this paper [2], we benchmark several SVM objective functions for large-scale image classification. We consider one-vs-rest, multi-class, ranking, and weighted approximate ranking SVMs. A comparison of online and batch methods for optimizing the objectives shows that online methods perform as well as batch methods in terms of classification accuracy, but with a significant gain in training speed. Using stochastic gradient descent, we can scale the training to millions of images and thousands of classes. Our experimental evaluation shows that ranking-based algorithms do not outperform the one-vs-rest strategy when a large number of training examples are used. Furthermore, the gap in accuracy between the different algorithms shrinks as the dimension of the features increases. We also show that learning through cross-validation the optimal rebalancing of positive and negative examples can result in a significant improvement for the one-vs-rest strategy. Finally, early stopping can be used as an effective regularization strategy when training with online algorithms. Following these "good practices", we were able to improve the state-of-the-art on a large subset of 10K classes and 9M images of ImageNet from 16.7% Top-1 accuracy to 19.1%.

6.1.3. Segmentation Driven Object Detection with Fisher Vectors

Participants: Ramazan Gokberk Cinbis, Jakob Verbeek, Cordelia Schmid.

In [18], we present an object detection system based on the Fisher vector (FV) image representation computed over SIFT and color descriptors. For computational and storage efficiency, we use a recent segmentation-based method to generate class-independent object detection hypotheses, in combination with data compression techniques. Our main contribution is a method to produce tentative object segmentation masks to suppress background clutter in the features. As illustrated in Figure 2 , re-weighting the local image features based on these masks is shown to improve object detection significantly. We also exploit contextual features in the form of a full-image FV descriptor, and an inter-category rescoring mechanism. Our experiments on the VOC 2007 and 2010 datasets show that our detector improves over the current state-of-the-art detection results.

6.1.4. Image Classification with the Fisher Vector: Theory and Practice

Participants: Jorge Sánchez, Florent Perronnin, Thomas Mensink, Jakob Verbeek.

Figure 1. Much work in computer vision has been devoted to image embedding (left): how to extract suitable features from an image? We focus on label embedding (right): how to embed class labels in a Euclidean space? We use attributes as side information for the label embedding and measure the "compatibility" between the embedded inputs and outputs with a function F

../../../../projets/lear/IMG/cinbis1.png

Figure 2. The image on the left and the one on the right show the top detection without and with using our segmentation-driven descriptors, respectively.

A standard approach to describe an image for classification and retrieval purposes is to extract a set of local patch descriptors, encode them into a high-dimensional vector and pool them into an image-level signature. The most common patch encoding strategy consists in quantizing the local descriptors into a finite set of prototypical elements. This leads to the popular Bag-of-Visual words (BOV) representation. In [10], we propose to use the Fisher Kernel framework as an alternative patch encoding strategy: we describe patches by their deviation from a "universal" generative Gaussian mixture model. This representation, which we call Fisher Vector (FV) has many advantages: it is efficient to compute, it leads to excellent results even with efficient linear classifiers, and it can be compressed with a minimal loss of accuracy using product quantization. We report experimental results on five standard datasets – PASCAL VOC 2007, Caltech 256, SUN 397, ILSVRC 2010 and ImageNet10K – with up to 9M images and 10K classes, showing that the FV framework is a state-of-the-art patch encoding technique. In figure 3 we show a representative benchmark performance comparison between BOV and FV representations.

Figure 3. Accuracy of the BOV and the FV as a function of the number of Gaussians (left) and feature dimensionality (right) on PASCAL VOC 2007 with SIFT descriptors only.

6.2. Learning and statistical models

6.2.1. *Kernel-Based Methods for Hypothesis Testing: A Unified View*
Participants: Zaid Harchaoui, Francis Bach, Olivier Cappe, Eric Moulines.

Kernel-based methods provide a rich and elegant framework for developing nonparametric detection procedures for signal processing. Several recently proposed procedures can be simply described using basic concepts of reproducing kernel Hilbert space embeddings of probability distributions, namely mean elements and covariance operators. In [5], we propose a unified view of these tools, and draw relationships with information divergences between distributions (see Figure 4).

6.2.2. *Supervised Feature Selection in Graphs with Path Coding Penalties and Network Flows*
Participants: Julien Mairal, Bin Yu.

../../../../projets/lear/IMG/zaid1.png

Figure 4. A schematic view of kernel embedding and mean element

In this paper [6], we consider supervised learning problems where the features are embedded in a graph, such as gene expressions in a gene network. In this context, it is of much interest to automatically select a subgraph with few connected components; by exploiting prior knowledge, one can indeed improve the prediction performance or obtain results that are easier to interpret. Regularization or penalty functions for selecting features in graphs have recently been proposed, but they raise new algorithmic challenges. For example, they typically require solving a combinatorially hard selection problem among all connected subgraphs. In this paper, we propose computationally feasible strategies to select a sparse and well-connected subset of features sitting on a directed acyclic graph (DAG), see Figure 5 . We introduce structured sparsity penalties over paths on a DAG called "path coding" penalties. Unlike existing regularization functions that model long-range interactions between features in a graph, path coding penalties are tractable. The penalties and their proximal operators involve path selection problems, which we efficiently solve by leveraging network flow optimization. We experimentally show on synthetic, image, and genomic data that our approach is scalable and leads to more connected subgraphs than other regularization functions for graphs.

Figure 5. Network Flow Model with Costs on Arcs for the Path Selection Problem

6.2.3. Structured Penalties for Log-linear Language Models

Participants: Anil Nelakanti, Cédric Archambeau, Julien Mairal, Francis Bach, Guillaume Bouchard.

Language models can be formalized as loglinear regression models where the input features represent previously observed contexts up to a certain length m. The complexity of existing algorithms to learn the parameters by maximum likelihood scale linearly in nd, where n is the length of the training corpus and d is the number of observed features. In this paper [26], we present a model that grows logarithmically in d, making it possible to efficiently leverage longer contexts (see Figure 6). We account for the sequential structure of natural language using tree-structured penalized objectives to avoid overfitting and achieve better generalization.

6.2.4. Optimization with First-Order Surrogate Functions

Participant: Julien Mairal.

../../../../projets/lear/IMG/julien2.png

Figure 6. The classical measure of performance for natural language models is the perplexity (lower is better). Our models are denoted by ℓ_2^T and ℓ_{inf}^T.

In this paper [23], we study optimization methods consisting of iteratively minimizing surrogates of an objective function, as illustrated in Figure 7 . By proposing several algorithmic variants and simple convergence analyses, we make two main contributions. First, we provide a unified viewpoint for several first-order optimization techniques such as accelerated proximal gradient, block coordinate descent, or Frank-Wolfe algorithms. Second, 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.

../../../../projets/lear/IMG/julien3.png

Figure 7. Illustration of the basic majorization-minimization principle. We compute a surrogate g_n of the objective function f around a current estimate θ_{n-1}. The new estimate θ_n is a minimizer of g_n. The approximation error h_n is smooth.

6.2.5. Stochastic Majorization-Minimization Algorithms for Large-Scale Optimization
Participant: Julien Mairal.

Majorization-minimization algorithms consist of iteratively minimizing amajorizing surrogate of an objective function. Because of its simplicity and its wide applicability, this principle has been very popular in statistics and in signal processing. In this paper [24], we intend to make this principle scalable. We introduce a stochastic majorization-minimization scheme which is able to deal with largescale or possibly infinite data sets. When applied to convex optimization problems under suitable assumptions, we show that it achieves an expected convergence rate of $O(1/\sqrt{n})$ after n iterations, and of $O(1/n)$ for strongly convex functions. Equally important, our scheme almost surely converges to stationary points for a large class of non-convex problems. We develop several efficient algorithms based on our framework. First, we propose a new stochastic proximal gradient method, which experimentally matches state-of-the-art solvers for large-scale ℓ_1- logistic regression.

Second, we develop an online DC programming algorithm for non-convex sparse estimation. Finally, we demonstrate the effectiveness of our approach for solving large-scale structured matrix factorization problems.

6.3. Recognition in video

6.3.1. *Temporal Localization of Actions with Actoms*

Participants: Adrien Gaidon, Zaid Harchaoui, Cordelia Schmid.

In this paper [4], we address the problem of localizing actions, such as opening a door, in hours of challenging video data. We propose a model based on a sequence of atomic action units, termed "actoms", that are semantically meaningful and characteristic for the action. Our Actom Sequence Model (ASM) represents an action as a sequence of histograms of actom-anchored visual features, which can be seen as a temporally structured extension of the bag-of-features. Training requires the annotation of actoms for action examples. At test time, actoms are localized automatically based on a non-parametric model of the distribution of actoms, which also acts as a prior on an action's temporal structure. We present experimental results on two recent benchmarks for action localization "Coffee and Cigarettes" and the "DLSBP" dataset. We also adapt our approach to a classification-by-localization set-up, and demonstrate its applicability on the challenging "Hollywood 2" dataset. We show that our ASM method outperforms the current state of the art in temporal action localization, as well as baselines that localize actions with a sliding window method (see Figure 8).

Figure 8. Illustration of actoms-based decomposition of actions.

6.3.2. Activity representation with motion hierarchies

Participants: Adrien Gaidon, Zaid Harchaoui, Cordelia Schmid.

Complex activities, e.g., pole vaulting, are composed of a variable number of sub-events connected by complex spatio-temporal relations, whereas simple actions can be represented as sequences of short temporal parts. In [3], we learn hierarchical representations of activity videos in an unsupervised manner. These hierarchies of mid-level motion components are data-driven decompositions specific to each video. We introduce a spectral divisive clustering algorithm to efficiently extract a hierarchy over a large number of tracklets (i.e., local trajectories). We use this structure to represent a video as an unordered binary tree. We model this tree using nested histograms of local motion features. We provide an efficient positive definite kernel that computes the structural and visual similarity of two hierarchical decompositions by relying on models of their parent-child relations. We present experimental results on four recent challenging benchmarks: the High Five dataset, the Olympics Sports dataset, the Hollywood 2 dataset, and the HMDB dataset. We show that per-video hierarchies provide additional information for activity recognition. Our approach improves over unstructured activity models, baselines using other motion decomposition algorithms, and the state of the art (see Figure 9).

../../../../projets/lear/IMG/zaid2.png

Figure 9. Illustration of motion hierarchies for weight-lifting.

6.3.3. DeepFlow: Large displacement optical flow with deep matching

Participants: Philippe Weinzaepfel, Jerome Revaud, Zaid Harchaoui, Cordelia Schmid.

Optical flow computation is a key component in many computer vision systems designed for tasks such as action detection or activity recognition. However, despite several major advances over the last decade, handling large displacement in optical flow remains an open problem. Inspired by the large displacement optical flow of Brox and Malik, our approach, termed DeepFlow, blends a matching algorithm with a variational approach for optical flow. We propose in [31] a descriptor matching algorithm, tailored to the optical flow problem, that allows to boost performance on fast motions. The matching algorithm builds upon a multi-stage architecture with 6 layers, interleaving convolutions and max-pooling, a construction akin to deep convolutional nets. Figure 10 shows an outline of our approach. Using dense sampling, it allows to efficiently retrieve quasi-dense correspondences, and enjoys a built-in smoothing effect on descriptors matches, a valuable asset for integration into an energy minimization framework for optical flow estimation. DeepFlow efficiently handles large displacements occurring in realistic videos, and shows competitive performance on optical flow benchmarks. Furthermore, it sets a new state-of-the-art on the MPI-Sintel dataset.
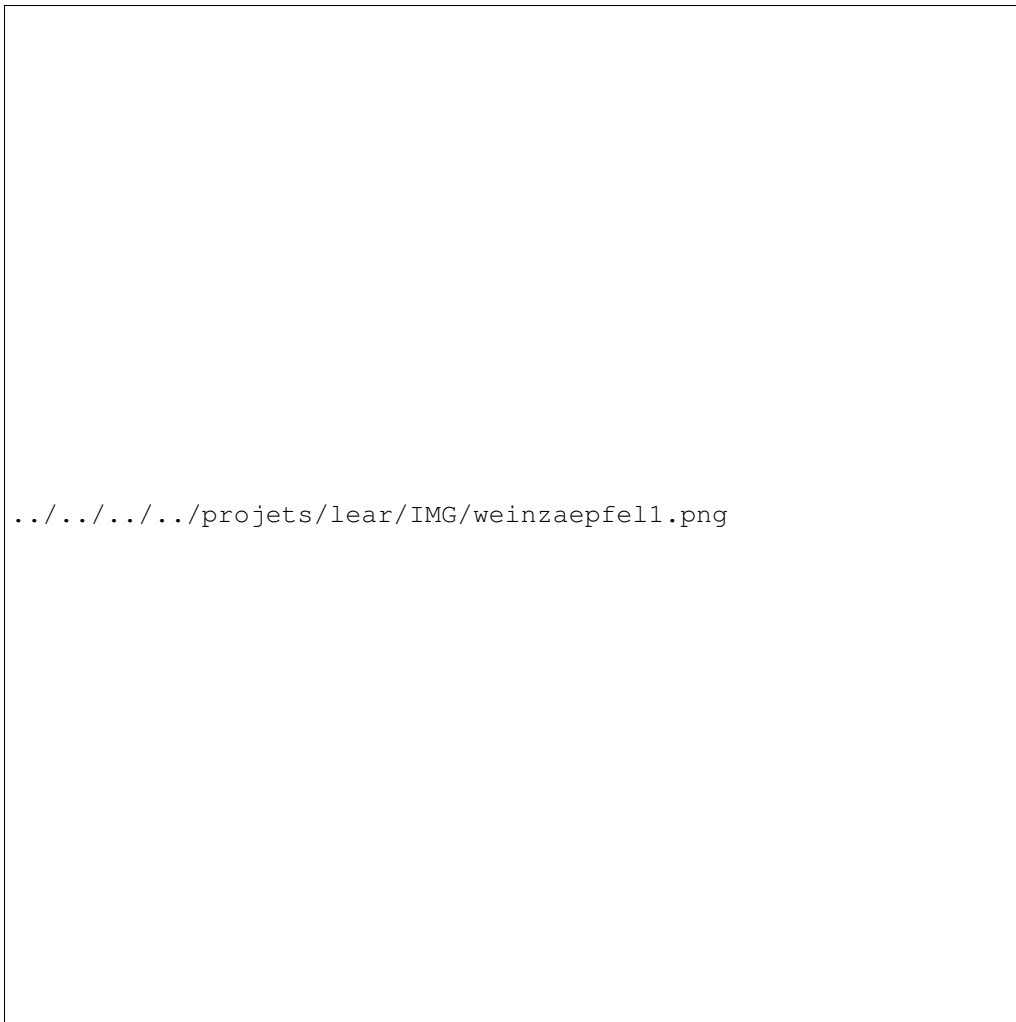
../../../../projets/lear/IMG/weinzaepfel1.png

Figure 10. Outline of DeepFlow.

6.3.4. *Event retrieval in large video collections with circulant temporal encoding*

Participants: Jerome Revaud, Matthijs Douze, Cordelia Schmid, Hervé Jégou.

This paper [28] presents an approach for large-scale event retrieval. Given a video clip of a specific event, e.g., the wedding of Prince William and Kate Middleton, the goal is to retrieve other videos representing the same event from a dataset of over 100k videos. 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 compare the videos in the frequency domain. This offers a significant gain in complexity and accurately localizes the matching parts of videos, see Figure 11 . Furthermore, we extend product quantization to complex vectors in order to compress our descriptors, and to compare them in the compressed domain. Our method outperforms the state of the art both in search quality and query time on two large-scale video benchmarks for copy detection, Trecvid and CCweb. Finally, we introduce a challenging dataset for event retrieval, EVVE, and report the performance on this dataset.

Figure 11. Example of correctly aligned videos. Each row is a different video, and each column corresponds to temporally aligned frames from the videos.

6.3.5. Dense trajectories and motion boundary descriptors for action recognition
Participants: Heng Wang, Alexander Kläser, Cordelia Schmid, Cheng-Lin Liu.

Figure 12. Illustration of the information captured by HOG, HOF, and MBH descriptors. Gradient/flow orientation is indicated by color (hue) and magnitude by saturation. The optical flow (top, middle) shows constant motion in the background, which is due to the camera movements. The motion boundaries (right) encode the relative motion between the person and the background.

This paper [11] introduces a video representation based on dense trajectories and motion boundary descriptors. Trajectories capture the local motion information of the video. A state-of-the-art optical flow algorithm enables a robust and efficient extraction of the dense trajectories. As descriptors we extract features aligned with the trajectories to characterize shape (point coordinates), appearance (histograms of oriented gradients) and motion (histograms of optical flow). Additionally, we introduce a descriptor based on motion boundary histograms (MBH) (see the visualization in Figure 12), which is shown to consistently outperform other state-of-the-art descriptors, in particular on real-world videos that contain a significant amount of camera motion.

We evaluate our video representation in the context of action classification on nine datasets, namely KTH, YouTube,Hollywood2, UCF sports, IXMAS, UIUC, Olympic Sports, UCF50 and HMDB51. On all datasets our approach outperforms current state-of-the-art results.

6.3.6. Action Recognition with Improved Trajectories

Participants: Heng Wang, Cordelia Schmid.

Figure 13. Visualization of human detection and inlier matches (top, left) as well as removed background trajectories, which are due to camera motion (top, right). The bottom row compares the original optical flow (bottom, left) and the warped version (bottom, right).

This paper [30] improves dense trajectories by taking into account camera motion to correct them. To estimate camera motion, we match feature points between frames using SURF descriptors and dense optical flow, which are shown to be complementary. These matches are, then, used to robustly estimate a homography with RANSAC. Human motion is in general different from camera motion and generates inconsistent matches. To improve the estimation, a human detector is employed to remove these matches. Given the estimated camera motion, we remove trajectories consistent with it. We also use this estimation to cancel out camera motion from the optical flow. This significantly improves motion-based descriptors, such as HOF and MBH (see Figure 13). Experimental results on four challenging action datasets (i.e., Hollywood2, HMDB51, Olympic Sports and UCF50) significantly outperform the current state of the art.

6.3.7. Action and event recognition with Fisher vectors on a compact feature set

Participants: Dan Oneaţă, Jakob Verbeek, Cordelia Schmid.

Action recognition in uncontrolled video is an important and challenging computer vision problem. Recent progress in this area is due to new local features and models that capture spatio-temporal structure between local features, or human-object interactions. Instead of working towards more complex models, we focus in this paper [27] on the low-level features and their encoding. We evaluate the use of Fisher vectors as an alternative to bag-of-word histograms to aggregate a small set of state-of-the-art low-level descriptors, in combination with linear classifiers. We present a large and varied set of evaluations, considering (i) classification of short actions in five datasets, (ii) localization of such actions in feature-length movies, and (iii) large-scale recognition of complex events. We find that for basic action recognition and localization MBH features alone are enough for state-of-the-art performance. For complex events we find that SIFT and MFCC features provide complementary cues. On all three problems we obtain state-of-the-art results, while using fewer features and less complex models.

6.3.8. *Stable hyper-pooling and query expansion for event detection*

Participants: Matthijs Douze, Jerome Revaud, Cordelia Schmid, Hervé Jégou.

This work [19] makes two complementary contributions to event retrieval in large collections of videos. First, we compare different ways of quantizing video frame descriptors in terms of temporal stability. Our best choices compare favorably with the standard pooling technique based on k-means quantization, see Figure 14 . Second, we introduce a technique to improve the ranking. It can be interpreted either as a query expansion method or as a similarity adaptation based on the local context of the query video descriptor. Experiments on public benchmarks show that our methods are complementary and improve event retrieval results, without sacrificing efficiency.

6.3.9. *Finding Actors and Actions in Movies.*

Participants: Piotr Bojanowski, Francis Bach, Ivan Laptev, Jean Ponce, Cordelia Schmid, Josef Sivic.

This work [16] addresses the problem of learning a joint model of actors and actions in movies using weak supervision provided by scripts. Specifically, we extract actor/action pairs from the script and use them as constraints in a discriminative clustering framework. The corresponding optimization problem is formulated as a quadratic program under linear constraints. People in video are represented by automatically extracted and tracked faces together with corresponding motion features. First, we apply the proposed framework to the task of learning names of characters in movies and demonstrate significant improvements over previous methods used for this task. Second, we explore joint actor/action constraints and show their advantage for weakly supervised action learning. We validate our method in the challenging setting of localizing and recognizing characters and their actions in the feature length movies Casablanca and American Beauty. Figure 15 shows an example of our results.

../../../../projets/lear/IMG/douze1.png

Figure 14. Several quantizations of video frame descriptors (left) to a color-coded index in $\{0, ..., 31\}$. Leftmost column: standard k-means, right: the proposed SSC. Time runs vertically.

../../../../projets/lear/IMG/bojanowski1.png

Figure 15. Automatic detection and annotation of characters and their actions in the movie Casablanca. The automatically resolved correspondence between video and script is color-coded.

<h2 style="text-align:center; color:red;">MAVERICK Project-Team</h2>

6. New Results

6.1. Visual perception

6.1.1. Decomposing intensity gradients into information about shape and material

Participants: Pascal Barla, Romain Vergne, Roland W. Fleming.

Recent work has shown that the perception of 3D shapes, material properties and illumination are inter-dependent, although for practical reasons, each set of experiments has probed these three causal factors independently. Most of these studies share a common observation though: that variations in image intensity (both their magnitude and direction) play a central role in estimating the physical properties of objects and illumination. Our aim is to separate retinal image intensity gradients into contributions of different shape and material properties, through a theoretical analysis of image formation. We find that gradients can be understood as the sum of three terms: variations of surface depth conveyed through surface-varying reflectance and near-field illumination effects (shadows and inter-reflections); variations of surface orientation conveyed through reflections and far-field lighting effects; and variations of surface micro-structures conveyed through anisotropic reflections. We believe our image gradient decomposition constitutes a solid and novel basis for perceptual inquiry. We first illustrate each of these terms with synthetic 3D scenes rendered with global illumination. We then show that it is possible to mimic the visual appearance of shading and reflections directly in the image, by distorting patterns in 2D. Finally, we discuss the consistency of our mathematical relations with observations drawn by recent perceptual experiments, including the perception of shape from specular reflections and texture. In particular, we show that the analysis can correctly predict certain specific illusions of both shape and material.

6.1.2. Predicting the effects of illumination in shape from shading

Participants: Roland W. Fleming, Romain Vergne, Steven Zucker.

Shading depends on different interactions between surface geometry and lighting. Under collimated illumination, shading is dominated by the 'direct' term, in which image intensities vary with the angle between surface normals and light sources. Diffuse illumination, by contrast, is dominated by 'vignetting effects' in which image intensities vary with the degree of self-occlusion (the proportion of incoming direction that each surface point 'sees'). These two types of shading thus lead to very different intensity patterns, which raises the question of whether shading inferences are based directly on image intensities. We show here that the visual system uses 2D orientation signals ('orientation fields') to estimate shape, rather than raw image intensities and an estimate of the illuminant. We rendered objects under varying illumination directions designed to maximize the effects of illumination on the image. We then passed these images through monotonic, non-linear intensity transfer functions to decouple luminance information from orientation information, thereby placing the two signals in conflict (Figure 6). In Task 1 subjects adjusted the 3D shape of match objects to report the illusory effects of changes of illumination direction on perceived shape. In Task 2 subjects reported which of a pair of points on the surface appeared nearer in depth. They also reported perceived illumination directions for all stimuli. We find that the substantial misperceptions of shape are well predicted by orientation fields, and poorly predicted by luminance-based shape from shading. For the untransformed images illumination could be estimated accurately, but not for the transformed images. Thus shape perception was, for these examples, independent of the ability to estimate the lighting. Together these findings support neurophysiological estimates of shape from the responses of orientation selective cell populations, irrespective of the illumination conditions.

6.1.3. Evaluation of Depth of Field for Depth Perception in DVR

Participants: Pascal Grosset, Charles Hansen, Georges-Pierre Bonneau.

../../../../projets/maverick/IMG/orientation-fields.jpg

Figure 6. Top: one single shape is shaded using multiple light source directions. Bottom: one single light source is used to shade multiple shapes. Each of them has approximatively the same percept as for the corresponding light direction. Similarity between orientation fields can be seen in the insets.

We study the use of Depth of Field for depth perception in Direct Volume Rendering (Figure 7). Direct Volume Rendering with Phong shading and perspective projection is used as the baseline. Depth of Field is then added to see its impact on the correct perception of ordinal depth. Accuracy and response time are used as the metrics to evaluate the usefulness of Depth of Field. The on site user study has two parts: static and dynamic. Eye tracking is used to monitor the gaze of the subjects. From our results we see that though Depth of Field does not act as a proper depth cue in all conditions, it can be used to reinforce the perception of which feature is in front of the other. The best results (high accuracy & fast response time) for correct perception of ordinal depth is when the front feature (out of the users were to choose from) is in focus and perspective projection is used. Our work has been published in the proceedings of the Pacific Graphics conference in 2013 [16].

../../../../projets/maverick/IMG/DepthOfFieldExperiment.jpg

Figure 7. A user study is conducted to evaluate the use of Depth of Field for depth perception in Direct Volume Rendering. The user has to say which of the two highlighted features is in the front. Statistical analysis of the results is performed to assess the effect of Depth of Field in the visualization. Results are detailed in the Pacific Graphics paper [16].

6.2. Visualization

6.2.1. Morse-Smale complexes

Participants: Léo Allemand-Giorgis, Georges-Pierre Bonneau, Stefanie Hahmann.

Preserving meaningful local extrema of scalar data in a visualization while removing nearby extrema with similar values is a powerful way for enhancing the appearance of significant features. For the special case of monotonic data, e.g. data with no local extrema in the interior of the domain, the visualization should not introduce spurious local extrema. We study a new piecewise polynomial interpolant that preserves the monotonicity of scalar data defined on a 2D uniform grid. Based on this interpolant, we also plan to

introduce a new method for visualizing data that has been simplified according to its Morse-Smale complex, a combinatorial structure connecting the critical points and partitioning the domain into a set of monotonic regions. In contrast with previous analogous works, our approach uses piecewise polynomial functions defined in each monotonic region instead of optimizing values on the original mesh vertices. We have presented our first results in a workshop and have submitted a paper for a book chapter about our new monotonic interpolant.

6.2.2. *Computation of components' interfaces in highly complex assemblies*

Participants: François Jourdes, Georges-Pierre Bonneau, Stefanie Hahmann, François Faure.

The preparation of CAD models from complex assemblies for simulation purposes is a very time-consuming and tedious process, since many tasks such as meshing and idealization are still completed manually. Herein, the detection and extraction of geometric interfaces between components of the assembly is of central importance not only for the simulation objectives but also for all necessary shape transformations such as idealizations or detail removals. It is a repetitive task in particular when complex assemblies have to be dealt with. This paper proposes a method to rapidly and fully automatically generate a precise geometric description of interfaces in generic B-Rep CAD models. The approach combines an efficient GPU ray-casting technique commonly used in computer graphics with a graph-based curve extraction algorithm. Not only is it able to detect a large number of interfaces efficiently, but it also provides an accurate Nurbs geometry of the interfaces, that can be stored in a plain STEP file for further downstream treatment. We demonstrate our approach on examples from aeronautics and automotive industry, see Figure 8 . Our results have been funded in by the ANR Project ROMMA. They have been published as a journal paper in [13], and presented at the Solid and Physical Modeling conference in 2013.

6.3. Image creation and editing

6.3.1. *Accurate Binary Image Selection from Inaccurate User Input*

Participants: Subr Kartic, Paris Sylvain, Soler Cyril, Kautz Jan.

Selections are central to image editing, since they are the starting point of common operations such as copy-pasting and local edits. Creating them by hand is particularly tedious and scribble-based techniques have been introduced to assist the process. By interpolating a few strokes specified by users, these methods generate precise selections. However, most of the algorithms assume a 100% accurate input, and even small inaccuracies in the scribbles often degrade the selection quality, which imposes an additional burden on users. In this work, we propose a selection technique tolerant to input inaccuracies (See example on Figure 9). We use a dense conditional random field (CRF) to robustly infer a selection from possibly inaccurate input. Further, we show that patch-based pixel similarity functions yield more precise selection than simple point-wise metrics. However, efficiently solving a dense CRF is only possible in low-dimensional Euclidean spaces, and the metrics that we use are high-dimensional and often non-Euclidean. We address this challenge by embedding pixels in a low-dimensional Euclidean space with a metric that approximates the desired similarity function. The results show that our approach performs better than previous techniques and that two options are sufficient to cover a variety of images depending on whether the objects are textured. This work has been published to the Eurographics Conference [15].

6.3.2. *Discrete Texture Design Using a Programmable Approach*

Participants: Hugo Loi, Thomas Hurtut, Romain Vergne, Thollot Joëlle.

Many rendering methods use discrete textures (planar arrangements of vector elements) instead of classic bitmaps. Discrete textures are resolution-insensitive and easily allow to modify the elements' geometry or spatial distribution. However, manually drawing such textures is a time-consuming task. Automating this production is a long-time studied subject. The methods designed for this purpose deal with a difficult tradeoff between the reachable variety of textures and the usability for a community of users. In this work, we show that considering discrete textures as programs allow for a larger variety of textures than relying on a given model. This work has been as a Siggraph 2013 talk [18].
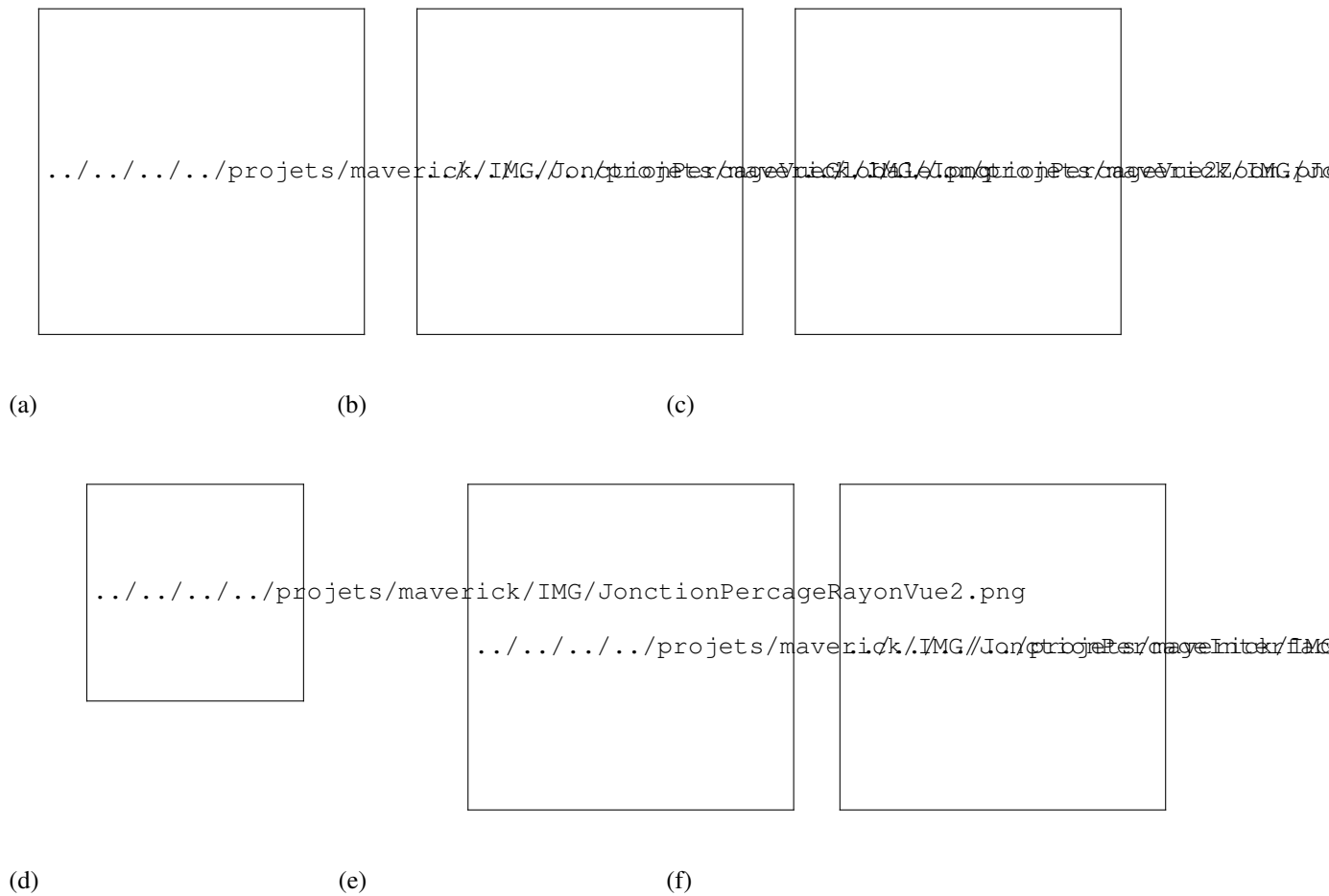
(a)

(b)

(c)

(d)

(e)

(f)

Figure 8. Computing Interfaces of an Aircraft part assembling the wings with the body of an aircraft (model courtesy of EADS). (a,b) two views of the components, (c) exploded view, (d) ray casting for detecting proximities between parts, (e) boundary reconstruction, (f) final interfaces

../../../../projets/maverick/IMG/scribbles.png

Figure 9. Accurate selection from inacurrate input. Although our algorithm is fed with inaccurate selection, it succeeds in separating the background from the foreground components in the image.

Figure 10. Discrete element texture generated with our programmable approach.

6.4. Complex scenes

6.4.1. Appearance pre-filtering

Participants: Heitz Eric, Neyret Fabrice.

Last year work and HPG'12 paper "Representing Appearance and Pre-filtering Subpixel Data in Sparse Voxel Octrees" was dealing with the light and view dependant aspect of complex surfaces due to sub-pixels details. This was done by replacing sub-voxel height fields by gaussian slope distribution and height-correlated colors by its gradient, feeding a Cook-Torrance-like microfacet brdf.

In continuation of this and in the same spirit of replacing sub-pixel values by gaussian distributions to be shaded using the frame of microfacets brdf, this year we addressed the filtering of color maps (on surfaces and per se), displacement map, and reflectance maps, thus obtaining a complete model of the local rendering integral (see Figure 11).

Note that Eric did his work partly during his 6-monthes stay of University of Montreal in the scope of Exploradoc regional founding. He also colaborated with nVIDIA for an on-going work related to animation of GigaVoxels, and we were invited for a stay of several weeks at Weta Digital, NZ to help them applying our techniques.

6.4.2. Filtering Color Mapped Textures and Surfaces

Participants: Heitz Eric, Neyret Fabrice, Nowrouzezahrai Derek, Poulin Pierre.

Indeed, several ubiquitus CG operations like filtering non-linear functions of the data are still mostly unsolved despite their important flaws. Typically, density, noise data, normals or height are filtered before feeding a color look-up texture, despite the strong non-linearity of the transform forbids factoring it out of the integral. This result on very visible flaws such as thin blue bones+air foams appearing as red muscle at distance in volume visualisation, silhouettes and horizon getting the middle tint instead of the integral of tints, procedural noise bump maps and height fields appearing as smooth instead of rough.

../../../../projets/maverick/IMG/prefiltrage1.png ../../../projets/maverick/IMG/prefiltrage2

Figure 11. The resulting pixel color is the integral of the local rendering on the surface, which combines 4 fields which are: the color texture parameter, the visibility from eye, the visibility from light, and the shading. As these are varying and non-linear, the ubicuitous simplification of averaging each term separately is not valid.

Assuming Gaussian distribution of colors within a pixel or voxel, the filtered colors values can be represented as color lobes (i.e. histogramms) instead of scalars. In all the cases where the subpixel/voxel raw data can also be represented as gaussian distribution (e.g. Perlin noise), the filtering is just the inner product of the two lobes. It can easily be tabuled as a 1D LUT MIP-map which LOD corresponds to the standard deviation and thus the scale. Since microfacet brdf models allow to estimate the *visible* slope statistics accounting for light and view visibility, this allows for emerging light-view dependant color effect both acurately and very efficiently. Note that the same scheme applies for colors corelated to orientation rather than heights (see Figure 12). This provides a multiscale representation where subpixel/subvoxel data is represented through lobes which can be precalculated or calculated on demand from the thiner level.

This work was published at ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games (I3D) 2013 [17]. An extended version "Filtering Non-Linear Transfer Functions on Surfaces" was published at IEEE Transactions on Visualization and Computer Graphics 2013 [11].

6.4.3. *Linear Efficient Antialiased Displacement and Reflectance Mapping*
Participants: Dupuy Jonathan, Heitz Eric, Neyret Fabrice.

Here, the last term of the local rendering integration is addressed: the filtering of subpixel/subvoxel geometry and brdf as an appearant brdf applied on a macro-geometry. By re-derivating accurately the brdf of a displacement map assumed to have sub-pixel gaussian distribution (with an exact masking term, more accurate cross-correlated light-view, and offseted appearant lobe), and by noting that the reflectance of the environment can be pre-filtered like the textures of the previous paper, we finally obtain a complete model of pre-fitltered appearance of surfaces (see Figure 13). This work, co-first-authored with Jonathan Dupuy, was published at ACM Transactions on Graphics and presented at Siggraph Asia [8].

6.5. Realistic rendering

6.5.1. *Interactive Rendering of Acquired Materials on Dynamic Geometry Using Frequency Analysis*
Participants: Bagher M. Mahdi, Soler Cyril, Subr Kartic, Belcour Laurent, Holzschuch Nicolas.

../../../../projets/maverick/IMG/teaser0.png ../../../projets/maverick/IMG/teaser1.png

Figure 12. (Left:) Correct pre-filtering of non-linear functions (e.g., LUT) of a noise texture. (Right:) Correct pre-filtering of the appearance of details, with light and view macroscopic dependency and color correlated with depth.

Shading acquired materials with high-frequency illumination is computationally expensive. Estimating the shading integral requires multiple samples of the incident illumination. The number of samples required may vary across the image, and the image itself may have high- and low-frequency variations, depending on a combination of several factors. Adaptively distributing computational budget across the pixels for shading is a challenging problem. In this work, we depict complex materials such as acquired reflectances, interactively, without any precomputation based on geometry. In each frame, we first estimate the frequencies in the local light field arriving at each pixel, as well as the variance of the shading integrand. Our frequency analysis [1] accounts for combinations of a variety of factors: the reflectance of the object projecting to the pixel, the nature of the illumination, the local geometry and the camera position relative to the geometry and lighting. We then exploit this frequency information (bandwidth and variance) to adaptively sample for reconstruction and integration. For example, fewer pixels per unit area are shaded for pixels projecting onto diffuse objects, and fewer samples are used for integrating illumination incident on specular objects (See Figure 14). This work has been published in IEEE Transactions on Visualization and Computer Graphics [3], as a follow up of a previous paper published at the I3D conference.

6.5.2. 5D Covariance Tracing for Efficient Defocus and Motion Blur

Participants: Belcour Laurent, Soler Cyril, Subr Kartic, Holzschuch Nicolas, Durand Frédo.

The rendering of effects such as motion blur and depth-of-field requires costly 5D integrals. We dramatically accelerate their computation through adaptive sampling and reconstruction based on the prediction of the anisotropy and bandwidth of the integrand. For this, we develop a new frequency analysis of the 5D temporal light-field, and show that first-order motion can be handled through simple changes of coordinates in 5D. We further introduce a compact representation of the spectrum using the co- variance matrix and Gaussian approximations. We derive update equations for the 5×5 covariance matrices for each atomic light transport event, such as transport, occlusion, BRDF, texture, lens, and motion. The focus on atomic operations makes our work general, and removes the need for special-case formulas. We present a new rendering algorithm that computes 5D covariance matrices on the image plane by tracing paths through the scene, focusing on the single-bounce case. This allows us to reduce sampling rates when appropriate and perform reconstruction of images with complex depth-of-field and motion blur effects (See Figure 15). This work was published at ACM Transactions on Graphics [5] and presented at Siggraph'2013.

../../../../projets/maverick/IMG/LEAD-R.png

Figure 13. (Left:) Correct pre-filtering of non-linear functions (e.g., LUT) of a noise texture. (Right:) Correct pre-filtering of the appearance of details, with light and view macroscopic dependency and color correlated with depth.

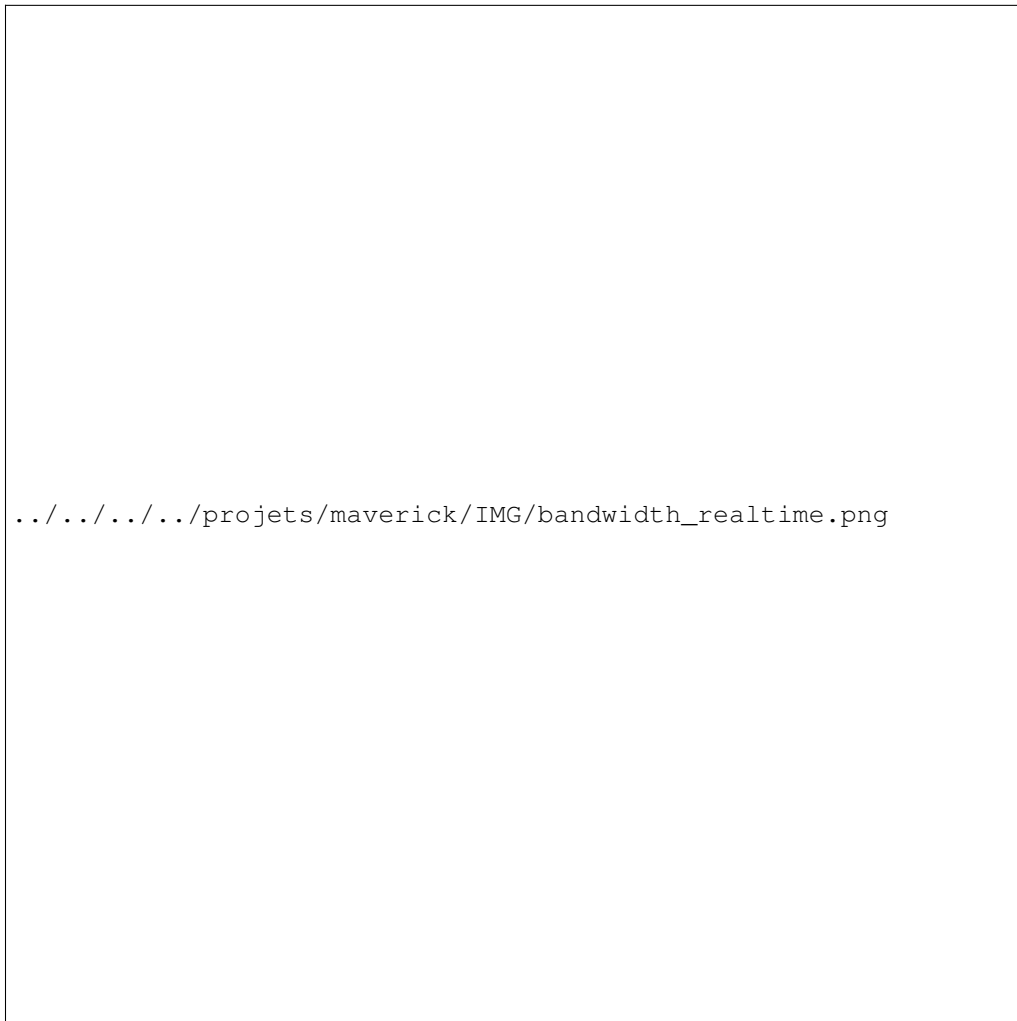../../../../projets/maverick/IMG/bandwidth_realtime.png

Figure 14. Our simplified bandwidth prediction technique is suitable for real-time rendering. It allows us to only compute a subset of the image pixels while concentrating integration cost to the pixels that have the largest variance of the integrand.
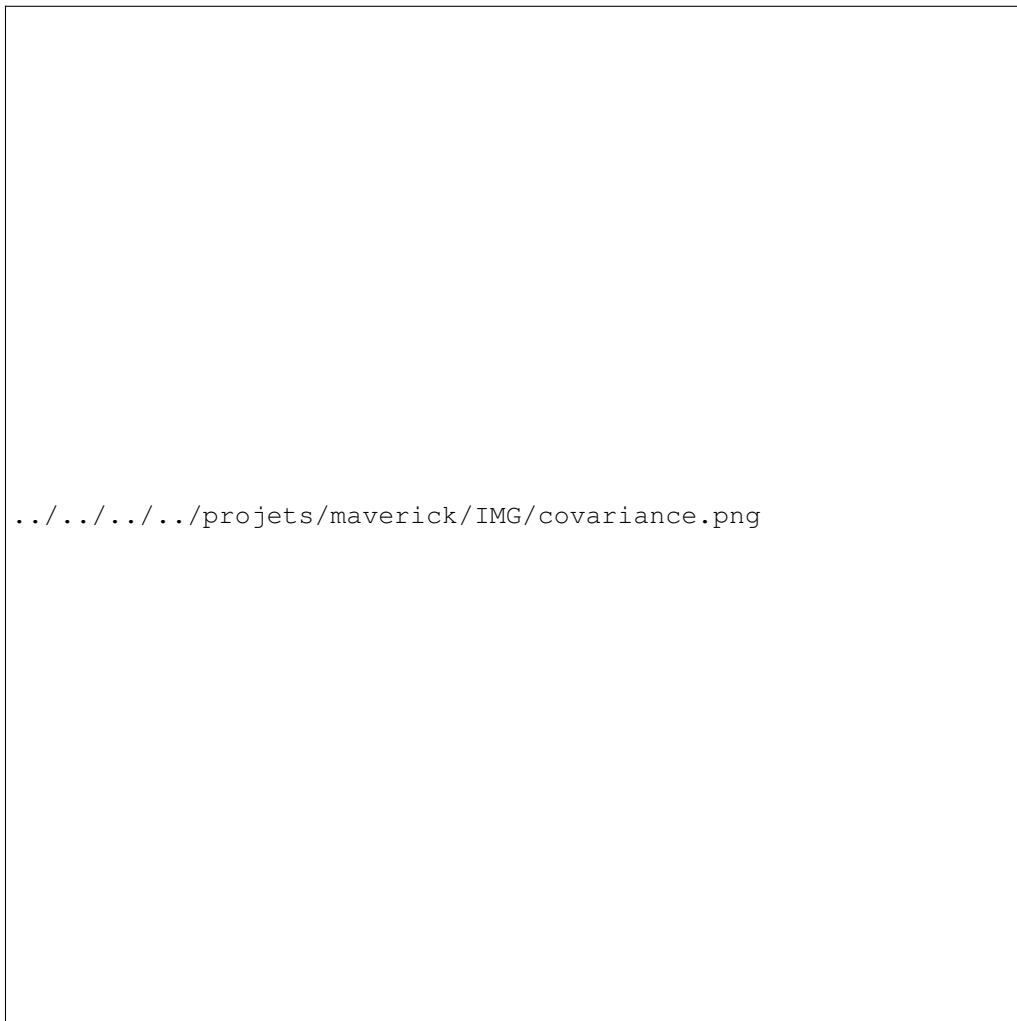
../../../../projets/maverick/IMG/covariance.png

Figure 15. Our covariance analysis of the power spectrum of local illumination allows us to accurately predict the sampling rates an reconstruction filters to significantly increase the convergence of path tracing.

6.5.3. *Accurate and Efficient Filtering using Anisotropic Filter Decomposition*

Participants: Soler Cyril, Bagher Mahdi, Nowrouzezahrai Derek.

Efficient filtering remains an important challenge in computer graphics, particularly when filters are spatially-varying, have large extent, and/or exhibit complex anisotropic profiles. We explored an efficient filtering approach for these difficult cases based on anisotropic filter decomposition (IFD). By decomposing complex filters into linear combinations of simpler, displaced isotropic kernels, and precomputing a compact prefiltered dataset, we are able to interactively apply any number of—potentially transformed—filters to a signal (See Figure 16). Our performance scales linearly with the size of the decomposition, not the size nor the dimensionality of the filter, and our prefiltered data requires reasonnable storage, comparing favorably to the state-of-the-art. We apply IFD to interesting problems in image processing and realistic rendering. This work is currently under submission and a technical report is already available [21].

Figure 16. Isotropic filter decomposition allows us to efficiently perform anisotropic filtering at the cost of minimal pre-computation.

6.5.4. Double- and Multiple-scattering Effects in Translucent Materials

Participants: Holzschuch Nicolas, Gascuel Jean-Dominique.

Some materials, such as coffee, milk or marble, have a soft translucent aspect because of sub-surface scattering: light enters them, is scattered several times inside before leaving in a different place. A full representation of sub-surface scattering effects in illumination simulation is computationally expensive. The main difficult comes from multiple scattering events: the high number of events increases the uncertainty on the result, forcing us to allocate more time for the computations. Recently, we showed that there is a strong correlation between the surface effects of multiple scattering inside the material and the effects after just two scatter events. This knowledge will help in accelerating multiple scattering effects (see figure 17 . We exploited this knowledge to provide a model and implementation for fast computation of double-scattering events, using a precomputed density function stored in a compact way. This work has been published in IEEE Computer Graphics and Applications [12].
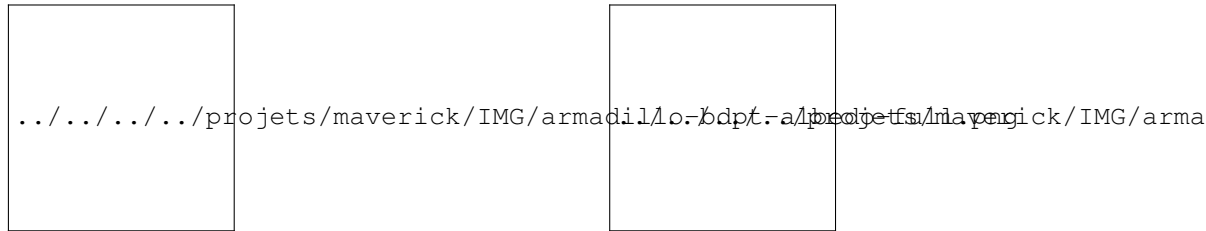

Figure 17. Our approximation achieves results very close to the reference solution: (left) the BDPT reference, in 32min; (right) our solution, in 1.7 min. .

6.6. Inverse problems

6.6.1. Sparse BRDF Approximation using Compressive Sensing

Participants: Zupancic Benoit, Soler Cyril.

BRDF acquisition is a tedious operation, since it requires measuring 4D data. On one side of the spectrum lie explicit methods, which perform many measurements to potentially produce very accurate reflectance data after interpolation. These methods are generic but practically difficult to setup and produce high volume data. On the other side, acquisition methods based on parametric models implicitly reduce the infinite dimensionality of the BRDF space to the number of parameters, allowing acquisition with few samples. However, parametric methods require non linear optimization. They become unstable when the number of parameters is large, with no guaranty that a given parametric model can ever fit particular measurements.

We experiment a new acquisition method where the measurement of the BRDF is performed from a single image, knowing the normals and illumination. To tackle such a severely underconstrained problem, we express the BRDF in a high dimensional basis, and perform the reconstruction using *compressive sensing*, looking for the most sparse solution to the linear problem of fitting the measurement image. Doing so, we leverage the coherency between the measured pixels, while keeping the high dimension of the space the BRDF is searched into.

This work is a very first attempt at reconstructing BRDFs using compressive sensing. In Fig 18 we used a synthetic input image, for the sake of checking the feasibility of the recovery algorithm, in the particular case of an isotropic spatially constant BRDFs. The possibility to extend our theory to non spatially varying and anisotropic BRDFs is currently under investigation. We would like to orient our work toward BRDF acquisition with consumer hardware. In particular, our preliminary results indicate that compressive sensing

could achieve an very accurate acquisition with additional input, such as a video of a static object under probed lighting.

This word has been published as a poster to the Siggraph Asia'2013 conference [24].

../../../../projets/maverick/IMG/ssbrdf.png

Figure 18. Reconstruction of a BRDF from a single image. This work stands up as a preliminary experiment to prove that BRDF acquisition can be performed with a much lower amount of data than what the Nyquist rate predicts for data reconstruction, if using compressive sensing methods.

6.6.2. *Floating tangents for approximating spatial curves with G1 piecewise helices*

Participants: Derouet-Jourdan Alexandre, Bertails-Descoubes Florence, Thollot Joëlle.

Curves are widely used in computer science to describe real-life objects such as slender deformable structures. Using only 3 parameters per element, piecewise helices offer an interesting and compact way of representing digital curves. In our work [7], we present a robust and fast algorithm to approximate Bezier curves with G1 piecewise helices. Our approximation algorithm takes a Bezier spline as input along with an integer N

and returns a piecewise helix with N elements that closely approximates the input curve. The key idea of our method is to take N+1 evenly distributed points along the curve, together with their tangents, and interpolate these tangents with helices by slightly relaxing the points. Building on previous work, we generalize the proof for Ghosh's co-helicity condition, which serves us to guarantee the correctness of our algorithm in the general case. Finally, we demonstrate both the efficiency and robustness of our method by successfully applying it to various datasets of increasing complexity, ranging from synthetic curves created by an artist to automatic image-based reconstructions of real data such as hair, heart muscular fibers or magnetic field lines of a star.

6.6.3. *Inverse Dynamic Hair Modeling with Frictional Contact*

Participants: Derouet-Jourdan Alexandre, Daviet Gilles, Bertails-Descoubes Florence, Thollot Joëlle.

In the latest years, considerable progress has been achieved for accurately acquiring the geometry of human hair, thus largely improving the realism of virtual characters. In parallel, rich physics-based simulators have been successfully designed to capture the intricate dynamics of hair due to contact and friction. However, at the moment there exists no consistent pipeline for converting a given hair geometry into a realistic physics-based hair model. Current approaches simply initialize the hair simulator with the input geometry in the absence of external forces. This results in an undesired sagging effect when the dynamic simulation is started, which basically ruins all the efforts put into the accurate design and/or capture of the input hairstyle. In this work [6] we propose the first method which consistently and robustly accounts for surrounding forces — gravity and frictional contacts, including hair self-contacts — when converting a geometric hairstyle into a physics-based hair model. Taking an arbitrary hair geometry as input together with a corresponding body mesh, we interpret the hair shape as a static equilibrium configuration of a hair simulator, in the presence of gravity as well as hair-body and hair-hair frictional contacts. Assuming that hair parameters are homogeneous and lie in a plausible range of physical values, we show that this large underdetermined inverse problem can be formulated as a well-posed constrained optimization problem, which can be solved robustly and efficiently by leveraging the frictional contact solver of the direct hair simulator. Our method was successfully applied to the animation of various hair geometries, ranging from synthetic hairstyles manually designed by an artist to the most recent human hair data automatically reconstructed from capture.

<p align="center" style="color:red">MESCAL Project-Team</p>

6. New Results

6.1. Simulation

6.1.1. Simulation of Parallel Computing Systems

Researchers in the area of distributed computing conduct many of their experiments in simulation. While packet-level simulation is often used to study network protocols, it can be too costly to simulate network communications for large-scale systems and applications. The alternative chosen in SimGrid and a few other simulation frameworks is to simulate the network based on less costly flow-level models. Surprisingly, in the literature, validation of these flow-level models is at best a mere verification for a few simple cases. Consequently, although distributed computing simulators are widely used, their ability to produce scientifically meaningful results is in doubt. In [13] we focus on the validation of state-of-the-art flow-level network models of TCP communication on Wide Area Networks, via comparison to packet-level simulation. While it is straightforward to show cases in which previously proposed models lead to good results, instead we systematically seek cases that lead to invalid results. Careful analysis of these cases reveal fundamental flaws and also suggest improvements. One contribution of this work is that these improvements lead to a new model that, while far from being perfect, improves upon all previously proposed models. A more important contribution, perhaps, is provided by the pitfalls and unexpected behaviors encountered in this work, leading to a number of enlightening lessons. In particular, this work shows that model validation cannot be achieved solely by exhibiting (possibly many) "good cases." Confidence in the quality of a model can only be strengthened through an invalidation approach that attempts to prove the model wrong.

The previous results assume steady-state and provide thus a reasonable model when message size is very large. Although, such assumptions may be reasonable when studying grid applications, when simulating HPC applications message sizes are often much smaller and phenomenon like slow-start or how communications and computations overlap have to be accurately modeled. Simulation and modeling for performance prediction and profiling is yet essential for developing and maintaining HPC code that is expected to scale for next-generation exascale systems. In [15], [34] we describe an implementation of a flow-based hybrid network model that accounts for factors such as network topology and contention, which are commonly ignored by the LogP models. Although, this may seem like a strange choice, we focus on large-scale, Ethernet-connected systems, as these currently compose 37.8% of the TOP500 index, and this share is expected to increase as higher-speed 10 and 100GbE become more available. Furthermore, the European Mont-Blanc project to study exascale computing by developing prototype systems with low-power embedded devices will also use Ethernet-based interconnect [28]. Our model is implemented within SMPI, an open-source MPI implementation that connects real applications to the SimGrid simulation framework. SMPI provides implementations of collective communications based on current versions of both OpenMPI and MPICH. SMPI and SimGrid also provide methods for easing the simulation of large-scale systems, including shadow execution, memory folding, and support for both online and offline (i.e., post-mortem) simulation. We validate our proposed model by comparing traces produced by SMPI with those from real world experiments, as well as with those obtained using other established network models. Our study shows that SMPI has a consistently better predictive power than classical LogP-based models for a wide range of scenarios including both established HPC benchmarks and real applications.

6.1.2. Perfect Simulation

Perfect simulation is a very efficient technique that uses coupling arguments to provide a sample from the stationary distribution of a Markov chain in a finite time without ever computing the distribution. In [7], we consider Jackson queueing networks (JQN) with finite buffer constraints and analyze the efficiency of sampling from their stationary distribution. In the context of exact sampling, the monotonicity structure of JQNs ensures

that such efficiency is of the order of the coupling time (or meeting time) of two extremal sample paths. In the context of approximate sampling, it is given by the mixing time. Under a condition on the drift of the stochastic process underlying a JQN, which we call *hyper-stability*, in our main result we show that the coupling time is polynomial in both the number of queues and buffer sizes. Then, we use this result to show that the mixing time of JQNs behaves similarly up to a given precision threshold. Our proof relies on a recursive formula relating the coupling times of trajectories that start from network states having 'distance one', and it can be used to analyze the coupling and mixing times of other Markovian networks, provided that they are monotone. An illustrative example is shown in the context of JQNs with blocking mechanisms.

In [35], we extend the technique to handle situations with infinite space state. We consider open JQN with losses with mixed finite and infinite queues and analyze the efficiency of sampling from their exact stationary distribution. Although the underlying Markov chain may have an infinite state space, we show that perfect sampling is possible. The main idea is to use a JQN 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 hyperstable 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 non-monotone networks such as queueing networks with negative customers.

6.2. Interactive Analysis and Visualization of Large Distributed Systems

6.2.1. Interactive Visualization

High performance applications are composed of many processes that are executed in large-scale systems with possibly millions of computing units. A possible way to conduct a performance analysis of such applications is to register in trace files the behavior of all processes belonging to the same application. The large number of processes and the very detailed behavior that we can record about them lead to a trace size explosion both in space and time dimensions. The performance visualization of such data is very challenging because of the quantities involved and the limited screen space available to draw them all. If the amount of data is not properly treated for visualization, the analysis may give the wrong idea about the behavior registered in the traces.

In [33], we detail data aggregation techniques that are fully configurable by the user to control the level of details in both space and time dimensions. We also present two visualization techniques that take advantage of the aggregated data to scale. These features are part of the Viva and Triva open-source tools and framework.

The performance of parallel and distributed applications is also highly dependent on the characteristics of the execution environment. In such environments, the network topology and characteristics directly impact data locality and movements as well as contention, which are key phenomena to understand the behavior of such applications and possibly improve it. Unfortunately few visualizations available to the analyst are capable of accounting for such phenomena. In [26], we propose an interactive topology-based visualization technique based on data aggregation that enables to correlate network characteristics, such as bandwidth and topology, with application performance traces. We claim that such kind of visualization enables to explore and understand non trivial behavior that are impossible to grasp with classical visualization techniques. We also claim that the combination of multi-scale aggregation and dynamic graph layout allows our visualization technique to scale seamlessly to large distributed systems. We support these claims through a detailed analysis of a high performance computing scenario and of a grid computing scenario.

6.2.2. Entropy Based Analysis

Although the previous approaches already improve upon state of the art and are useful on current scenarios, it is clear that at very large scale they would probably not be as effective, which led us to change perspective and to investigate how entropy can help building tractable macroscopic descriptions. Indeed, data aggregation can provide such abstractions by partitioning the systems dimensions into aggregated pieces of information. This process leads to information losses, so the partitions should be chosen with the greatest caution, but

in an acceptable computational time. While the number of possible partitions grows exponentially with the size of the system, we propose in [25] an algorithm that exploits exogenous constraints regarding the system semantics to find best partitions in a linear or polynomial time. We detail two constrained sets of partitions that are respectively applied to temporal and spatial aggregation of an agent-based model of international relations. The algorithm succeeds in providing meaningful high-level abstractions for the system analysis.

Our approach is able to evaluate geographical abstractions used by the domain experts in order to provide efficient and meaningful macroscopic descriptions of the world global state [23]. We also successfully applied this technique to identify international media events by spatially and temporally aggregating RSS Flows of Newspapers [22], in particular with the case of the Syrian civil war between May 2011 and December 2012 [31], [21].

We also applied this technique to the analysis of large distributed systems and combined it with the treemap visualization technique [40], [14]. These features have been integrated in the Viva and Triva open-source tools and framework.

6.3. Trace Management and Analysis

6.3.1. *Embedded Systems*

The growing complexity of embedded system hardware and software makes their behavior analysis a challenging task. In this context, tracing provides relevant information about the system execution and appears to be a promising solution. However, trace management and analysis are hindered by several issues like the diversity of trace formats, the incompatibility of trace analysis methods, the problem of trace size and its storage as well as by the lack of visualization scalability. In [42], [27], [41], we present FrameSoC, a new trace management infrastructure that solves all the above issues together. It provides generic solutions for trace storage and defines interfaces and plugin mechanisms for integrating diverse analysis tools. We illustrate the benefit of FrameSoC with a case study of a visualization module that enables representation scalability of large traces by using an aggregation algorithm. Temporal aggregation techniques based on entropy are also currently integrated to the FrameSoC framework.

6.3.2. *Jobs Resource Utilization*

In HPC community the System Utilization metric enables to determine if the resources of the cluster are efficiently used by the batch scheduler. This metric considers that all the allocated resources (memory, disk, processors, etc) are full-time utilized. To optimize the system performance, we have to consider the effective physical consumption by jobs regarding the resource allocations. This information gives an insight into whether the cluster resources are efficiently used by the jobs. In [20], [30], we propose an analysis of production clusters based on the jobs resource utilization. The principle is to collect simultaneously traces from the job scheduler (provided by logs) and jobs resource consumption. The latter has been realized by developing a job monitoring tool, whose impact on the system has been measured as lightweight (0.35% speed-down). The key point is to statistically analyze both traces to detect and explain underutilization of the resources. This could enable to detect abnormal behavior, bottlenecks in the cluster leading to a poor scalability, and justifying optimizations such as gang scheduling or best effort scheduling. This method has been applied to two medium sized production clusters on a period of eight months.

6.4. Reconstructing the Software Environment of an Experiment

In the scientific experimentation process, an experiment result needs to be analyzed and compared with several others, potentially obtained in different conditions. Thus, the experimenter needs to be able to redo the experiment. Several tools are dedicated to the control of the experiment input parameters and the experiment replay. In parallel concurrent and distributed systems, experiment conditions are not only restricted to the input parameters, but also to the software environment in which the experiment was carried out. It is therefore essential to be able to reconstruct this type of environment. The task can quickly become complex for experimenters, particularly on research platforms dedicated to scientific experimentation, where both hardware and software are in constant rapid evolution. In [19] we discuss the concept of the reconstructability of software environments and propose a tool, Kameleon, for dealing with this problem.

6.5. Performance Evaluation

6.5.1. *Computing the Throughput of Probabilistic and Replicated Streaming Applications*

In [8], we investigate how to compute the throughput of probabilistic and replicated streaming applications. We are given (i) a streaming application whose dependence graph is a linear chain; (ii) a one-to-many mapping of the application onto a fully heterogeneous target platform, where a processor is assigned at most one application stage, but where a stage can be replicated onto a set of processors; and (iii) a set of random variables modeling the computation and communication times in the mapping. We show how to compute the throughput of the application, i.e., the rate at which data sets can be processed, under two execution models, the Strict model where the actions of each processor are sequentialized, and the Overlap model where a processor can compute and communicate in parallel. The problem is easy when application stages are not replicated, i.e., assigned to a single processor: in that case the throughput is dictated by the critical hardware resource. However, when stages are replicated, i.e., assigned to several processors, the problem becomes surprisingly complicated: even in the deterministic case, the optimal throughput may be lower than the smallest internal resource throughput. The first contribution of the paper is to provide a general method to compute the throughput when mapping parameters are constant or follow I.I.D. exponential laws. The second contribution is to provide bounds for the throughput when stage parameters (computation and communication times) form associated random sequences, and are N.B.U.E. (New Better than Used in Expectation) variables: the throughput is bounded from below by the exponential case and bounded from above by the deterministic case. An extensive set of simulation allows us to assess the quality of the model, and to observe the actual behavior of several distributions.

6.5.2. *Optimization of Cloud Task Processing with Checkpoint-Restart Mechanism*

In [17], we explain how to optimize fault-tolerance techniques based on a checkpointing/restart mechanism, in the context of cloud computing. Our contribution is three-fold. (1) We derive a fresh formula to compute the optimal number of checkpoints for cloud jobs with varied distributions of failure events. Our analysis is not only generic with no assumption on failure probability distribution, but also attractively simple to apply in practice. (2) We design an adaptive algorithm to optimize the impact of checkpointing regarding various costs like checkpointing/restart overhead. (3) We evaluate our optimized solution in a real cluster environment with hundreds of virtual machines and Berkeley Lab Checkpoint/Restart tool. Task failure events are emulated via a production trace produced on a large-scale Google data center. Experiments confirm that our solution is fairly suitable for Google systems. Our optimized formula outperforms Young's formula by 3-10 percent, reducing wallclock lengths by 50-100 seconds per job on average.

6.6. Game Theory and Applications

6.6.1. *Fair Scheduling in Large Distributed Computing Sytems*

Fairly sharing resources of a distributed computing system between users is a critical issue that we have investigated in two ways.

Our first proposal specifically addresses the question of designing a distributed sharing mechanism. A possible answer resorts to Lagrangian optimization and distributed gradient descent. Under certain conditions, the resource sharing problem can be formulated as a global optimization problem, which can be solved by a distributed self-stabilizing demand and response algorithm. In the last decade, this technique has been applied to design network protocols (variants of TCP, multi-path network protocols, wireless network protocols) and even distributed algorithms for smart grids. In [9], we explain how to use this technique for scheduling Bag-of-Tasks (BoT) applications on a Grid since until now, only simple mechanisms have been used to ensure a fair sharing of resources amongst these applications. Although the resulting algorithm is in essence very similar to previously proposed algorithms in the context of flow control in multi-path networks, we show using carefully designed experiments and a thorough statistical analysis that the grid context is surprisingly more difficult than the multi-path network context. Interestingly, we can show that, in practice, the convergence of the algorithm is hindered by the heterogeneity of application characteristics, which is completely overlooked

in related theoretical work. Our careful investigation provides enough insights to understand the true difficulty of this approach and to propose a set of non-trivial adaptations that enable convergence in the grid context. The effectiveness of our proposal is proven through an extensive set of complex and realistic simulations.

Our second proposal is centralized but more fine grain as it does drop the steady-state hypothesis and considers sequences of campaigns. Campaign Scheduling is characterized by multiple job submissions issued from multiple users over time. The work in [18] presents a new fair scheduling algorithm called OStrich whose principle is to maintain a virtual time-sharing schedule in which the same amount of processors is assigned to each user. The completion times in the virtual schedule determine the execution order on the physical processors. Then, campaigns are interleaved in a fair way by OStrich. For independent sequential jobs, we show that OStrich guarantees the stretch of a campaign to be proportional to campaign's size and the total number of users. The theoretical performance of our solution is assessed by simulating OStrich compared to the classical FCFS algorithm, issued from synthetic workload traces generated by two different user profiles. This is done to demonstrate how OStrich benefits both types of users, in contrast to FCFS.

6.6.2. *Fundamentals of Continuous Games*

We have made the following contributions:

1. Continuous-time game dynamics are typically first order systems where payoffs determine the growth rate of the players' strategy shares. In [12], we investigate what happens beyond first order by viewing payoffs as higher order forces of change, specifying e.g., the acceleration of the players' evolution instead of its velocity (a viewpoint which emerges naturally when it comes to aggregating empirical data of past instances of play). To that end, we derive a wide class of higher order game dynamics, generalizing first order imitative dynamics, and, in particular, the replicator dynamics. We show that strictly dominated strategies become extinct in n-th order payoff-monotonic dynamics n orders as fast as in the corresponding first order dynamics; furthermore, in stark contrast to first order, weakly dominated strategies also become extinct for $n \geq 2$. All in all, higher order payoff-monotonic dynamics lead to the elimination of weakly dominated strategies, followed by the iterated deletion of strictly dominated strategies, thus providing a dynamic justification of the well-known epistemic rationalizability process of Dekel and Fudenberg. Finally, we also establish a higher order analogue of the folk theorem of evolutionary game theory, and we show that convergence to strict equilibria in n-th order dynamics is n orders as fast as in first order.

2. In [37] we introduce a new class of game dynamics made of a pay-off replicator-like term modulated by an entropy barrier which keeps players away from the boundary of the strategy space. We show that these *entropy-driven* dynamics are equivalent to players computing a score as their on-going exponentially discounted cumulative payoff and then using a quantal choice model on the scores to pick an action. This dual perspective on *entropy-driven* dynamics helps us to extend the folk theorem on convergence to quantal response equilibria to this case, for potential games. It also provides the main ingredients to design a discrete time effective learning algorithm that is fully distributed and only requires partial information to converge to QRE. This convergence is resilient to stochastic perturbations and observation errors and does not require any synchronization between the players.

6.6.3. *Application to Wireless Networks*

We have made the following contributions:

1. Starting from an entropy-driven reinforcement learning scheme for multi-agent environments, we develop in [36] a distributed algorithm for robust spectrum management in Gaussian multiple-input, multiple-output (MIMO) uplink channels. In continuous time, our approach to optimizing the transmitters' signal distribution relies on the method of matrix exponential learning, adjusted by an entropy-driven barrier term which generates a distributed, convergent algorithm in discrete time. As opposed to traditional water-filling methods, the algorithm's convergence speed can be controlled by tuning the users' learning rate; accordingly, entropy-driven learning algorithms in MIMO systems converge arbitrarily close to the optimum signal covariance profile within a few iterations (even for large numbers of users and/or antennas per user), and this convergence remains robust even in the

presence of imperfect (or delayed) measurements and asynchronous user updates.

2. Consider a wireless network of transmitter-receiver pairs where the transmitters adjust their powers to maintain a target SINR level in the presence of interference. In [46], we analyze the optimal power vector that achieves this target in large, random networks obtained by "erasing" a finite fraction of nodes from a regular lattice of transmitter-receiver pairs. We show that this problem is equivalent to the so-called Anderson model of electron motion in dirty metals which has been used extensively in the analysis of diffusion in random environments. A standard approximation to this model is the so-called coherent potential approximation (CPA) method which we apply to evaluate the first and second order intra-sample statistics of the optimal power vector in one- and two-dimensional systems. This approach is equivalent to traditional techniques from random matrix theory and free probability, but while generally accurate (and in agreement with numerical simulations), it fails to fully describe the system: in particular, results obtained in this way fail to predict when power control becomes infeasible. In this regard, we find that the infinite system is always unstable beyond a certain value of the target SINR, but any finite system only has a small probability of becoming unstable. This instability probability is proportional to the tails of the eigenvalue distribution of the system which are calculated to exponential accuracy using methodologies developed within the Anderson model and its ties with random walks in random media. Finally, using these techniques, we also calculate the tails of the system's power distribution under power control and the rate of convergence of the Foschini-Miljanic power control algorithm in the presence of random erasures.

MISTIS Project-Team

6. New Results

6.1. Mixture models

6.1.1. *Parameter estimation in the heterogeneity linear mixed model*

Participant: Marie-José Martinez.

Joint work with: Emma Holian (National University of Ireland, Galway)

In studies where subjects contribute more than one observation, such as in longitudinal studies, linear mixed models have become one of the most used techniques to take into account the correlation between these observations. By introducing random effects, mixed models allow the within-subject correlation and the variability of the response among the different subjects to be taken into account. However, such models are based on a normality assumption for the random effects and reflect the prior belief of homogeneity among all the subjects. To relax this strong assumption, Verbeke and Lesaffre (1996) proposed the extension of the classical linear mixed model by allowing the random effects to be sampled from a finite mixture of normal distributions with common covariance matrix. This extension naturally arises from the prior belief of the presence of unobserved heterogeneity in the random effects population. The model is therefore called the heterogeneity linear mixed model. Note that this model does not only extend the assumption about the random effects distribution, indeed, each component of the mixture can be considered as a cluster containing a proportion of the total population. Thus, this model is also suitable for classification purposes.

Concerning parameter estimation in the heterogeneity model, the use of the EM-algorithm, which takes into account the incomplete structure of the data, has been considered in the literature. Unfortunately, the M-step in the estimation process is not available in analytic form and a numerical maximisation procedure such as Newton-Raphson is needed. Because deriving such a procedure is a non-trivial task, Komarek et al. (2002) proposed an approximate optimization. But this procedure proved to be very slow and limited to small samples due to requiring manipulation of very large matrices and prohibitive computation.

To overcome this problem, we have proposed in [28], [52] an alternative approach which consists of fitting directly an equivalent mixture of linear mixed models. Contrary to the heterogeneity model, the M-step of the EM-algorithm is tractable analytically in this case. Then, from the obtained parameter estimates, we can easily obtain the parameter estimates in the heterogeneity model.

6.1.2. *Taking into account the curse of dimensionality*

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

Joint work with: C. Bouveyron (Univ. Paris 1), 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) [64], we propose 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 new 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.

6.1.3. Mixture modelling using skewed multivariate heavy tailed distributions with variable amounts of tailweight

Participants: Florence Forbes, Darren Wraith.

Clustering concerns the assignment of each of N, possibly multidimensional, observations $y_1, ..., y_N$ to one of K groups. A popular way to approach this task is via a parametric finite mixture model. While the vast majority of the work on such mixtures has been based on Gaussian mixture models in many applications the tails of normal distributions are shorter than appropriate or parameter estimations are affected by atypical observations (outliers). The family of location and scale mixtures of Gaussians has the ability to generate a number of flexible distributional forms. It nests as particular cases several important asymmetric distributions like the Generalised Hyperbolic distribution. The Generalised Hyperbolic distribution in turn nests many other well known distributions such as the Normal Inverse Gaussian (NIG) whose practical relevance has been widely documented in the literature. In a multivariate setting, we propose to extend the standard location and scale mixture concept into a so called multiple scaled framework which has the advantage of allowing different tail and skewness behaviours in each dimension of the variable space with arbitrary correlation between dimensions. The approach builds upon, and develops further, previous work on scale mixtures of Gaussians [25]. Estimation of the parameters is provided via an EM algorithm with a particular focus on NIG distributions. Inference is then extended to cover the case of mixtures of such multiple scaled distributions for application to clustering. Assessments on simulated and real data confirm the gain in degrees of freedom and flexibility in modelling data of varying tail behaviour and directional shape.

6.1.4. High-Dimensional Regression with Gaussian Mixtures and Partially-Latent Response Variables

Participant: Florence Forbes.

Joint work with: Antoine Deleforge and Radu Horaud from the Inria Perception team.

In this work we address the problem of approximating high-dimensional data with a low-dimensional representation. We make the following contributions. We propose an inverse regression method which exchanges the roles of input and response, such that the low-dimensional variable becomes the regressor, and which is tractable. We introduce a mixture of locally-linear probabilistic mapping model that starts with estimating the parameters of inverse regression, and follows with inferring closed-form solutions for the forward parameters of the high-dimensional regression problem of interest. Moreover, we introduce a partially-latent paradigm, such that the vector-valued response variable is composed of both observed and latent entries, thus being able to deal with data contaminated by experimental artifacts that cannot be explained with noise models. The proposed probabilistic formulation could be viewed as a latent-variable augmentation of regression. We devise expectation-maximization (EM) procedures based on a data augmentation strategy which facilitates the maximum-likelihood search over the model parameters. We propose two augmentation schemes and we describe in detail the associated EM inference procedures that may well be viewed as generalizations of a number of EM regression, dimension reduction, and factor analysis algorithms. The proposed framework is validated with both synthetic and real data. We provide experimental evidence that our method outperforms several existing regression techniques.

6.1.5. Acoustic space learning via variational EM for Sound-Source Separation and Localization

Participant: Florence Forbes.

Joint work with: Antoine Deleforge and Radu Horaud from the Inria Perception team.

In this paper we address the problems of modeling the acoustic space generated by a full-spectrum sound source and of using the learned model for the localization and separation of multiple sources that simultaneously emit sparse-spectrum sounds. We lay theoretical and methodological grounds in order to introduce the *binaural manifold* paradigm. We perform an in-depth study of the latent low-dimensional structure of the high-dimensional interaural spectral data, based on a corpus recorded with a human-like audiomotor robot head. A

non-linear dimensionality reduction technique is used to show that these data lie on a two-dimensional (2D) smooth manifold parameterized by the motor states of the listener, or equivalently, the sound source directions. We propose a *probabilistic piecewise affine mapping* model (PPAM) specifically designed to deal with high-dimensional data exhibiting an intrinsic piecewise linear structure. We derive a closed-form expectation-maximization (EM) procedure for estimating the model parameters, followed by Bayes inversion for obtaining the full posterior density function of a sound source direction. We extend this solution to deal with missing data and redundancy in real world spectrograms, and hence for 2D localization of natural sound sources such as speech. We further generalize the model to the challenging case of multiple sound sources and we propose a variational EM framework. The associated algorithm, referred to as *variational EM for source separation and localization* (VESSL) yields a Bayesian estimation of the 2D locations and time-frequency masks of all the sources. Comparisons of the proposed approach with several existing methods reveal that the combination of acoustic-space learning with Bayesian inference enables our method to outperform state-of-the-art methods.

6.2. Statistical models for Neuroscience

6.2.1. *Hemodynamically informed parcellation of cerebral fMRI data*

Participants: Florence Forbes, Aina Frau-Pascual, Thomas Vincent.

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

Standard detection of evoked brain activity in functional MRI (fMRI) relies on a fixed and known shape of the impulse response of the neurovascular coupling, namely the hemodynamic response function (HRF). To cope with this issue, the joint detection-estimation (JDE) framework has been proposed. This formalism enables to estimate a HRF per region but for doing so, it assumes a prior brain partition (or parcellation) regarding hemodynamic territories (eg. [14]). This partition has to be accurate enough to recover accurate HRF shapes but has also to overcome the detection-estimation issue: the lack of hemodynamics information in the non-active positions. During the internship of A. Frau Pascual at Neurospin, we proposed an hemodynamically-based parcellation, consisting first of a feature extraction step, followed by a Gaussian Mixture-based parcellation, which considers the injection of the activation levels in the parcellation process, in order to overcome the detection-estimation issue and find the underlying hemodynamics. The work has been submitted to the ICASSP conference in 2014.

6.2.2. *Variational variable selection to assess experimental condition relevance in event-related fMRI*

Participants: Florence Forbes, Christine Bakhous, Lotfi Chaari, Thomas Vincent, Farida Enikeeva.

Joint work with: Michel Dojat (Grenoble Institute of Neuroscience) and Philippe Ciuciu from Neurospin, CEA in Saclay.

Brain functional exploration investigates the nature of neural processing following cognitive or sensory stimulation. This goal is not fully accounted for in most functional Magnetic Resonance Imaging (fMRI) analysis which usually assumes that all delivered stimuli possibly generate a BOLD response everywhere in the brain although activation is likely to be induced by only some of them in specific brain regions. Generally, criteria are not available to select the relevant conditions or stimulus types (e.g. visual, auditory, etc.) prior to activation detection and the inclusion of irrelevant events may degrade the results, particularly when the Hemodynamic Response Function (HRF) is jointly estimated as in the JDE framework mentioned in the previous section. To face this issue, we propose an efficient variational procedure that automatically selects the conditions according to the brain activity they elicit. It follows an improved activation detection and local HRF estimation that we illustrate on synthetic and real fMRI data. This approach is an alternative to our previous approach based on Monte-Carlo Markov Chain (MCMC) inference [63]. Corresponding papers [31], [45]. A synthesis can also be found in the PhD manuscript of C. Bakhous (Grenoble University, December 2013) [11].

6.2.3. *Bayesian Joint Detection-Estimation of cerebral vasoreactivity from ASL fMRI data*

Participants: Florence Forbes, Thomas Vincent.

In the context of ARC AINSI project, joint work with: Philippe Ciuciu from Neurospin, CEA in Saclay.

Functional MRI (fMRI) is the method of choice to non-invasively probe cerebral activity evoked by a set of controlled experimental conditions. A rising fMRI modality is Arterial Spin Labeling (ASL) which enables to quantify the cerebral perfusion, namely the cerebral blood flow (CBF) and emerges as a more direct biomarker of neuronal activity than the standard BOLD (Blood Oxygen Level Dependent) fMRI.

Although the study of cerebral vasoreactivity using fMRI is mainly conducted through the BOLD fMRI modality (see the two previous sections), owing to its relatively high signal-to-noise ratio (SNR), ASL fMRI provides a more interpretable measure of cerebral vasoreactivity than BOLD fMRI. Still, ASL suffers from a low SNR and is hampered by a large amount of physiological noise. Our contribution, described in [43], [44] aims at improving the recovery of the vasoreactive component from the ASL signal. To this end, a Bayesian hierarchical model is proposed, enabling the recovery of perfusion levels as well as fitting their dynamics. On a single-subject ASL real data set involving perfusion changes induced by hypercapnia, the approach is compared with a classical GLM-based analysis. A better goodness-of-fit is achieved, especially in the transitions between baseline and hypercapnia periods. Also, perfusion levels are recovered with higher sensitivity and show a better contrast between gray- and white matter.

6.2.4. *Physiologically-inspired Bayesian analysis of BOLD and ASL fMRI data*
Participants: Florence Forbes, Thomas Vincent, Jennifer Sloboda.

In the context of ARC AINSI project, joint work with: Philippe Ciuciu from Neurospin, CEA in Saclay.

The ASL modality is most commonly used as a static measure where the average perfusion is computed over a volume sequence lasting several minutes. Recently, ASL has been used in functional activation protocols and hence gives access to a dynamic measure of perfusion, namely the variations of CBF which are elicited by specific tasks. ASL MRI mainly consists of acquiring pairs of control and label images and looking at the average control-label difference. The Signal-to-Noise Ratio (SNR) of this difference is very low so that several hundreds of image pairs need to be acquired, thus increasing significantly the time spent by the subject in the scanner and making the acquisition very sensitive to the patient's movement. In addition, this averaging requires that the perfusion signal is at a steady state, limiting the scope of fMRI task experiments to baseline perfusion measurements or long block designs. In contrast, it is highly desirable to measure change in perfusion due to an effect of interest in activation paradigms from event-related designs. It is technically possible to collect event-related ASL data but most approaches to functional ASL data analysis use a standard linear model (GLM-based) formulation with regressors encoding differences in control/tag scans and both ASL and BOLD activation signals being associated with the same canonical response function. The canonical hemodynamic response function (HRF) is generally used although it has been been calibrated on BOLD experiments only, thus reflecting simultaneous variations of CBF, cerebral blood volume (CBV) and cerebral oxygen consumption (CMRO2). In contrast, the perfusion signal only reflects variation in CBF so that the associated response, the perfusion response function (PRF), is likely to differ from the HRF. In the internship proposal of Jennifer Sloboda, we proposed to recover both a hemodynamic (BRF for BOLD response function) and a perfusion (PRF) response functions from event-related functional ASL data. To do so, a joint detection estimation (JDE) formalism was used. In the BOLD context, the JDE framework has proven to successfully extract the HRF while also performing activation detection. We had recently extended this formalism (see Section 6.2.3 and [43], [44]) to model an additional perfusion component linked to the BOLD one through a common activation detection. The main issue addressed then was to characterize the link between BOLD and perfusion components. To establish this link, we proposed a methodological axis which consists of developing a physiologically-inspired approach. To do so, dynamical non-linear equations available in physiological models were linearized and approximated in a parsimonious way so as to establish prior relations between the perfusion and BOLD responses which can be injected in our Bayesian setting. The inference of the initial model is currently done through a Markov Chain Monte Carlo approach but a Variational Expectation-Maximization implementation is also conceivable. As such, the tasks were two-fold: (1) investigate the physiological model and (2) inject it into the JDE setting. Investigation of the physiological model allows for: (1) creation of artificial fMRI data, (2) investigation of the relationship between physiological changes

and the resulting simulated BOLD or ASL signal, and (3) characterization of the link between BOLD and perfusion responses. Injection of the physiologically inspired prior into the JDE model, is to (1) improve perfusion response recovery, (2) determine physiologically quantified units to the JDE recovered values This work is going to serve as a preliminary investigation into the incorporation of physiological information in the Bayesian JDE setting from which to determine the trajectory of future model developments.

6.3. Markov models

6.3.1. *Spatial modelling of plant diversity from high-throughput environmental DNA sequence data*

Participants: Florence Forbes, Angelika Studeny.

This is joint work with: Eric Coissac and Pierre Taberlet from LECA (Laboratoire d'Ecologie Alpine) and Alain Viari from Inria team Bamboo.

This work [48] considers a statistical modelling approach to investigate spatial cross-correlations between species in an ecosystem. A special feature is the origin of the data from high-troughput environmental DNA sequencing of soil samples. Here we use data collected at the Nourague CNRS Field Station in French Guiana. We describe bivariate spatial relationships in these data by a separable linear model of coregionalisation and estimate a cross-correlation parameter. Based on this estimate, we visualise plant taxa co-occurrence pattern in form of 'interaction graphs' which can be interpreted in terms of ecological interactions. Limitations of this approach are discussed along with possible alternatives in [48].

6.3.2. *Modelling multivariate counts with graphical Markov models.*

Participants: Jean-Baptiste Durand, Florence Forbes, Marie-José Martinez, Angelika Studeny.

Joint work with: Pierre Fernique (Montpellier 2 University, CIRAD and Inria Virtual Plants), Yann Guédon (CIRAD and Inria Virtual Plants) and Iragaël Joly (INRA-GAEL and Grenoble INP).

Multivariate count data are defined as the number of items of different categories issued from sampling within a population, which individuals are grouped into categories. The analysis of multivariate count data is a recurrent and crucial issue in numerous modelling problems, particularly in the fields of biology and ecology (where the data can represent, for example, children counts associated with multitype branching processes), sociology and econometrics. Denoting by K the number of categories, multivariate count data analysis relies on modelling the joint distribution of the K-dimensional random vector $N = (N_0, ..., N_{K-1})$ with discrete components. Our work focused on I) Identifying categories that appear simultaneously, or on the contrary that are mutually exclusive. This was achieved by identifying conditional independence relationships between the K variables; II) Building parsimonious parametric models consistent with these relationships; III) Characterizing and testing the effects of covariates on the distribution of N, particularly on the dependencies between its components.

Our context of application was characterised by zero-inflated, often right skewed marginal distributions. Thus, Gaussian and Poisson distributions were not *a priori* appropriate. Moreover, the multivariate histograms typically had many cells, most of which were empty. Consequently, nonparametric estimation was not efficient.

To achieve these goals, we proposed an approach based on graphical probabilistic models (Koller & Friedman, 2009 [70]) to represent the conditional independence relationships in N, and on parametric distributions to ensure model parsimony [51]. The considered graphs were partially directed, so as to represent both marginal independence relationships and cyclic dependencies between quadruplets of variables (at least).

Graph search was achieved by a stepwise approach, issued from unification of previous algorithms presented in Koller & Friedman (2009) for DAGs: Hill climbing, greedy search, first ascent and simulated annealing. The search algorithm was improved by taking into account our parametric distribution assumptions, which led to caching overlapping graphs at each step. An adaptation to PDAGs of graph search algorithms for DAGs was developed, by defining new operators specific to PDAGs.

Comparisons between different algorithms in the literature for directed and undirected graphical models was performed on simulated datasets to: (i) Assess gain in speed induced by caching; (ii) Compare the graphs obtained under parametric and nonparametric distributions assumptions; (iii) Compare different strategies for graph initialization. Strategies based on several random graphs were compared to those based on a fast estimation of an undirected graph, assumed to be the moral graph.

First results were obtained in modelling individual daily activity program [50] and interactions between flowering and vegetative growth in plants (see sections below).

6.3.3. *Statistical characterization of tree structures based on Markov tree models and multitype branching processes, with applications to tree growth modelling.*

Participant: Jean-Baptiste Durand.

Joint work with: Pierre Fernique (Montpellier 2 University and CIRAD) and Yann Guédon (CIRAD), Inria Virtual Plants.

The quantity and quality of yields in fruit trees is closely related to processes of growth and branching, which determine ultimately the regularity of flowering and the position of flowers. Flowering and fruiting patterns are explained by statistical dependence between the nature of a parent shoot (*e.g.* flowering or not) and the quantity and natures of its children shoots – with potential effect of covariates. Thus, better characterization of patterns and dependences is expected to lead to strategies to control the demographic properties of the shoots (through varietal selection or crop management policies), and thus to bring substantial improvements in the quantity and quality of yields.

Since the connections between shoots can be represented by mathematical trees, statistical models based on multitype branching processes and Markov trees appear as a natural tool to model the dependencies of interest. Formally, the properties of a vertex are summed up using the notion of vertex state. In such models, the numbers of children in each state given the parent state are modeled through discrete multivariate distributions. Model selection procedures are necessary to specify parsimonious distributions. We developed an approach based on probabilistic graphical models (see Section 6.3.2) to identify and exploit properties of conditional independence between numbers of children in different states, so as to simplify the specification of their joint distribution [51], [32].

This work was carried out in the context of Pierre Fernique's first year of PhD (Montpellier 2 University and CIRAD). It was applied to model dependencies between short or long, vegetative or flowering shoots in apple trees. The results highlighted contrasted patterns related to the parent shoot state, with interpretation in terms of alternation of flowering (see paragraph 6.3.4). It was also applied to the analysis of the connections between cyclic growth and flowering of mango trees [32]. This work will be continued during Pierre Fernique's PhD thesis, with extensions to other fruit tree species and other parametric discrete multivariate families of distributions, including covariates and mixed effects.

6.3.4. *Statistical characterization of the alternation of flowering in fruit tree species*

Participant: Jean-Baptiste Durand.

Joint work with: Jean Peyhardi and Yann Guédon (Mixed Research Unit DAP, Virtual Plants team), Baptiste Guitton, Yan Holtz and Evelyne Costes (DAP, AFEF team), Catherine Trottier (Montpellier University)

A first study was performed to characterize genetic determinisms of the alternation of flowering in apple tree progenies [37], [21]. Data were collected at two scales: at whole tree scale (with annual time step) and a local scale (annual shoot or AS, which is the portions of stem that were grown during the same year). Two replications of each genotype were available.

Indices were proposed to characterize alternation at tree scale. The difficulty is related to early detection of alternating genotypes, in a context where alternation is often concealed by a substantial increase of the number of flowers over consecutive years. To separate correctly the increase of the number of flowers due to aging of young trees from alternation in flowering, our model relied on a parametric hypothesis for the trend (fixed slopes specific to genotype and random slopes specific to replications), which translated into mixed effect

modelling. Then, different indices of alternation were computed on the residuals. Clusters of individuals with contrasted patterns of bearing habits were identified.

To model alternation of flowering at AS scale, a second-order Markov tree model was built. Its transition probabilities were modelled as generalized linear mixed models, to incorporate the effects of genotypes, year and memory of flowering for the Markovian part, with interactions between these components.

Asynchronism of flowering at AS scale was assessed using an entropy-based criterion. The entropy allowed for a characterisation of the roles of local alternation and asynchronism in regularity of flowering at tree scale.

Moreover, our models highlighted significant correlations between indices of alternation at AS and individual scales.

This work was extended by the Master 2 internship of Yan Holtz, supervised by Evelyne Costes and Jean-Baptiste Durand. New progenies were considered, and a methodology based on a lighter measurement protocol was developed and assessed. It consisted in assessing the accuracy of approximating the indices computed from measurements at tree scale by the same indices computed as AS scale. The approximations were shown sufficiently accurate to provide an operational strategy for apple tree selection.

As a perspective of this work, patterns in the production of children ASs (numbers of flowering and vegetative children) depending on the type of the parent AS must be analyzed using branching processes and different types of Markov trees, in the context of Pierre Fernique's PhD Thesis (see paragraph 6.3.3).

6.4. Semi and non-parametric methods

6.4.1. *Modelling extremal events*

Participants: Stéphane Girard, El-Hadji Deme.

Joint work with: L. Gardes (Univ. Strasbourg) and E. Deme (Univ. Gaston Berger, Sénégal)

We are working on the estimation of the second order parameter ρ (see paragraph 3.3.1). We proposed a new family of estimators encompassing the existing ones (see for instance [69], [68]). This work is in collaboration with El-Hadji Deme who obtained a grant (IBNI price) to work within the Mistis team on extreme-value statistics. The results are published in [18].
In addition to this work, we have established a review on the Weibull-tail distributions [29].

6.4.2. *Conditional extremal events*

Participants: Stéphane Girard, Gildas Mazo, Jonathan El-Methni.

Joint work with: L. Gardes (Univ. Strasbourg) and A. Daouia (Univ. Toulouse I and Univ. Catholique de Louvain)

The goal of the PhD thesis of Alexandre Lekina 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 [66] with extreme-value methods in order to obtain efficient estimators of the conditional tail index and conditional extreme quantiles. When the covariate is functional and random (random design) we focus on kernel methods [16].

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.

6.4.3. *Estimation of extreme risk measures*

Participants: Stéphane Girard, Jonathan El-Methni, El-Hadji Deme.

Joint work with: L. Gardes and A. Guillou (Univ. Strasbourg)

One of the most popular risk measures is the Value-at-Risk (VaR) introduced in the 1990's. In statistical terms, the VaR at level $\alpha \in (0, 1)$ corresponds to the upper α-quantile of the loss distribution. The Value-at-Risk however suffers from several weaknesses. First, it provides us only with a pointwise information: VaR(α) does not take into consideration what the loss will be beyond this quantile. Second, random loss variables with light-tailed distributions or heavy-tailed distributions may have the same Value-at-Risk . Finally, Value-at-Risk is not a coherent risk measure since it is not subadditive in general. A coherent alternative risk measure is the Conditional Tail Expectation (CTE), also known as Tail-Value-at-Risk, Tail Conditional Expectation or Expected Shortfall in case of a continuous loss distribution. The CTE is defined as the expected loss given that the loss lies above the upper α-quantile of the loss distribution. This risk measure thus takes into account the whole information contained in the upper tail of the distribution. It is frequently encountered in financial investment or in the insurance industry. In [36], we have established the asymptotic properties of the classical 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 [36]. Similar developments have been achieved in the case of the Proportional Hazard Premium measure of risk [19]. In [22], we study the situation where some covariate information is available. We thus has to deal with conditional extremes (see paragraph 6.4.2). We also proposed a new risk measure (called the Conditional Tail Moment) which encompasses various risk measures, such as the CTE, as particular cases.

6.4.4. *Multivariate extremal events*

Participants: Stéphane Girard, Gildas Mazo, Florence Forbes, Van Trung Pham.

Joint work with: C. Amblard (TimB in TIMC laboratory, Univ. Grenoble I) and L. Menneteau (Univ. Montpellier II)

Copulas are a useful tool to model multivariate distributions [72]. At first, we developed an extension of some particular copulas [1]. It followed a new class of bivariate copulas defined on matrices [56] and some analogies have been shown between matrix and copula properties.

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 [61]. 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. A pairwise moment-based inference procedure has also been proposed and the asymptotic normality of the corresponding estimator has been established [53].

6.4.5. *Level sets estimation*

Participant: Stéphane Girard.

Joint work with: A. Guillou and L. Gardes (Univ. Strasbourg), G. Stupfler (Univ. Aix-Marseille) and A. Daouia (Univ. Toulouse I and Univ. Catholique de Louvain)

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 [60] and thus should not be used to detect outliers.

In collaboration with A. Daouia, we investigate the application of such methods in econometrics [17]: A new characterization of partial boundaries of a free disposal multivariate support is introduced by making use of large quantiles of a simple transformation of the underlying multivariate distribution. Pointwise empirical and smoothed estimators of the full and partial support curves are built as extreme sample and smoothed quantiles. The extreme-value theory holds then automatically for the empirical frontiers and we show that some fundamental properties of extreme order statistics carry over to Nadaraya's estimates of upper quantile-based frontiers.

In collaboration with G. Stupfler and A. Guillou, new estimators of the boundary are introduced. The regression is performed on the whole set of points, the selection of the "highest" points being automatically performed by the introduction of high order moments [26], [27].

6.4.6. *Retrieval of Mars surface physical properties from OMEGA hyperspectral images.*

Participants: Stéphane Girard, Alessandro Chiancone.

Joint work with: S. Douté from Laboratoire de Planétologie de Grenoble, J. Chanussot (Gipsa-lab and Grenoble-INP) and J. Saracco (Univ. Bordeaux).

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 (ill-conditioned problems) [57]. We have also defined an adaptive version of the method which is able to deal with block-wise evolving data streams [15].

6.4.7. *High-dimensional change-point detection with sparse alternatives*

Participant: Farida Enikeeva.

Joint work with: Zaid Harchaoui from LEAR team Inria Grenoble

The change-point problem is a classical problem of statistics that arises in various applications as signal processing, bioinformatics, financial market analysis. The goal of change-point problems is to make an inference about the moment of a change in the distribution of the observed data. We consider the problem of detection of a simultaneous change in mean in a sequence of Gaussian vectors.

The state-of-the-art approach to the change-point detection/estimation is based on the assumption of growing number of observations and fixed dimension of the signal. We work in high-dimensional setting assuming that the vector dimension tends to infinity and the length of the sequence grows slower than the dimension of the signal. Assuming that the change occurs only in a subset of the vector components of unknown cardinality we can reduce our problem to the problem of testing non-zero components in a sequence of sparse Gaussian vectors. We construct a testing procedure that is adaptive to the number of components with a change. This testing procedure is based on combination of two chi-squared type test statistics. This combined test provides an optimal performance of the test both in the cases of high and moderate sparsity. We obtain the detection boundary of the test and show its rate-optimality in minimax sense.

The results of the paper [59] were presented at

- NIPS 2013, Workshop on Modern Nonparametric Methods in Machine Learning (Dec. 2013)
- Conference on Structural Inference in Statistics, Potsdam, Germany (Sept. 2013)

6.4.8. *Yield Improvement by the Redundancy Method for Component Calibration*
Participant: Farida Enikeeva.

Joint work with: Dominique Morche (CEA-LETI) and Alp Oguz (CEA-LETI)

This work [23] was done in the framework of the Optimyst II project of MINALOGIC in collaboration with CEA-LETI and LJK-UJF. In this project we explore the benefits of the redundant channels methodology for the calibration of electronic components.

The demand for high data rate in communication puts stringent requirements on components' dynamic range. However, the extreme size reduction in advanced technology results inadvertently in increased process variability, which inherently limits the performances. The redundancy approach is based on the idea of dividing an elementary component (capacitor, resistor, transistor) into several subsets and then choosing an optimal combination of such subsets to provide the production of a component with very precise characteristics. For several years, the redundancy method has been identified as complementary to digital calibration to improve the performances. On practice, it is hard for a designer to select an optimal number of redundant components to provide the desired production yield and to minimize the area occupied by the components. The usual way to solve this problem is to resort to statistical simulations which are time consuming and sometimes misleading. We propose a normal approximation of the yield in order to estimate the number of redundant components needed to provide a minimal area occupied by the components.

MOAIS Project-Team

6. New Results

6.1. Distributed Art Performance

Moais collaborated with partners fromI2cat, Barcelona, Psnc, Poznan and Grenoble-INP to setup a live distributed art performance for the ICT 2013 conference at Vilnius. This distributed performance gathered musicians located a Poznan, Barcelona and Vilnius, as well as a dancer modeled in 3D on the Grimage platform at Inria Grenoble. Though physically present in different cities these artists performed together for a numerical dance and music performance numerically assembled and transmitted in real-time at Vilnius. This joint effort relies on the FlowVR framework from Moais and the UltraGrid software from CESNET. This event received a significant attention from the medias (In France: FR3 and Tele-grenoble, France inter, etc.). A video is available at http://cyan1.grenet.fr/podcastmedia/Visionair/ICT2013_promo.m4v.

6.2. VTK Parallelization Framework

Moais developed a framework for the parallelization of scientific visualization algorithms based on on-demand task extraction and work stealing techniques. This work is developed for the VTK software and supports the OpenMP, Intel TBB and Kaapi runtime environments. Mathias Ettinger visited the Kitware company, NY, for two months to prepare the integration of this work in the next release of VTK. This work is performed in collaboration with the EDF company.

6.3. Parallel Sorting Algorithm

We developed a novel adaptive sorting algorithm, called PaVO, relying on a Packed Mermory Array data structure. Maintaining gaps in the array of elements enable to reduce the span of modifications needed when reordering elements. This is particularly relevant in a parallel context to reduce the data dependencies. Performance results on a NUMA architecute show that PaVo outperfoms standard parallel sorting algorithms even for a large amount of disorder.

6.4. High bandwidth IPSec gateways and ICMP

Internet Control Message Protocol (ICMP)i is essential for performance aspects in particular for Path Maximum Transmission Unit discovery but is also known to be a cause of attacks. In collaboration with Planet, we demonstrate, through a real exploit on a testbed, that an external attacker having eavesdropping and traffic injection capabilities in the black untrusted network, without any access to clear-text (thesis of Ludovic Jacquin). This impacts out current research on trusted outsourced computations.

6.5. Efficient Parallel multi-GPUs execution

We developed a novel scheduling algorithm in Kaapi to perform multi-GPUs execution of task' based program [19]. Performance results on Cholesky factorization on up to 8-GPUs shows that Kaapi outperfoms similar runtime systems and even hand code parallel version.

6.6. Porting Kaapi for Native Mode on Intel Xeon Phi

Kaapi was ported natively on Intel Xeon Phi co-processor. Specific memory hierarchy was managed transparently to the applicaiton by the development of specific hierarchical work stealing scheduler. Experimentations on dense linear algebra kernels (Cholesky, LU and QR factorization) shows a very promising gain compared to the standard parallel implementation available in the Intel MKL [16]. Extension of theses resultats are under publication process.

6.7. Adaptive loop scheduling in GCC OpenMP runtime library

We port an adaptive loop scheduler from Kaapi into the OpenMP runtime library of GCC called libGOMP [12]. The loop scheduler is consencious of the bloc data mapping to improve locality of computation.

6.8. Kaapi in EPX standard distribution

Kaapi software developped by the MOAIS team was included in the standard EPX distribution. EPX has won the 2013 Grand Prix SFEN (http://www-epx.cea.fr).

MOISE Project-Team

6. New Results

6.1. Mathematical Modelling of the Ocean Dynamics

6.1.1. *Coupling Methods for Oceanic and Atmospheric Models*

Participants: Eric Blayo, Mehdi-Pierre Daou, Laurent Debreu, Florian Lemarié, Antoine Rousseau, Manel Tayachi.

6.1.1.1. *Interface conditions for coupling ocean models*

Many physical situations require coupling two models with not only different resolutions, but also different physics. Such a coupling can be studied within the framework of global-in-time Schwarz methods. However, the efficiency of these iterative algorithms is strongly dependent on interface conditions. As a first step towards coupling a regional scale primitive equations ocean model with a local Navier-Stokes model, a study on the derivation of interface conditions for 2-D $x - z$ Navier-Stokes equations has been performed in D. Cherel PhD thesis. It has been shown theoretically that several usual conditions lead to divergent algorithms, and that a convergent algorithm is obtained when using transmission conditions given by a variational calculation.

D. Cherel has implemented a Schwarz-based domain decomposition method, for which he developed optimized absorbing boundary conditions that mix the velocity and pressure variables on an Arakawa-C grid. The numerical results confirm the rate of convergence that has been obtained theoretically, thanks to a Fourier analysis of the semi-discretized problem. New optimized conditions offer much better convergence than classical Dirichlet-Dirichlet conditions, even when domains overlap. A paper is now ready for submission.

6.1.1.2. *Coupling dimensionally heterogeneous models*

The coupling of different types of models is gaining more and more attention recently. This is due, in particular, to the needs of more global models encompassing different disciplines (*e.g.* multi-physics) and different approaches (*e.g.* multi-scale, nesting). Also, the possibility to assemble different modeling units inside a friendly modelling software platform is an attractive solution compared to developing more and more global complex models. More specifically one may want to couple 1D to 2D or 3D models, such as Shallow Water and Navier Stokes models: this was the framework of our partnership with EDF in the project MECSICO, now extended with ARTELIA Group. In her PhD (defended Oct. 28th, 2013) [4], M. Tayachi built a theoretical and numerical framework to couple 1D, 2D and 3D models for river flows.

In [103] (now accepted for publication), we propose and analyze an efficient iterative coupling method for a dimensionally heterogeneous problem. We consider the case of a 2-D Laplace equation with non symmetric boundary conditions with a corresponding 1-D Laplace equation. We first show how to obtain the 1-D model from the 2-D one by integration along one direction, by analogy with the link between shallow water equations and the Navier-Stokes system. Then we focus on the design of a Schwarz-like iterative coupling method. We discuss the choice of boundary conditions at coupling interfaces. We prove the convergence of such algorithms and give some theoretical results related to the choice of the location of the coupling interface, and to the control of the difference between a global 2-D reference solution and the 2-D coupled one. These theoretical results are illustrated numerically. The extension of this work to shallow water equations has been started in 2013 with the PhD thesis of Medhi Pierre Daou (funded by ARTELIA). An extension to primitive equations is envisaged: a post-doc position has been proposed in 2013 (not funded) and will renewed in 2014.

6.1.1.3. 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 approached one [75]. This finding has motivated a deep numerical analysis of multi-physics coupling problems, first on simplified academic cases based on diffusion equations. In this context, Schwarz-like iterative domain decomposition methods have been analyzed and efficient interface conditions have been determined to optimize the convergence rate of the method [19], [20], [53], [79]. This method has then been applied to the coupling of realistic oceanic and atmospheric models to simulate the propagation of a tropical cyclone (cyclone erica, Fig. 1). Sensitivity tests to the coupling method have been been carried out in an ensemblist approach. We showed that with a mathematically consistent coupling, compared to coupling methods en vogue in existing coupled models, the spread of the ensemble is reduced, thus indicating a much reduced uncertainty in the physical solution [24], [75].

The next step is now to complete the theoretical work done on a diffusion problem by including the formulation of physical parameterizations to tackle a problem more representative of the realistic models: a PhD thesis should start on this subject in fall 2014. In parallel, an important perspective is to assess the impact of our work on IPCC-like climate models, this task will be initiated in 2014 through a collaboration between the MOISE project-team and the LSCE (Laboratoire des Sciences du Climat et de l'Environnement).

In collaboration with geophysicists, number of studies to investigate important small-scales air-sea feedbacks are in progress [45]. Through those studies, the aim will be to mathematically derive a metamodel able to represent important processes in the marine atmospheric boundary layers. The medium term objective will be to use this 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.

6.1.2. Numerical schemes for ocean modelling

Participants: Laurent Debreu, Jérémie Demange, Florian Lemarié.

Reducing the traditional errors in terrain-following vertical coordinate ocean models (or sigma models) has been a focus of interest for the last two decades. The objective is to use this class of model in regional domains which include not only the continental shelf, but the slope and deep ocean as well. Two general types of error have been identified: 1) the pressure-gradient error and 2) spurious diapycnal diffusion associated with steepness of the vertical coordinate. In ([92],[46]) we have studied the problem of diapycnal mixing. The solution to this problem requires a specifically designed advection scheme. We propose and validate a new scheme, where diffusion is split from advection and is represented by a rotated biharmonic diffusion scheme with flow-dependent hyperdiffusivity. The main numerical development was to render the biharmonic diffusion operator scheme unconditionally stable. This is particularly needed when the slopes between coordinates lines and isopycnal surfaces are important so that the rotation of the biharmonic leads to strong stability condition along the vertical coordinate where the grid size is relatively small. This work also extends more classical results on the stability of Laplacian diffusion with mixed derivatives.

In his PhD, Jérémie Demange works on advection-diffusion schemes for ocean models (Supervisors : L. Debreu, P. Marchesiello (IRD)). His work focuses on the link between tracers (temperature and salinity) and momentum advection and diffusion in the non hyperbolic system of equations typically used in ocean models (the so called primitive equations with hydrostatic and Boussinesq assumptions). We also investigated the use of a depth dependent barotropic mode in free surface ocean models. When most ocean models assume that this mode is vertically constant, we have shown that the use of the true barotropic mode, derived from a normal mode decomposition, allows more stability and accuracy in the representation of external gravity waves.

Salinity at 1000 m in the Southwest Pacific ocean is shown in figure 2 . The use of traditional upwind biased schemes (middle) exhibits a strong drift in the salinity field in comparison with climatology (left). The introduction of high order diffusion rotated along geopotential surfaces prevents this drift while maintaining high resolution features (right).

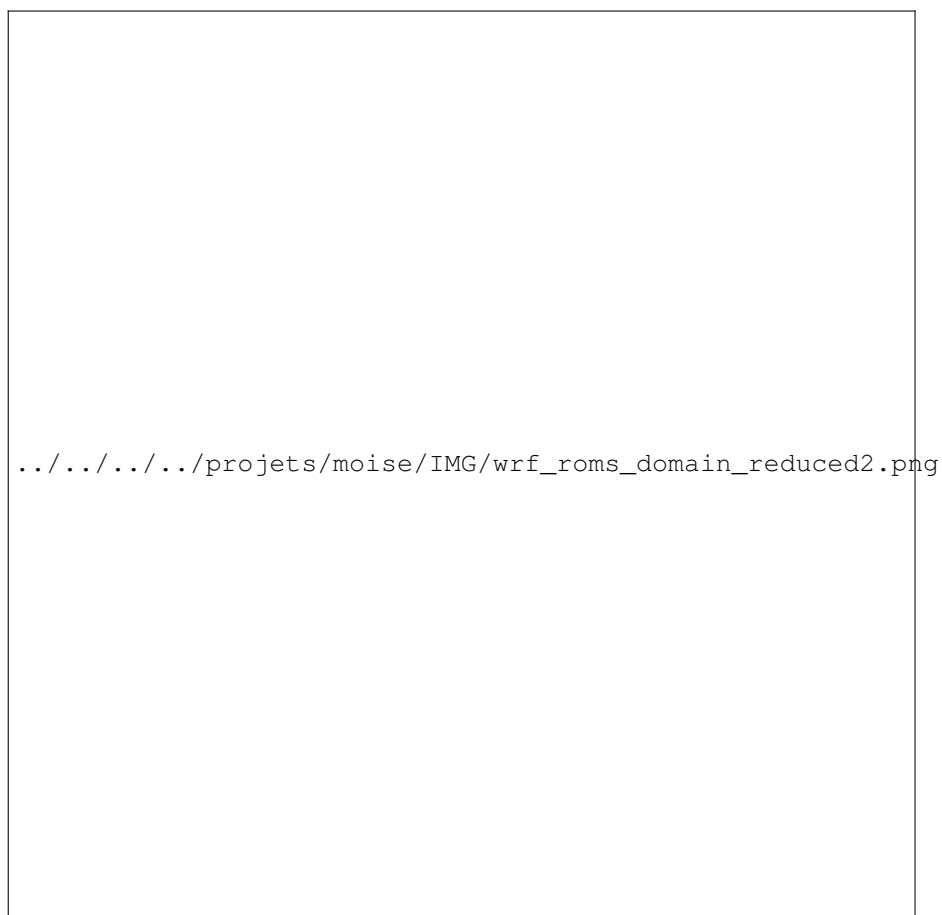../../../../projets/moise/IMG/wrf_roms_domain_reduced2.png

Figure 1. Snapshots (March 12, 2003 at 8 p.m. GMT) of (a) oceanic sea surface temperature (b) atmospheric 10 meter winds during a coupled simulation.
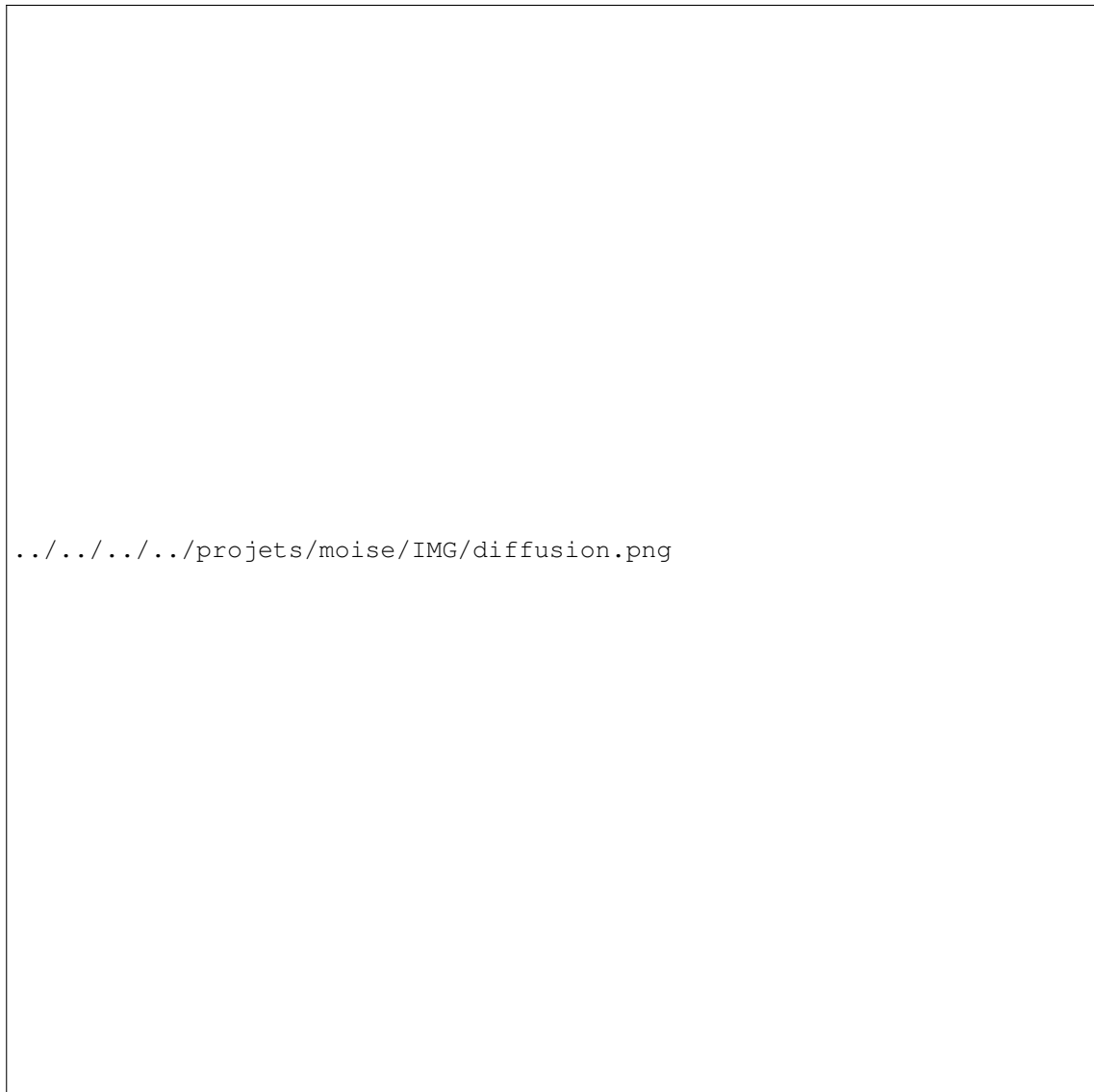
../../../../projets/moise/IMG/diffusion.png

Figure 2. Salinity at 1000m in the Southwest Pacific ocean.

6.2. Data Assimilation for Geophysical Models

6.2.1. *Development of a Variational Data Assimilation System for OPA9/NEMO*

Participants: Arthur Vidard, Bénédicte Lemieux-Dudon, Pierre-Antoine Bouttier, Eric Blayo.

We are heavily involved in the development of NEMOVAR (Variational assimilation for NEMO). For several years now, we built a working group (coordinated by A. Vidard) in order to bring together various NEMOVAR user-groups with diverse scientific interests (ranging from singular vector and sensitivity studies to specific issues in variational assimilation). It has led to the creation of the VODA (Variational Ocean Data Assimilation for multi scales applications) ANR project (ended in 2012). A new project, part of a larger EU-FP7 project (ERA-CLIM2) has been submitted late 2012 and will start early 2014.

The project aims at delivering a common NEMOVAR platform based on NEMO platform for 3D and 4D variational assimilation. Following 2009-11 VODA activities, a fully parallel version of NEMOTAM (Tangent and Adjoint Model for NEMO) is now available for the community in the standard NEMO version. This version is based on the released 3.4.1 version of NEMO.

We are also investigating variational data assimilation methods applied to high resolution ocean numerical models (see figure 3). This part of the project is now well advanced and encouraging preliminary results are available on an idealised numerical configuration of an oceanic basin. Several novative diagnostics have been also developed in this framework as part of P.A. Bouttier's PhD that will be defended early 2014

Lastly, multi resolution algorithms have been developed to solve the variational problem, and preliminary results were presented in two international communications [52], [51].

6.2.2. *Ensemble Kalman filtering for large scale ice-sheet models*

Participants: Bertrand Bonan, Maëlle Nodet, Catherine Ritz.

In collaboration with C. Ritz (CNRS, Laboratoire de Glaciologie et Geophysique de l'Environnement (LGGE), Grenoble), we aim to develop inverse methods for ice cap models.

In the framework of global warming, the evolution of sea level is a major but ill-known phenomenon. It is difficult to validate the models which are used to predict the sea level elevation, because observations are heterogeneous and sparse.

Data acquisition in polar glaciology is difficult and expensive. Satellite data have a good spatial coverage, but they allow only indirect observation of the interesting data. Moreover, ice dynamics processes are highly non linear and involve many feedback loops, so that classical linear data assimilation give poor results.

B. Bonan defended his PhD [1] in November 2013 on this subject. We implemented the Ensemble Transform Kalman Filter (ETKF) algorithm for a flowline Shallow-Ice model, called Winnie, developed by C. Ritz at LGGE. On twin experiments we got interesting results. Figures 4 show the reconstruction of the bedrock topography for various ensemble sizes. We can see that the obtained bedrock is very close to the true one, even for small ensemble sizes. This is very promising for the future, as we want to implement this method into a full 3D model. A journal paper has been submitted on this subject, and the results have been presented at many conferences [27], [37], [48], [38], [39].

6.2.3. *Inverse methods for full-Stokes glaciology models*

Participants: Olivier Gagliardini, Maëlle Nodet, Catherine Ritz.

We are investigating the means to apply inverse modeling to another class of glaciology models, called full-Stokes model. Such a model is developed by LGGE and CSC in Finland, called Elmer/Ice. Contrary to large scale models, Elmer/Ice is based on the full Stokes equations, and no assumptions regarding aspect ratio are made, so that this model is well adapted to high resolution small scale modelling, such as glaciers (and more recently the whole Greenland ice-sheet).

../../../../projets/moise/IMG/seabass.jpg

Figure 3. Snapshot of the relative vorticity field (1/s) for an academic oceanic basin model at 1/100° horizontal resolution.

../../../../projets/moise/IMG/ice_sheet/ETKF_N_30_50_100.png ../../../projets/moise/IMG/deviation_bedrock_

Figure 4.

Left: Bedrock topography after 20 years of the LETKF with inflation. The background (green) is compared to reference (blue) and the analyses for various ensemble sizes: 100 members (purple), 50 members (cyan) and 30 members (orange).
Right: Standard deviation of the errors (compared to the reference) for the bedrock topography after 20 years of the LETKF with inflation. The background (green) is compared to the analyses for various ensemble sizes: 100 members (purple), 50 members (cyan) and 30 members (orange).

In collaboration with O. Gagliardini, F. Gillet-Chaulet and C. Ritz (Laboratoire de Glaciologie et Géophysique de l'Environnement (LGGE), Grenoble), we investigated a new method to solve inverse problems for a Full-Stokes model of Groenland, which consisted in solving iteratively a sequence of Neumann and Dirichlet problems within a gradient descent algorithm. We also compared this method to an approximate variational algorithm, using the fact that the full Stokes equations are almost self-adjoint. These results have been presented at an international conference [44].

6.3. Development of New Methods for Data Assimilation

6.3.1. *Variational Data Assimilation with Control of Model Error*

Participants: Bénédicte Lemieux-Dudon, Arthur Vidard.

One of the main limitations of the current operational variational data assimilation techniques is that they assume the model to be perfect, mainly because of computing cost issues. Numerous researches have been carried out to reduce the cost of controlling model errors by controlling the correction term only in certain privileged directions or by controlling only the systematic and time correlated part of the error.

Both the above methods consider the model errors as a forcing term in the model equations. Trémolet (2006) describes another approach where the full state vector (4D field: 3D spatial + time) is controlled. Because of computing cost one cannot obviously control the model state at each time step. Therefore, the assimilation window is split into sub-windows, and only the initial conditions of each sub-window are controlled, the junctions between each sub-window being penalized. One interesting property is that, in this case, the computation of the gradients, for the different sub-windows, are independent and therefore can be done in parallel.

This method is now implemented in a realistic oceanic framework using OPAVAR/ NEMOVAR. An extensive documentation has been produced and we are now assessing the improvement over the previous scheme

6.3.2. *Direct assimilation of sequences of images*

Participants: François-Xavier Le dimet, Maëlle Nodet, Nicolas Papadakis, Arthur Vidard, Vincent Chabot.

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 just 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 [101]. 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 [77]. 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 submitted paper late 2013 [63] and presented in a national conference [34]

We are also part of TOMMI, another ANR project started mid 2011, where we are investigating the possibility to use optimal transportation based distances for images assimilation.

6.3.3. *Image processing, Optimal transport*

Participants: Maëlle Nodet, Nicolas Papadakis, Arthur Vidard, Nelson Feyeux.

Within the optimal transport project TOMMI funded by the ANR white program, a new optimization scheme based on proximal splitting method has been proposed to solve the dynamic optimal transport problem. This work allows the computation of generalized optimal transports and will be published in SIAM Journal on Imaging Sciences [96]. We investigate also the use of optimal transport based distances for data assimilation. N. Feyeux just started his PhD on this subject, and his PhD project has been presented in a regional workshop [49].

6.3.4. *A Nudging-Based Data Assimilation Method: the Back and Forth Nudging*

Participants: Maëlle Nodet, Jacques Blum, Didier Auroux.

The Back and Forth Nudging (BFN) algorithm has been recently introduced for simplicity reasons, as it does not require any linearization, nor adjoint equation, or minimization process in comparison with variational schemes. Nevertheless it provides a new estimation of the initial condition at each iteration.

Previous theoretical results showed that BFN was often ill-posed for viscous partial differential equations. To overcome this problem, we proposed a new version of the algorithm, which we called the Diffusive BFN, and which showed very promising results on one-dimensional viscous equations. Experiments on more sophisticated geophysical models, such as Shallow-Water equations and NEMO ocean model are still in progress, in collaboration with University of Nice, and have been presented at the MAMERN conference [30].

6.3.5. *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 presented in a paper submitted to QJRMS ([68]).

6.3.6. *Variational Data Assimilation and Control of Boundary Conditions*

Participant: Eugène Kazantsev.

A variational data assimilation technique is applied to the identification of the optimal boundary conditions for two configurations of the NEMO model.

The first one is a full-physics low-resolution configuration, known as ORCA-2 model. In this experiment we identify optimal parametrizations of boundary conditions on the lateral boundaries as well as on the bottom and on the surface of the ocean [17]. The influence of boundary conditions on the solution is analyzed as in the assimilation window and beyond the window. It is shown that the influence of the lateral boundaries is not significant in this configuration, while optimal surface and bottom boundary conditions allow us to better represent the jet streams, such as Gulf Stream and Kuroshio. Analyzing the reasons of the jets reinforcement, we notice that data assimilation has a major impact on parametrization of the bottom boundary conditions for u and v [23].

The second configuration of the Nemo model is devoted to the identification of the optimal parametrization of lateral boundary conditions. The model in a rectangular box placed in mid-latitudes and subjected 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 that optimal boundary has a rather complicated geometry which is neither a staircase, nor a strait line. The boundary conditions found in the data assimilation procedure brings the solution toward the reference solution allowing to correct the influence of the rotated grid (see fig. 5).

Adjoint models, necessary to variational data assimilation, have been produced by the TAPENADE software, developed by the TROPICS 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.
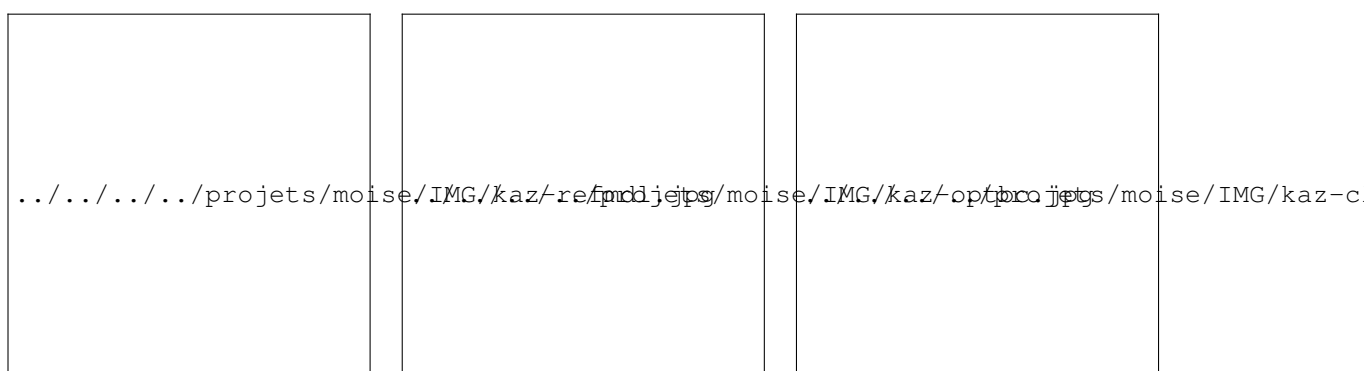

Figure 5. Sea surface elevation: reference solution on the aligned grid (left), solutions on the $30°$ rotated grid with optimal (center) and classical (right) boundary conditions.

6.4. Quantifying Uncertainty

6.4.1. *Sensitivity analysis for forecasting ocean models*

Participants: Anestis Antoniadis, Eric Blayo, Gaëlle Chastaing, Céline Helbert, Alexandre Janon, François-Xavier Le Dimet, Simon Nanty, Maëlle Nodet, Clémentine Prieur, Federico Zertuche, Simon Nanty, Laurent Gilquin.

6.4.1.1. Scientific context

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

6.4.1.2. 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' [100] 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 [98] and more recently by Owen [95] but the quantities they estimate still require $O(d)$ samples. In a recent work [104] we introduce a new approach to estimate for any k all the k-th order Sobol' indices by using only two samples based on replicated latin hypercubes. 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. It is the first step in the PhD of Laurent Gilquin (started in October 2013). Laurent Gilquin is supervised by Clémentine Prieur and Elise Arnaud (EPI STEEP) and his PhD is funded by the ANR project CITIES.

We can now wonder what are the asymptotic properties of these new estimators, or also of more classical ones. In [67], the authors deal with asymptotic properties of the estimators. In [70], the authors establish also a multivariate central limit theorem and non asymptotic properties.

6.4.1.3. Intrusive sensitivity analysis, reduced models

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 [10], 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 [90]. We are now considering problems related to more general models such as Shallow-Water models. In [15], 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.

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 submitted, dealing with goal oriented uncertainties assessment [89].

6.4.1.4. 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 [88] in the case where input parameters are correlated. Clémentine

Prieur supervised Gaëlle Chastaing's PhD thesis on the topic (defended in September 2013) [2]. We obtained first results [81], 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 submitted [64], [66].

Céline Helbert and Clémentine Prieur supervise the PhD thesis of Simon Nanty (funded by CEA Cadarache). 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 implies functional dependent inputs. A first step is the modeling of these inputs.

6.4.1.5. 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 ([91], [97]) : a new estimating 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 year 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 year of the thesis was dedicated to the development of a new sequential approach based on a course to fine wavelets algorithm.

6.4.2. Evaluation of a posteriori covariance errors

In the context of data assimilation, taking into account the a priori covariance error on the prediction and on the observations, the model and the observations, an analysis can be obtained followed by a prediction. This one makes sense only if an estimation of the error can be provided. The tendency is to use "ensemble methods" i.e. to realize a large number of predictions and estimate statistics on the results. This method raises two problems: the high computational cost and the weak theoretical justification. We have proposed a new method based on the fact that in the linear case the covariance is the inverse of the Hessian. The principle of our method is to add a correcting term to the Hessian in the non linear case. This work has been published in 2013 [14]. This paper has also be presented at the 6th WMO Symposium on Data Assimilation held in College Park, MD, USA in October 2013 [73].

6.4.3. Second Order Information in Variational Data Assimilation

This theme is centered around sensitivity analysis with respect to the observations. The link between data and models is made only in the Optimality System. Therefore a sensitivity analysis on the observations must be carried out on the Optimality System thus using second order information. This research is done in cooperation with Victor Shutyaev (Institute of Numerical Mathematics, Moscow), Tran Thu Ha (Institute of Mechanics, Ha Noi, Vietnam). One paper is published in the Russ. J. Of Numerical Analysis [18]. Another application to identification of parameters in a hydrological model is submitted [105].

6.5. Tracking of mesoscale convective systems

Participants: Clémentine Prieur, Alexandros Makris.

6.5.1. Scientific context

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).

6.5.2. Results

A first approach involves adapting the multiple hypothesis tracking (MHT) model originally designed by the NCAR (National Centre for Atmospheric Research) for tracking storms [102] to the data for West Africa. With A. Makris (working on a post-doctoral position), we proposed a Bayesian approach [76], 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 and must now be investigated in the specific context of real data on West Africa. Using PHD (probability hypothesis density) filters adapted to our problem, generalising recent developments in particle filtering for spatio-temporal branching processes (e.g. [80]) 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.

6.6. Multivariate risk indicators

Participants: Clémentine Prieur, Patricia Tencaliec.

6.6.1. Scientific context

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.

6.6.2. Results

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 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 [82] for a review of recent extensions of the notion of return level to the multivariate framework. In the context of environmental risk, [99] 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 [83] 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 [9], [86], 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 [10]. 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.

6.7. Non-parametric estimation for kinetic diffusions

Participant: Clémentine Prieur.

This research is the subject of a collaboration with Venezuela (Professor Jose R. Leon, Caracas Central University) and is partly funded by an ECOS Nord project.

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 coercivitiy, 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 [8] and when only partial observational data are available. A paper on the non parametric estimation of the drift has been submitted recently [62] (see Samson et al., 2012, for results for parametric models). As far as the estimation of the diffusion term is concerned, we obtained promising results, in collaboration with J.R. León (Caracas, Venezuela) and P. Cattiaux (Toulouse). These results should be submitted shortly.

6.8. CO$_2$ Storage

Participant: Céline Helbert.

In collaboration with Bernard Guy (EMSE, Saint-Etienne) and more specifically in the context the PhD of Joharivola Raveloson (EMSE, Saint-Etienne), we are interested in the study of the water-rock interactions in the case of CO$_2$ storage in geological environment. This work is following the study of Franck Diedro in the same subject [87]. In this study we focus on the scale of observation of geochemical phenomena while taking into account the heterogeneity of the reservoir. This heterogeneity at small and large scale helps to maintain a local variability of the chemical composition and influence reaction rates at the pore as well at the reservoir scale. To connect the parameters at both scale (pore and reservoir) we use deterministic and stochastic simulations of a reactive transport code developed by IFPEN.

6.9. Land Use and Transport models calibration

Participants: Thomas Capelle, Laurent Gilquin, Clémentine Prieur, Nicolas Papadakis, Arthur Vidard.

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 lead by the STEEP EPI. This project has just started early 2013 and Two PhD about sensitivity analysis and calibration were launched this fall.

6.10. Mathematical modelling for CFD-environment coupled systems

Participants: Antoine Rousseau, Maëlle Nodet.

6.10.1. *Minimal-time bioremediation of natural water resources*

The objective of this work is to provide efficient strategies for the bioremediation of natural water resources. The originality of the approach is to couple minimal time strategies that are determined on a simplified model with a faithful numerical model for the hydrodynamics. Based on a previous paper that deals with an implicit representation of the spatial inhomogenity of the resource with a small number of homogeneous compartments (with a system of ODEs), we implement a coupled ODE-PDE system that accounts for the spatial non-homogeneity of pollution in natural resources. The main idea is to implement a Navier-Stokes model in the resource (such as a lake), with boundary conditions that correspond to the output feedback that

has been determined to be optimal for the simple ODEs model of a (small) bioreactor. A first mathematical model has been introduced and numerical simulations have been performed in academic situations. We built a reduced model that approximates the reference PDE model thanks to a set of ODEs with parameters. Numerical optimization is performed on these parameters in order to better fit the reference model. In addition, bioremediation algorithms proposed by the authors have been sent to Inria Technology Transfer Services for a patent registration. Two publications (ready for submission) will be sent as soon as the patent submission process is complete.

Finally, A. Rousseau spent 2 weeks in Santiago (April 2013) upon Inria Chile's invitation in order to work on the bioremediation of natural resources. AR and Inria Chile made a common answer to a chilean funding program (by COPEC) that was not chosen.

6.10.2. Mathematical modelling for the confinement of lagoons

This work deals with the concept of confinement of paralic ecosystems. It is based on a recent paper by E. Frénod that presents a modelling procedure in order to compute the confinement field of a lagoon.

A. Rousseau and E. Frénod improve in 2012 the existing model in order to account for tide oscillations in any kind of geometry such as a non-rectangular lagoons with a non-flat bottom. The new model, that relies on PDEs rather than ODEs, is then implemented thanks to the finite element method. Numerical results confirm the feasibility of confinement studies thanks to the introduced model. During the internship of J.-P. Bernard, we implemented the proposed method in a realistic situation, namely the Etang de Thau in Languedoc-Roussillon, France (see Figure 6). This led to two publications in 2013 [5] and [13], plus one accepted paper in 2014 [6].


```
../../../../projets/moise/IMG/Thau.png
```

Figure 6. Confinement map in the Thau Lagoon (France).

MORPHEO Team

6. New Results

6.1. Robust human body shape and pose tracking

This work considers markerless human performance capture from multiple camera videos and, in particular, the recovery of both shape and parametric motion information, as often required in applications that produce and manipulate animated 3D contents using multiple videos. To this aim, an approach is proposed that jointly estimates skeleton joint positions and surface deformations by fitting a reference surface model to 3D point reconstructions. The approach is based on a probabilistic deformable surface registration framework coupled with a bone binding energy. The former makes soft assignments between the model and the observations while the latter guides the skeleton fitting. The main benefit of this strategy lies in its ability to handle outliers and erroneous observations frequently present in multi view data. For the same purpose, we also introduce a learning based method that partitions the point cloud observations into different rigid body parts that further discriminate input data into classes in addition to reducing the complexity of the association between the model and the observations. We argue that such combination of a learning based matching and of a probabilistic fitting framework efficiently handle unreliable observations with fake geometries or missing data and hence, it reduces the need for tedious manual interventions. The work was presented at the 3DV conference [7] where it received the best paper runner up award.

Figure 4. Human pose recovery with 3 different standard datasets.

6.2. Inverse dynamics on rock climbing with and without measurement of contact forces

Rock climbing involves complex interactions of the body with the environment (Figure 5). It represents an interesting problem in biomechanics as multiple contacts in the locomotion task make it an underconstrained problem. In this study we are interested in evaluating how a climber transfers weight through the holds. The motivation of this study is also technical as we are developing an inverse dynamics method that automatically estimates in 3D, not only the usual torques at joint angles, but also the wrenches at contacts [9].

../../../../projets/morpheo/IMG/climbInvDyn.png

Figure 5. Inverse dynamics on rock climbing with and without measurement of contact forces.

6.3. Video-based methodology for markerless human motion analysis

This study presents a video-based experiment for the study of markerless human motion. Silhouettes are extracted from a multi-camera video system to reconstruct a 3D mesh for each frame using a reconstruction method based on visual hull. For comparison with traditional motion analysis results, we set up an experiment integrating video recordings from 8 video cameras and a Vicon™ marker-based motion capture system (Figure 6). Our preliminary data provided distances between the 3D trajectories from the Vicon system and the 3D mesh extracted from the video cameras. In the long term, the main ambition of this method is to provide measurement of skeleton motion for human motion analyses while eliminating markers [8].
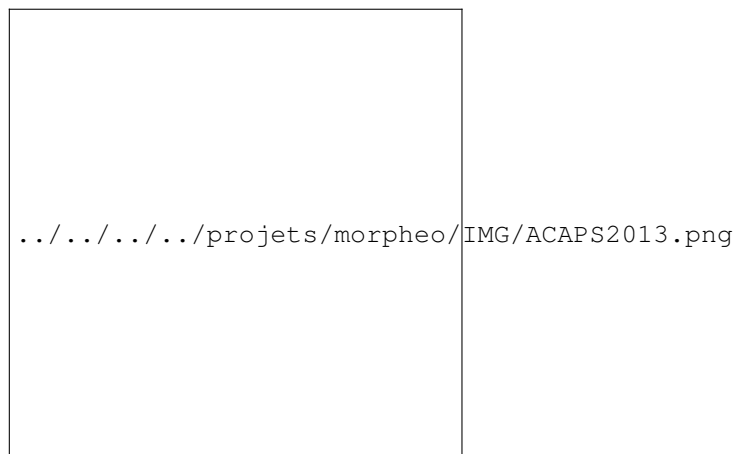
../../../../projets/morpheo/IMG/ACAPS2013.png

Figure 6. Video-based methodology for markerless human motion analysis.

6.4. 3D shape cropping

We introduce shape cropping as the segmentation of a bounding geometry of an object as observed by sensors with different modalities. Segmenting a bounding volume is a preliminary step in many multi-view vision applications that consider or require the recovery of 3D information, in particular in multi-camera environments. Recent vision systems used to acquire such information often combine sensors of different types, usually color and depth sensors. Given depth and color images we present an efficient geometric algorithm to compute a polyhedral bounding surface that delimits the region in space where the object lies. The resulting cropped geometry eliminates unwanted space regions and enables the initialization of further processes including surface refinements. Our approach exploits the fact that such a region can be defined as the intersection of 3D regions identified as non empty in color or depth images. To this purpose, we propose a novel polyhedron combination algorithm that overcomes computational and robustness issues exhibited by traditional intersection tools in our context. We show the correction and effectiveness of the approach on various combination of inputs. This work was presented at the Vision Modeling and Visualization workshop 2013 [6].

6.5. Multi-view object segmentation in space and time

In this work, we address the problem of object segmentation in multiple views or videos when two or more viewpoints of the same scene are available. We propose a new approach that propagates segmentation coherence information in both space and time, hence allowing evidences in one image to be shared over the complete set. To this aim the segmentation is cast as a single efficient labeling problem over space and time with graph cuts. In contrast to most existing multi-view segmentation methods that rely on some form of dense reconstruction, ours only requires a sparse 3D sampling to propagate information between viewpoints. The approach is thoroughly evaluated on standard multi-view datasets, as well as on videos. With static views, results compete with state of the art methods but they are achieved with significantly fewer viewpoints. With multiple videos, we report results that demonstrate the benefit of segmentation propagation through temporal cues, in ICCV 2013 [5].

6.6. Segmentation of temporal mesh sequences into rigidly moving components

This work considers the segmentation of meshes into rigid components given temporal sequences of deforming meshes (Figure 9). We have proposed a fully automatic approach that identifies model parts that consistently move rigidly over time. This approach can handle meshes independently reconstructed at each time instant. It allows therefore for sequences of meshes with varying connectivities as well as varying topology. It incrementally adapts, merges and splits segments along a sequence based on the coherence of motion information within each segment. In order to provide tools for the evaluation of the approach, we also introduce new criteria to quantify a mesh segmentation. Results on both synthetic and real data as well as comparisons are provided in the paper [1].

6.7. Segmentation of plant point cloud models into elementary units

High-resolution terrestrial Light Detection And Ranging (tLiDAR), a 3-D remote sensing technique, has recently been applied for measuring the 3-D characteristics of vegetation from grass to forest plant species. The resulting data are known as a point cloud which shows the 3-D position of all the hits by the laser beam giving a raw sketch of the spatial distribution of plant elements in 3-D, but without explicit information on their geometry and connectivity. In this study we propose a new approach based on a delineation algorithm that clusters a point cloud into elementary plant units. The algorithm creates a graph (points + edges) to recover plausible neighbouring relationships between the points and embed this graph in a spectral space in order to segment the point-cloud into meaningful elementary plant units. Our approach is robust to inherent geometric outliers and/or noisy points and only considers the x, y, z coordinate tLiDAR data as an input. It has been presented at the FSPM conference [4].

../../../../projets/morpheo/IMG/teaser.png

Figure 7. Result of shape cropping using three input depth maps for polyhedral reconstruction.

../../../../projets/morpheo/IMG/teaser2.png

Figure 8. Multi-view object segmentation using our method with the 3 wide-baseline views shown only, with no photo-consistency hypothesis and no user interaction.

../../../../projets/morpheo/IMG/these_romain.png

Figure 9. Segmentation of temporal mesh sequences into rigidly moving components.

../../../../projets/morpheo/IMG/fspm.jpg

Figure 10. Segmentation of a plant point cloud model into elementary units.

NANO-D Team

5. New Results

5.1. Adaptively Restrained Particle Simulations for Isobaric-Isothermal Ensemble

Participants: Zofia Trstanova, Stephane Redon.

We continued working on the Adaptively Restrained Particles Simulations (ARPS) approach that was proposed by Svetlana Artemova and Stephane Redon [11] and that was designed to speed up the particles simulations by switching on and off the degrees of freedom based on the kinetic energy of the particle.

It has been shown, for the NVE and the NVT ensemble, that this method has many advantages [11]. We want to extend ARPS for the isobaric-isothermal ensemble (NPT) since this ensemble is very often used in particle simulations, because many chemical reactions happen under constant pressure. An adaptive method for this ensemble with advantages of ARPS might be very useful in many scientific domains (physics, biology, chemistry).

We combined the ARPS method with an existing method that describes the NPT ensemble. We already obtained very promising analytical and numerical results that support the main characteristically advantages of ARPS shown by Svetlana Artemova and Stephane Redon. For instance, Figure 4 shows preservation of the radial distribution function.

5.2. Interactive large-scale deformations of molecular structures

Participants: Jelmer Wolterink, Himani Singhal, Marc Piuzzi, Stephane Redon.

We have developed new interaction methods for large-scale deformation of molecular structures. These new methods allow a user to attach control points to molecules, and use these control points to easily deform the structures while preserving their realism (e.g. local interactions, etc.). The new methods may be applied to any type of molecule (e.g. proteins, carbon nanotubes, etc.), and may be used in combination with interactive simulation.

5.3. Towards parallel adaptive molecular simulations

Participants: Krishna Kant Singh, Benjamin Bouvier, Jean-Francois Mehaut, Stephane Redon.

The adaptive algorithms that we are developing have two main components. The first component determines when and how degrees of freedom can be deactivated and reactivated during a simulation. The second component takes advantage of the frozen degrees of freedom to accelerate the calculation of the potential energy and interatomic forces. Indeed, the potential energy and forces can often be expressed as a (potentially complex) sum of terms which only depend on relative atomic positions. When the relative positions do not change, it is not necessary to update the corresponding terms, which reduces the computation time. We have shown that it is possible to significantly speed up simulations using this approach, while being able to recover static equilibrium statistics [11].

We have now begun to study the possibility of developing adaptive *parallel* simulation algorithms, and have begun to review and benchmark popular simulation packages (GROMACS, NAMD, OpenMM, etc.), depending on the number of atoms, the number of available cores, etc.

5.4. Protein secondary structure prediction for dynamic simulations

Participants: Marc Piuzzi, Sergei Grudinin, Stephane Redon.

../../../../projets/nano-d/IMG/RDF.png

Figure 4. Radial distribution function obtained with different ARPS simulations compared to the full dynamics
(blue dash line)

There is a tight link between a protein's function and its molecular structure. Hence, global stability is essential for a protein to keep its role inside the cell. Various chemical interactions helps stabilizing the structure (covalent bonds, hydrogen bonds, etc.) but not all parts of a protein present the same stability. The most stable regions of a protein present numerous hydrogen bonds on backbone atoms composing geometrically distinguishable secondary structures (the primary structure being the amino acid sequence): helices and beta sheets.

These structures have been well studied and although important properties have been defined, there is no absolute definition of what is a helix or a beta sheet. Thus, various methods have been developed to predict the secondary structure of a protein using the amino acid sequence and/or the protein structure using different parameters and structural descriptors.

However, none of these methods have been made in the context of interactive simulation where the shape of the protein is dynamic: here the prediction has to be done at each time step on the whole protein. Moreover, the result is deterministic and returns only the type of structure without any information about the accuracy. We are developing a new approach that is appropriate in an interactive context, where secondary structure assignment has to continuously change during interaction.

5.5. Motion Planning for Quasi-Static Simulation

Participants: Leonard Jaillet, Stephane Redon.

Recently, motion planning methods inspired from Robotics have been applied to the study of biological molecular systems [8]. These approaches rely on compact graph representations that aim to capture large amplitude motions more efficiently than classic simulation techniques, despite their lower resolution.

We developed within the SAMSON's architecture a new motion planning strategy to perform quasi-static simulation at the nano-scale.

The user provides as inputs the initial and final state of the system he or she wants to simulate. Then, the method searches a transition path that follows the low-energy valleys of the conformational landscape (see figure 5).

The adaptation of motion planning approaches to quasi-static simulation at the nano-scale comes with several challenges. First, these approaches must be adapted to tackle the high dimensionality involved in the case of nanosystems, dimensionality that is directly related to the number of atoms considered. Second, these approaches must be extended to face the complexity of the underlying physics that comes from the various types of interactions between atoms.

The method we propose is able to perform simulations involving bonds breaking. This is, up to our knowledge, the first motion planning approach able to simulate chemical reactions.

5.6. Molecular Modeling

5.6.1. *Rapid determination of RMSDs corresponding to macromolecular rigid body motions*

Participants: Petr Popov, Sergei Grudinin.

Finding the root mean sum of squared deviations (RMSDs) between two coordinate vectors that correspond to the rigid body motion of a macromolecule is an important problem in structural bioinformatics, computational chemistry and molecular modeling. Standard algorithms compute the RMSD with time proportional to the number of atoms in the molecule. We developed *RigidRMSD*, a new algorithm that determines a set of RMSDs corresponding to a set of rigid body motions of a macromolecule in constant time with respect to the number of atoms in a molecule. Our algorithm is particularly useful for rigid body modeling applications such as rigid body docking, and also for high-throughput analysis of rigid body modeling and simulation results. A C++ implementation of our algorithm will be available at http://nano-d.inrialpes.fr/software/RigidRMSD.

../../../../projets/nano-d/IMG/Methane.png

Figure 5. Snapshots of the transition path obtained with our motion planning simulation method. It represents a chemical reaction where two molecules of methanes interact to form a an ethane and a dihydrogen.

To demonstrate the efficiency of the RigidRMSD library, we compared the clustering application implemented with our algorithm to the one from the Hex software. We chose Hex for the comparison because it is a very fast rigid body docking tool and also because it explicitly provides the clustering time. For the comparison, we collected a small benchmark of 23 protein dimers of various size. After, we launched Hex version 6.3 on this benchmark and collected docking solutions before clustering, sizes of clusters, and clustering time. We then also clustered these solutions using the *RigidRMSD* library. Figure 6 shows the clustering time of the HEX clustering algorithm with respect to our clustering using two rotation representations as a function of the number of atoms in the smaller protein (left) and the number of docking solutions before the clustering (right). We can clearly see that our implementation of the clustering algorithm is more than an order of magnitude faster compared to the Hex implementation. Also, the quaternion representation of rotation is on average three times more efficient compared to the matrix representation.

../../../../projets/nano-d/IMG/Graph-01-log.png ../../../projets/nano-d/IMG/Graph-02-log.png

Figure 6. Left: Time spent on clustering by Hex and RigidRMSD with respect to the number of atoms in the ligand protein. Number of considered solutions and the RMSD threshold was fixed to 10,000 and 10.0 Å, respectively. Right: Average time spent on clustering by Hex and RigidRMSD with respect to the number of docking solutions. For this plot we chose five structures with the number of atoms of about 2,000 and plotted the standard deviation of the running time. For both plots, the RMSD threshold was fixed to 10.0 Å.

5.6.2. *Fast fitting atomic structures into a low-resolution density map using 3D orthogonal Hermite functions*

Participants: Georgy Derevyanko, Sergei Grudinin.

We developed a new algorithm for fitting protein structures into a low resolution electron density (e.g. cryo-electron microscopy) map. The algorithm uses 3D orthogonal Hermite functions for fast operations on the electron density.

Orthogonal Hermite function of order n is defined as:

$$\psi_n(x;\lambda) = \frac{\sqrt{\lambda}}{\sqrt{2^n n! \sqrt{\pi}}} \exp\left(-\frac{\lambda^2 x^2}{2}\right) H_n(\lambda x),$$ (1)

where $H_n(x)$ is the Hermite polynomial and λ is the scaling parameter. In Fig. 7 we show several orthogonal Hermite functions of different orders with different parameters λ. These functions form an orthonormal basis set in $L^2(\mathbb{R})$. A 1D function $f(x)$ decomposed into the set of 1D Hermite functions up to an order N reads

$$f(x) = \sum_{i=0}^{N} \widehat{f}_i \psi_i(x; \lambda) \tag{2}$$

Here, $\widehat{f}_i$ are the decomposition coefficients, which can be determined from the orthogonality of the basis functions $\psi_i(x; \lambda)$. Decomposition in Eq. 2 is called the *band-limited decomposition* with $\psi_i(x; \lambda)$ basis functions. To decompose the electron density map and the protein structures, we employ the 3D Hermite functions:

$$\psi_{n,l,m}(x, y, z; \lambda) = \psi_n(x; \lambda)\psi_l(y; \lambda)\psi_m(z; \lambda), \tag{3}$$

which form an orthonormal basis set in $L^2(\mathbb{R}^3)$. A function $f(x, y, z)$ represented as a band-limited expansion in this basis reads

$$f(x, y, z) = \sum_{i=0}^{N} \sum_{j=0}^{N-i} \sum_{k=0}^{N-i-j} \widehat{f}_{i,j,k} \psi_{i,j,k}(x, y, z; \lambda) \tag{4}$$

Our algorithm accelerates rotation of the Fourier image of the electron density by using 3D orthogonal Hermite functions. As a part of the new method, we presented an algorithm for the rotation of the density in the Hermite basis and an algorithm for the conversion of the expansion coefficients into the Fourier basis. We implemented the program of fitting a protein structure to a low-resolution electron density map, which uses the cross-correlation or the Laplacian-filtered cross-correlation as the fitting criterion. We demonstrated that in the Hermite basis, the Laplacian filter has a particularly simple form. To assess the quality of density encoding in the Hermite basis, we uses two measures, the R-factor and the cross-correlation factor. Finally, we validated our algorithm using two examples and compare its efficiency with two widely used fitting methods, ADP_EM and *colores* from the Situs package.

5.6.3. Fast Rotational-Translation Matching of Rigid Bodies by Fast Fourier Transform Acceleration of Six Degrees of Freedom

Participants: Alexandre Hoffmann, Sergei Grudinin.

We introduced a new method for rigid molecular fitting. This problem is usually solved by maximizing the Cross Corelation Function (CCF), which is computed using the Fast Fourier Transform (FFT) algorithm. Our method handles six degrees of freedom at once and requires only one computation of the Cross Corelation Function, with the six-dimensional Fast Fourier Transform. Our method only requires a low pre-processing time ($O(N^7)$), which is comparable to the cost of the subsequent 6D FFT ($O(N^6 \log(N^6))$). It also uses a dual Hermite-Fourier representation, which allows to represent a small molecule with a fewer number of coefficients in the Hermite basis.

5.6.4. Prediction of complexes with point group symmetry using spherical polar Fourier docking correlations

Participants: David W. Ritchie, Sergei Grudinin.

Many proteins form symmetric homo-oligomers that perform a certain physiological function. We present the first point group symmetry docking algorithm that generates perfectly symmetrical protein complexes for arbitrary point group symmetry types (C_n, D_n, T, O, and I). We validate the algorithm on proteins from the 3D-Complex database, where it achieves on average the success rate of 55%. The running time of the algorithm is less than a minute on a modern workstation.

../../../../projets/nano-d/IMG/hermitePlotNLambda_01.png

Figure 7. Left: 1D Hermite functions of order 6 for three different scaling parameters λ. Right: 1D Hermite functions of two different orders for the scaling parameter $\lambda = 1$.

Many of the protein complexes in the protein Data bank (PDB) are symmetric homo-oligomers. According to the 3D-Complex database, C_2 homo-dimers comprise the majority of known homo-oligomers. However, many complexes have higher order rotational symmetry (i.e. C_n>2), and a good number have multiple rotational symmetry axes, namely those with dihedral (D_n), tetrahedral (T), octahedral (O), and icosahedral (I) point group symmetries. Although symmetrical complexes are often solved directly by X-ray crystallography, it would still be very useful to be able to predict whether or not a given monomer might self-assemble into a symmetrical structure. We present a new point group symmetry docking algorithm. In the last few years, several protein-protein docking programs have been adapted to predict symmetrical pair-wise docking orientations for C_n and D_n symmetries. However, to our knowledge, there does not yet exist an algorithm which can automatically generate perfectly symmetrical protein complexes for arbitrary point group symmetry types.

We introduce the notion of a "docking equation" in which the notation $A(\underline{x}) \longleftrightarrow B(\underline{x})$ represents an interaction between proteins A and B in 3D space. It is also useful introduce the operators $\widehat{T}(x, y, z)$ and $\widehat{R}(\alpha, \beta, \gamma)$, which represent the actions of translating an object by an amount (x, y, z) and rotating it according to the three Euler rotation angles (α, β, γ). Then, guided by Figure 8 , and assuming that we start with two identical monomers at the origin, we can write down a C_n docking equation for the two monomers as

$$\widehat{T}(0, y, 0)\widehat{R}(\alpha, \beta, \gamma)A(\underline{x}) \longleftrightarrow \widehat{R}(0, 0, \omega)\widehat{T}(0, y, 0)\widehat{R}(\alpha, \beta, \gamma)B(\underline{x}). \tag{5}$$

Then, we perform a series of fast Fourier transform (FFT) correlation searches using the Hex spherical polar Fourier docking algorithm to determine the four parameters $(y, \alpha, \beta, \gamma)$. For higher symmetries, D_n, T, O, and I, we introduce two more parameters and perform a series of FFT in a similar way. The calculation for each structure takes less than a minute on a modern workstation.
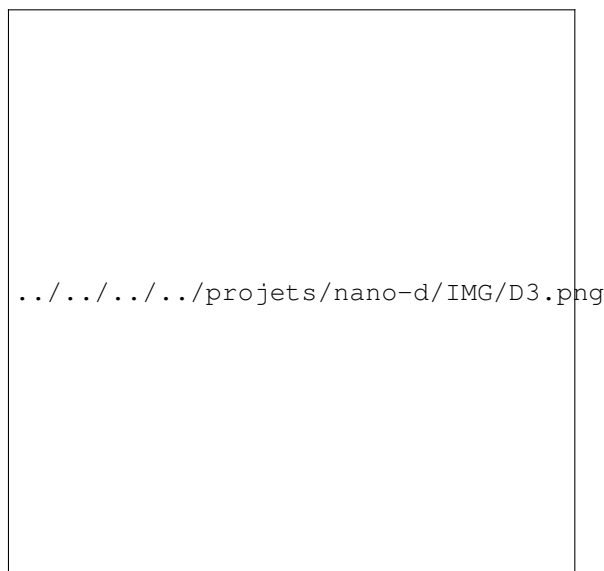

Figure 8. Illustrations of the C_3 and D_3 point group symmetries.

We validated our method on protein structures from the 3D-Complex database, which contains 17,183 protein complexes with assigned biological unit and symmetry type. It mostly contains cyclic and dihedral proteins, as well as 86 tetrahedral, 47 octahedral, and 6 icosahedral complexes (excluding all viral structures). Starting

from the structures of monomers, we generated symmetric biological units based on the symmetry type for each complex provided by 3D-Complex. Figure 9 summarizes the performance of our method on these proteins, where we say that the model is correct if all pair-wise RMSDs are smaller than 10 Ångstroms. On average, we found about 55% of correct predictions ranked first.

../../../../projets/nano-d/IMG/3DSIG-plot2.png

Figure 9. Summary of the correctly predicted complexes found on the first place (blue) and in the top ten solutions (green).

Figure 10 shows correctly predicted examples from each of the symmetry types. Each complex is perfectly symmetrical, although due to the soft docking function in Hex it is possible that some interfaces might contain minor steric clashes.

5.7. Software Engineering

5.7.1. *SAMSON User interface*

Participants: Jocelyn Gate, Maria Werewka, Stephane Redon.

We have continued the development of SAMSON, our open-architecture platform for modeling and simulation of nanosystems (SAMSON: Software for Adaptive Modeling and Simulation Of Nanosystems):

- We have moved to Qt5 to handle the Graphical User Interface.

../../../../projets/nano-d/IMG/3DSIG-plot3.png

Figure 10. Illustrations of the correctly predicted complexes. For each complex, the group symbol and the PDB code are shown.

- We now compile SAMSON is 64 bits only. This removes limitations of 32 bits applications, in particular concerning memory limits.
- We now have a complete installer mechanism for both users and developers.
- We extended the set of development tools (action generators, UUID generators, etc.)
- We have changed the windowing system to allow windows to move outside SAMSON.
- We have designed a coherent style for icons, windows, menus, etc.
- We have added 3D rendering.

There are now more than 40 modules in SAMSON (parsers, editors, models, apps, etc.).

The current user interface of SAMSON is visible in Figure 11 .

5.7.2. *SAMSON Elements*

Participants: Svetlana Artemova, Stephane Redon.

We have added new SAMSON Elements (modules).

We have been working on input and output for SAMSON. Precisely, we now have the possibility to download molecules to SAMSON and save them to external files in three possible formats:

- pdb (Protein Data Bank format, containing experimentally determined 3d structures and widely used for applications in biology);
- mol2 (Sybyl chemical modeler input file, containing chemical compounds and small ligands);
- xyz (basic format, containing atoms coordinates).

Basic properties of atoms, residues, and molecules have been determined and structures storing these properties were implemented.

Finally, since energy minimization is crucial for providing physically-correct structures while interactively editing molecules in SAMSON, we have implemented several fast and stable algorithms to perform such energy minimization in SAMSON.

5.7.3. *SAMSON Website*

Participants: Mohamed Yengui, Jocelyn Gate, Stephane Redon.

We are developing a web application aiming at distributing and valorizing SAMSON and SAMSON Elements (modules). The goal of the website is to develop a community of users and developers in all areas of nanoscience (physics, biology, chemistry, electronics, etc.). The website will:

- allow users and developers to create and manage accounts on the website.
- allow visualizing, searching and downloading SAMSON and SAMSON Elements.
- allow the creation, validation and dissemination of SAMSON Elements.
- provide tracking requests for the arrival of new SAMSON Elements or the modification of an existing SAMSON Element.

To achieve this, we have designed the architecture in a way that speeds development effort for a faster product release, while keeping in mind scalability, security and high reliability.

We have also implemented and tested locally the account validation process. A user can now sign up, confirm the registration from the received email and authenticate with the registered account to download SAMSON and SAMSON elements from the website. We will make the site public when we release SAMSON.

../../../../projets/nano-d/IMG/SAMSON-Current.jpg

Figure 11. The current user interface of SAMSON showing an app to download molecules directly from the Protein Data Bank. The data graph on the left shows the hierarchical structure of the structural model.

<p align="center" style="color:red">NECS Project-Team</p>

6. New Results

6.1. Networked systems and graph analysis

6.1.1. Distributed graph-discovery

Participants: A. Kibangou [Contact person], T.-M. D. Tran, F. Garin, A. de Almeida [UFC Brazil].

The availability of information on the communication topology of a wireless sensor network is essential for the design of the estimation algorithms. In the context of distributed self-organized sensor networks, there is no central unit with the knowledge of the network, and the agents must run some distributed network discovery algorithm.

We have studied the problem of estimating the eigenvalues of the Laplacian matrix associated with a graph modeling the interconnections between the nodes of a given network. Our approach is based on properties resulting from the factorization of the average consensus matrix. Indeed, as recently shown [45], the average consensus matrix can be written as a product of Laplacian based consensus matrices whose stepsizes are given by the inverse of the nonzero Laplacian eigenvalues. By distributively solving the factorization of the average consensus matrix, we have shown that the Laplacian eigenvalues can be computed as the inverse of the stepsizes in each estimated factor, where these factors are constrained to be structured as Laplacian based consensus matrices. A constrained optimization problem was formulated and distributed gradient descent methods have been formulated. As formulated, the problem can be viewed as a consensus problem with equality constraints. In contrast to the state-of-the-art, the proposed algorithm does not require great resources in both computation and storage. This algorithm can also be viewed as a way for decentralizing the design of finite-time average consensus protocol recently proposed in the literature.

Laplacian eigenvalues have several interesting properties that can help to study networks, however they cannot uniquely characterize the topology of the network. Therefore, we have directly studied the problem of topology identification in [20]. The considered set-up concerns a collaborative wireless sensor network where nodes locally exchange coded informative data before transmitting the combined data towards a remote fusion center equipped with an antenna array. For this communication scenario, a new blind estimation algorithm was developed for jointly recovering network transmitted data and connection topology at the fusion center. The proposed algorithm is based on a two-stage approach. The first stage is concerned with the estimation of the channel gains linking the nodes to the fusion center antennas. The second stage performs a joint estimation of network data and connection topology matrices by exploiting a constrained (PARALIND) tensor model for the collected data at the fusion center.

Distributed network-discovery algorithms become even more challenging in the case where the algorithm must be anonymous, namely in the case when the agents do not have or do not want to disclose their identifiers (id.s), either for technological reasons (in time-varying self-organized networks, assigning unique identifiers to agents is a challenge) or for privacy concerns. In anonymous networks, even simple tasks such as counting the number of agents are challenging problems. In [24] we have proposed an algorithm for node-counting in anoymous networks. It is based on a graph-constrained LTI system similar to linear consensus, and on system identification: the idea is that the order of the system is the number of agents, and based on local observations each agent tries to identify the order of the system, testing the rank of the Hankel matrix from the output data.

6.1.2. Observability in consensus networks

Participants: A. Kibangou [Contact person], C. Commault [Grenoble INP].

Studying the observability problem of a system consists in answering the question: is it possible, for a given node, to reconstruct the entire network state just from its own measurements and those of its neighbors? Studying observability for arbitrary graphs is particularly a tough task, therefore, studies are generally restricted to some families of graphs; for instance, recently, observability has been studied for paths and circular graphs and also grids where the study was carried out based on rules on number theory. We have considered families of graphs admitting an association scheme such that strongly regular graphs and distance regular graphs. The regularity properties of these kinds of graphs can particularly be useful for robustifying the network as for cryptographic systems. Based on the so-called Bose-Mesner algebra, we have stated observability conditions on consensus networks modeled with graphs modeled with strongly regular graphs and distance regular graphs; for this purpose, we have introduced the notion of local observability bipartite graph that allows characterizing the observability in consensus networks. We have shown that the observability condition is given by the nullity of the so-called "local bipartite observability graph"; when the nullity of the graph cannot be derived directly from the structure of the local bipartite observability graph, the rank of the associated bi-adjacency matrix allows evaluating the observability. The bi-adjacency matrix of the local bipartite observability graph must be full column rank for guaranteeing observability. From this general necessary and sufficient condition, we have deduced sufficient conditions for strongly regular graphs and distance regular graphs [25].

6.2. Collaborative and distributed algorithms

6.2.1. *Finite-time average consensus*

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

Nowadays, several distributed estimation algorithms are based on the average consensus concept. Average consensus can be reached using a linear iterations scheme where each node repeatedly updates its value as a weighted linear combination of its own value and those of its neighbors; the main benefit of using a linear iterations scheme is that, at each time-step, each node only has to transmit a single value to each of its neighbors. Based on such a scheme, several algorithms have been proposed in the literature; however, in the most proposed algorithms the weights are chosen so that all the nodes asymptotically converge to the same value. Sometimes, consensus can be embedded as a step of more sophisticated distributed; obviously, asymptotic convergence is not suitable for these kinds of distributed methods, and therefore it is interesting to address the question of exact consensus in finite-time. For time-invariant network topologies and in the perfect information exchange case, i.e., without channel noise nor quantization, we have shown that the finite-time average consensus problem can be solved as a matrix factorization problem with joint diagonalizable matrices depending on the graph Laplacian eigenvalues; moreover, the number of iterations is equal to the number of distinct nonzero eigenvalues of the graph Laplacian matrix. The design of such a protocol requires the knowledge of the Laplacian spectrum, which can be carried out in a distributed way (see Section 6.1.1 . The matrix factorization problem is solved in a distributed way, in particular a learning method was proposed and the optimization problem was solved by means of distributed gradient backpropagation algorithms. The factor matrices are not necessarily symmetric and the number of these factor matrices is exactly equal to the diameter of the graph [30].

6.2.2. *Linear consensus in large-scale geometric graphs*

Participants: F. Garin [Contact person], E. Lovisari [Lund], S. Zampieri [Padova].

Traditional analysis of linear average-consensus algorithms studies, for a given communication graph, the convergence rate, given by the essential spectral radius of the transition matrix (i.e., the second largest eigenvalues' modulus). For many graph families, such analysis predicts a performance which degrades when the number of agents grows, basically because spreading information across a larger graph requires a longer time; however, when considering other well-known quadratic performance indices (involving all the eigenvalues of the transition matrix), the scaling law with respect to the number of agents can be different. This is consistent with the fact that, in many applications, for example in estimation problems, it is natural to expect that a larger number of cooperating agents has a positive, not a negative effect on performance. It is

natural to use a different performance measure when the algorithm is used for different purposes, e.g., within a distributed estimation or control algorithm. We are interested in evaluating the effect of the topology of the communication graph on performance, in particular for large-scale graphs. We have focused on graph families which can describe sensor networks, and hence have geometric constraints, namely nodes can be connected only with nearby nodes in the sense of Euclidean distance [16].

6.2.3. *Distributed computation methods for multidimensional data*
Participants: A. Kibangou [Contact person], A. de Almeida [UFC Brazil].

In [19], we consider the issue of distributed computation of tensor decompositions. A central unit observing a global data tensor assigns different data sub-tensors to several computing nodes grouped into clusters. The goal is to distribute the computation of a tensor decomposition across the different computing nodes of the network, which is particularly useful when dealing with large-scale data tensors. However, this is only possible when the data sub-tensors assigned to each computing node in a cluster satisfies minimum conditions for uniqueness. By allowing collaboration between computing nodes in a cluster, we show that average consensus based estimation is useful to yield unique estimates of the factor matrices of each data sub-tensor. Moreover, an essentially unique reconstruction of the global factor matrices at the central unit is possible by allowing the subtensors assigned to different clusters to overlap in one or several modes. The proposed approach is useful to a number of distributed tensor-based estimation problems in signal and data processing.

6.2.4. *Collaborative source seeking*
Participants: C. Canudas de Wit [Contact person], R. Fabbiano, F. Garin, Y. Gaudfrin, J. Dumon.

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 research area is attracting a rapidly increasing interest, in particular in applications where 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 low sensitive to measurement noise. This work is the topic of the Ph.D. of Ruggero Fabbiano, and is described in papers under review.

Moreover, a hardware implementation of the source-seeking algorithm has been done during the internship of Yvan Gaudfrin, at GIPSA-LAB with the support of Jonathan Dumon. A description of the setup and videos of the source-seeking robot are available online: http://necs.inrialpes.fr/pages/platforms.php

6.3. Sensor networks: estimation and data fusion

6.3.1. *Data fusion approaches for motion capture by inertial and magnetic sensors*
Participants: H. Fourati [Contact person], A. Makni, A. Kibangou.

We are interested to motion capture (or attitude) by fusing Inertial and Magnetic Sensors. In [15], we present a viable quaternion-based Complementary Observer (CO) which is designed for rigid body attitude estimation. The CO processes data from a small inertial/magnetic sensor module containing tri-axial angular rate sensors, accelerometers, and magnetometers, without resorting to GPS data. The proposed algorithm incorporates a motion kinematic model and adopts a two-layer filter architecture. In the latter, the Levenberg Marquardt Algorithm (LMA) pre-processes acceleration and local magnetic field measurements, to produce what will be called the system's output. The system's output together with the angular rate measurements will become measurement signals for the CO. In this way, the overall CO design is greatly simplified. The efficiency of the CO is experimentally investigated through an industrial robot and a commercial IMU during human segment motion exercises. In a recent work [35], a viable quaternion-based Adaptive Kalman Filter (q-AKF) that is designed for rigid body attitude estimation. This approach is an alternative to overcome the limitations of the classical Kalman filter. The q-AKF processes data from a small inertial/magnetic sensor module containing triaxial gyroscopes, accelerometers, and magnetometers. The proposed approach addresses two challenges. The first one concerns attitude estimation during various dynamic conditions, in which external acceleration occurs. Although external acceleration is one of the main source of loss of performance in attitude estimation methods, this problem has not been sufficiently addressed in the literature. An adaptive algorithm compensating external acceleration from the residual in the accelerometer is proposed. At each step, the covariance matrix associated with the external acceleration is estimated to adaptively tune the filter gain. The second challenge is focused on the energy consumption issue of gyroscopes for long-term battery life of Inertial Measurement Units. We study the way to reduce the gyro measurement acquisition while maintaining acceptable attitude estimation. Through numerical simulations, under external acceleration and parsimonious gyroscope's use, the efficiency of the proposed q-AKF is illustrated.

6.3.2. *Pedestrian dead-reckoning navigation*

Participant: H. Fourati [Contact person].

We proposes a foot-mounted Zero Velocity Update (ZVU) aided Inertial Measurement Unit (IMU) filtering algorithm for pedestrian tracking in indoor environment [22]. The algorithm outputs are the foot kinematic parameters, which include foot orientation, position, velocity, acceleration, and gait phase. The foot motion filtering algorithm incorporates methods for orientation estimation, gait detection, and position estimation. A novel Complementary Filter (CF) is introduced to better pre-process the sensor data from a foot-mounted IMU containing tri-axial angular rate sensors, accelerometers, and magnetometers and to estimate the foot orientation without resorting to GPS data. A gait detection is accomplished using a simple states detector that transitions between states based on acceleration measurements [32]. Once foot orientation is computed, position estimates are obtained by using integrating acceleration and velocity data, which has been corrected at step stance phase for drift using an implemented ZVU algorithm, leading to a position accuracy improvement. We illustrate our findings experimentally by using of a commercial IMU during regular human walking trial in a typical public building. Experiment results show that the positioning approach achieves approximately a position accuracy less than 1 m and improves the performance regarding a previous work of literature [33].

6.3.3. *Sensor placement of unreliable sensors*

Participants: F. Garin [Contact person], P. Frasca [Twente].

In this work (see [23]), 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. However, as the number of sensors grows to infinity, the ratio between the cost of equally-spaced configurations and the optimal failure-free cost only grows as the logarithm of the number of sensors. We interpret this result as a confirmation of the intrinsic robustness of sensor networks.

6.4. Control design and co-design

6.4.1. *Energy-aware networked control*

Participants: C. Canudas de Wit [Contact person], F. Garin, N. Cardoso de Castro, D. Quevedo [Newcastle].

We have considered an event-based approach to energy-efficient management of the radio chip in the sensor node of a wireless networked control system. Indeed the radio is the main energy consumer, and intermittent data transmission allows one to reduce the use of the radio. While the existing literature in the control community on event-based control only addresses policies using two radio modes (transmitting/sleep), our work follows some considerations on the radio chip modes well-known in the communication networks literature, and introduces various radio-modes: different 'idle' non-transmitting modes, where only part of the radio chip is switched off (thus consuming more energy than 'sleep', but allowing for faster transition to transmission), and various transmitting modes, with different power levels. We propose an event-based radio-mode switching policy, which allows to perform a trade-off between energy saving and performance of the control application; to this end, a switched model describes the system, taking into account control and communication. The optimal switching policy is computed using dynamic programming, considering a cost either over an infinite time-horizon (see [31]) or over a finite receding horizon (joint work with D. Quevedo, Univ. Newcastle, Australia, described in a paper in preparation).

6.4.2. *Adaptive control strategy based reference model for spacecraft motion trajectory*

Participants: H. Fourati [Contact person], Z. Samigulina.

In aerospace field, the economic realization of a spacecraft is one of the main objectives which should be accomplished by conceiving the optimal propulsion system and the best control algorithms. Our work focuses on the development of a viable Adaptive Control Approach (ACA) for Spacecraft Motion Trajectory (SMT), see [39].The proposed strategy involves the nonlinear mathematical model of SMT expressed in the central field, which is linearized by the Taylor expansion, and the second Lyapunov method to offer a high rate and unfailing performance in the functioning. The adaptive control system is composed of the cascade of adaptation loop and feedback control loop. When the spacecraft deviates from its reference trajectory model, the ACA acts on the control system to correct this deviation and follow the optimal reference trajectory. Therefore, when the states of the adjustable model are different from its reference values, then the error signal is provided as an input to the adaptation law, which contains the adaptation algorithm. The output will be the state variable feedback control matrix which will be used to calculate the new control law vector. The efficiencies of the linearization procedure and the control approach are theoretically investigated through some realistic simulations and tests under MATLAB. The steady state errors of control between the reference model and the adjustable model of SMT converge to zero. This work is described in [38].

6.5. Transportation networks and vehicular systems

6.5.1. *Traffic estimation and prediction*

Participants: C. Canudas de Wit [Contact person], A. Kibangou, L. Leon Ojeda, F. Morbidi.

Reconstructing densities in portions of the road links not equipped with sensors constitutes an important task in traffic estimation, forecasting, and control problems. Among many other approaches, model-based observers is one popular technique to build this information. They can also be understood as *virtual sensors* deployed inside of the cells not equipped with *true sensors*. They are used to better track, in real-time, density variations with a fine degree of granularity in the space, as the *virtual cells* can be selected as small as desired.

In [43], a graph constrained-CTM observer was introduced. It allows reconstructing rather accurately the internal states (densities) of a road portion not equipped with sensors. This strategy for real-time density estimation was applied on Grenoble South Ring. In [27], this observer has been associated with an adaptive Kalman filtering approach for traffic prediction in terms of travel time. The adaptive Kalman filtering approach was also been used for predicting input flows in [26].

6.5.2. *Traffic control*

Participants: C. Canudas de Wit [Contact person], D. Pisarski.

This work has been conducted in two parallel directions, combining steady state analysis and design of an optimal ramp metering controller.

The first direction was to extend the preliminary results presented in the papers [10] and [46]. The goal was to implement the idea of optimal steady state balancing. A relevant software was built up and tested on the model representing the south ring of Grenoble. The results were published in [28]. A comprehensive study of steady state balancing was submitted as a journal paper, under review.

The second direction was to develop a distributed optimal ramp metering controller. This study is motivated by two following facts. The first one is to decentralize and parallelize computation for optimal freeway traffic control problem, and thus to reduce computational complexity. The second one is to reduce the lengths of the communication channels, in order to eliminate the probability of information delay or packet loss. The proposed new control objective provides a uniformly distributed (or balanced) vehicle density such that the usage of freeway (measured by the Total Travel Distance and Total Input Volume) is maximized. This optimal balancing objective is reached by taking a proper state feedback control structure and optimizing the set of gains. Here we imposed distributed condition for both, the feedback structure and the optimization process. We have focused the efforts to design the controller network architecture that is based on the common elements (ramp meter controllers), executing the same computational procedures and applying the control signals based on the same state-feedback structure. This meets a spirit of 'plug and play' (PnP), and is beneficial for both, architecture assembling and component replacement (in case of failure). In order to define the functionality for each of this PnP controller, the analysis on both system controllability and conditions for optimality were carried on. The preliminary work let us to determine the what type of information and upon which communication topology it is required to be sent in order to solve the posed optimization problem. Firstly, the feedback controls for each of the controllers require state information of the section that is controllable for it. In general, each of the controllers demands the state for its closest surrounding sections (downstream and upstream). Secondly, each of the controllers communicates with its closest active neighbors to exchange the information of optimal solution, namely optimal boundary flow or optimal control. We also observed that in any system mode there might be only one inactive controller (the controller that does not have any controllable section) surrounded by two active ones, and thus the maximum required information comes from the two closest neighbors for each of the directions. We noticed also that inactive controllers may serve to convey the information for the active ones, so the communication can be based on a path (or linear) graph. Part of this research was realized in UC Berkeley during the visit of Dominik Pisarski in PATH laboratory.

Traffic control is based on models of traffic, usually the so-called CTM – Cell transmission Model. Some work in the team aims at developing different models, more suitable for control. One such model is based on cells of variable length, as an alternative way to describe the congestion position. This model, proposed in [42], has been refined in the master thesis of Giulio Bontadini, taking into account mass conservation laws.

6.5.3. *Vehicle control for disabled people*

Participants: C. Canudas de Wit [Contact person], V. Ciarla, J. Dumon, F. Quaine [UJF], V. Cahouet [UJF].

Disabled people face the effort to turn the steering wheel while driving their vehicle. This study, funded by the VolHand project, focuses on the aspect of the assistance during driving maneuver at low speeds (for instance, parking). On common vehicles for healthy people, the system that improves the driver's steering feel in these situations is the power-steering stage, which is mounted at the basis of the steering column and is based on hydraulic technology; the new generation uses an electric motor instead of the hydraulic pump, with more advantages in terms of fuel consumption, better road-feel feedback to the driver and better return-to-center performances of the steering wheel. This work has developed a general methodology to adapt the current technology for disabled people, by introducing additional blocks that can be implemented via software without altering the hardware of the vehicle. In this way, it can be easily exported without additional costs in terms of design and technology for the industrial partner. The methodology has been studied theoretically, joining control aspects with bio-mechanical ones. Moreover, the theoretical study has been tested in laboratory on the

hardware-in-the loop setup, using the experimental platform NeCSCar (see Section 5.2). First, a real steering wheel has been linked to a real-time PC-unit and to an electrical motor. A graphical user interface has been implemented to facilitate the access to the software. Then, the last part of the study has been the experimental validation with a tele-operated real vehicle. The vehicle provided the measure of the friction torque to the PC-unit, simulating a real driving situation.

This work is described in [41] and in the Ph.D. thesis [11].

6.5.4. *Control of communicating vehicles in urban environment*

Participants: C. Canudas de Wit [Contact person], G. de Nunzio.

For a given vehicle there are different ways to travel on a given distance in a given time, associated to different levels of energy consumption; therefore, it is always possible to find an energy-optimal trajectory. Advising the driver via a suitable interface can reduce the energy consumed during the travel, and thus improve the energy efficiency: this is the principle of eco-driving. In urban areas, the optimal trajectory of the vehicle depends on interactions with other vehicles, also on passive signs (panels, priorities, etc.) and active signs (traffic lights); in each case, constraints are imposed on the command (vehicle speed). From the infrastructure perspective, traffic control in urban areas consists in determining the state of traffic signals in order to solve an optimization problem, for example minimizing average travel time of vehicles in the road network. If all vehicles could communicate with one another and with the active infrastructure (traffic lights), we could imagine benefits for each of the two problems which can be considered as a whole: on the one hand, from the vehicles' point of view, more information is available that can be integrated into the online optimization problem; on the other hand, there are new measures and new commands available to control traffic. Indeed, the estimation of the traffic is no longer necessary, as the position and speed of approaching vehicles is known and shared. More importantly, the traffic manager can send instructions to the vehicles. The aim of the research is to evaluate the potential in terms of energy saving and traffic improvement made possible by communicating vehicles. This work is the topic of the Ph.D. thesis of Giovanni De Nunzio, a CIFRE thesis with IFPEN. The paper [21] considers the scenario where vehicle and infrastructure (traffic lights) can communicate, and describes a suitable optimization algorithm that can be run on-board the vehicle so to optimize its energy consumption by avoiding stops and abrupt changes of speed at traffic lights, thanks to the information on upcoming traffic lights on the same road.

NUMED Project-Team (section vide)

<p align="center" style="color:red">OPALE Project-Team</p>

6. New Results

6.1. Mathematical analysis and control of macroscopic traffic flow models

6.1.1. Vehicular traffic

Participants: Alessandra Cabassi, Maria Laura Delle Monache, Paola Goatin, Alexandre Bayen [UC Berkeley, CA, USA], Legesse Lemecha Obsu [Addis Ababa University, Ethiopia].

In collaboration with UC Berkeley, and as part of the Associated Team ORESTE activity (see http://www-sop.inria.fr/members/Paola.Goatin/ORESTE/index.html), we have proposed a new junction model for ramp metering: we introduce a coupled PDE-ODE model, in which the PDE describes the evolution of the cars flow on the main lane and the ODE describes the evolution of the queue length on the on-ramp, modeled by a buffer, which ensures that boundary conditions are satisfied in strong sense. We were able to prove existence and uniqueness of the solution of the corresponding Riemann problem [41]. Relying on the above junction model, we have applied the Discrete Adjoint Method to efficiently compute (locally) optimal ramp-metering parameters to minimize the total travel time on a stretch of highway [80].

In parallel, we have proposed two optimization strategy for instantaneous optimization of total travel times and total waiting times at roundabouts, which give an estimate of the time spent by drivers on the network section. These cost functionals are minimized with respect to the right-of-way parameter of the incoming roads. For each cost functional, the analytical expression is given for each junction, see [72]. This work is part of L.L. Obsu's PhD thesis.

Finally, we designed a new finite volume algorithm to track the trajectory of a bus in the surrounding traffic using a locally non-uniform moving mesh, see [3, 4, 5].

As part of our TRAM3 activity, we also organized the workshop "TRAM2 - Traffic Modeling and Management: Trends and Perspectives", which successfully took place at Inria Sophia Antipolis on March 20-22, 2013 (see https://team.inria.fr/opale/workshop-tram2/).

In the framework of the EIT ITC Labs Multimodal Mobility activity, A. Cabassi's internship was devoted to the calibration and the validation of a first order traffic flow model against processed real data provided by the industrial partners Autoroutes Traffic and VINCI Autoroutes, see [69].

6.1.2. Crowd motion

Participants: Régis Duvigneau, Paola Goatin, Matthias Mimault, Debora Amadori [L'Aquila University, Italy], Christophe Chalons [LJLL, UP7], Massimiliano D. Rosini [ICM, Warsaw University, Poland], Nicolas Seguin [LJLL, UPMC], Monika Twarogowska.

From the analytical point of view, we have been studying the properties of some models in one space dimension. Concerning Hughes' scalar model, we have established a partial existence result in collaboration with D. Amadori and M.D. Rosini (see [75]). M. Mimault's internship in 2012 was devoted to develop a MATLAB code based on wave-front tracking to compute the solutions of Hughes' model with generalized running cost, see [42]. He is currently working on a mixed hyperbolic-elliptic 2x2 system of conservation laws describing two groups of people moving in opposite directions. Finally, in collaboration with C. Chalons and N. Seguin, we generalized previous results on conservation laws with local flux constraints [3], [5] to general flux functions and non-classical solutions arising in pedestrian flow modeling, see [39]. From the numerical point of view, we have implemented some macroscopic models in 2D on unstructured triangular meshes on the Num3sis platform. We provided a comparison between first and second order models in reproducing complex dynamics of crowd motion, such as formation of stop-and-go waves and clogging at bottlenecks. Then, we concentrated on the higher-order model and analyzed the dependence of the behavior of its solutions on some of the parameters of the system. In particular, we produced some examples where placing obstacles in front of the door prevents from blocking and decreases the evacuation time, see [73], [81].

The above researches were partially funded by the ERC Starting Grant "TRAM3 - Traffic management by macroscopic models".

6.2. Optimum design and control in fluid dynamics and its couplings

In computational sciences for physics and engineering, Computational Fluid Dynamics (CFD) are playing one of the major roles in the scientific community to foster innovative developments of numerical methodologies. Very naturally, our expertise in compressible CFD has led us to give our research on numerical strategies for optimum design a particular, but not exclusive focus on fluids.

The framework of our research aims to contribute to numerical strategies for PDE-constrained multi-objective optimization, with a particular emphasis on CPU-demanding computational applications in which the different criteria to be minimized (or reduced) originate from different physical disciplines that share the same set of design variables. These disciplines are often fluids, as a primary focus, coupled with some other disciplines, such as structural mechanics.

Our approach to *competitive optimization* is focused on the two-discipline problem. It is based on a particular construction of *Nash games*, relying on a *split of territory* in the assignment of individual strategies. A methodology has been proposed for the treatment of two-discipline optimization problems in which one discipline, the primary discipline, is preponderant, or fragile. Then, it is recommended to identify, in a first step, the optimum of this discipline alone using the whole set of design variables. Then, an orthogonal basis is constructed based on the evaluation at convergence of the Hessian matrix of the primary criterion and constraint gradients. This basis is used to split the working design space into two supplementary subspaces to be assigned, in a second step, to two virtual players in competition in an adapted Nash game, devised to reduce a secondary criterion while causing the least degradation to the first. The formulation has been proved to potentially provide a set of Nash equilibrium solutions originating from the original single-discipline optimum point by smooth continuation, thus introducing competition gradually [53]. (see also subsectionsubsect:helico).

Our approach to *cooperative optimization*, in theory, is not limited in number of objective functions. It is based on a result of convex analysis established for a general unconstrained mult-iobjective problem in which all the gradients are assumed to be known. The theorem [16] states that in the convex hull of the gradients, there exists a unique vector of minimal norm, ω; if it is nonzero, the vector ω is a descent direction common to all criteria; otherwise, the current design point is Pareto-stationary. This result led us to generalize the classical steepest-descent algorithm by using the vector ω as search direction. We refer to the new algorithm as the multiple-gradient descent algorithm (MGDA). The MGDA yields to a Pareto-stationary point, and actual Pareto-optimality is then tested [54] (see also subsection 6.2.4).

The two approaches have been combined to explore the Pareto front segment-wise as illustrated on Figure 2 .

6.2.1. Multiple-Gradient Descent Algorithm (MGDA)

Participants: Jean-Antoine Désidéri, Régis Duvigneau, Matteo Giacomini, Abderrahmane Habbal, Adrien Zerbinati.

6.2.1.1. Theory and numerical experimentation of the MGDA construction

In multi-objective optimization, the knowledge of the Pareto set provides valuable information on the reachable optimal performance. A number of evolutionary strategies (PAES, NSGA-II, etc), have been proposed in the literature and proved to be successful to identify the Pareto set. However, these derivative-free algorithms are very demanding in terms of computational time. Today, in many areas of computational sciences, codes are developed that include the calculation of the gradient, cautiously validated and calibrated.

The notion of Pareto-stationarity, originally established to be a necessary condition of optimality in differentiable multi-objective optimization of unconstrained problems, has been extended to problems subject to equality constraints. On this basis, we were able to establish that by augmenting, in a classical manner, the objective-functions of a penalty term equal to the square of the constraint violation, and applying the standard MGDA to it, would result in converged solutions that are Pareto-stationary in the extended sense. Numerical experimentation on this is on-going.

../../../../projets/opale/IMG/nash-mgda.png

Figure 2. ***Two-discipline optimization of a generic geometry of a supersonic aircraft, for concurrent drag and sonic-boom reduction*** *(from A. Minelli's doctoral thesis). The wave drag is calculated by the ONERA elsA code in 3D finite-volume Eulerian flow mode over a 6M-node mesh and the sonic boom using a three-layer approach. The Nash-game paths have been devised by appropriate territory splitting in order to be tangent to the Pareto front, and they are interrupted whenever the Pareto-stationarity condition is judged excessively violated. The MGDA paths converge rapidly back to the front. The simulation demonstrates how the two algorithms complement each other and provide a potential for a piecewise description of the Pareto front, evaluated more economically than a stochastic algorithm operating on a large population.*

6.2.1.2. Meta-model-assisted CFD optimization by MGDA

Using MGDA in a multi objective optimization problem requires the evaluation of a large number of points with regard to criteria, and their gradients. In the particular case of a CFD problems, each point evaluation is very costly since it involves a flow computation, possibly the solution of an adjoint-equation. To alleviate this difficulty, we have proposed to construct meta-models of the functionals of interest (lift, drag, etc) and to calculate approximate gradients by local finite differences. These meta-models are updated throughout the convergence process to the evaluation of the new design points by the high-fidelity model, here the 3D compressible Euler equations.

This variant of MGDA has been tested successfully over a problem of external aerodynamic optimum-shape design of an aircraft wing consisting of reducing wave-drag, and augmenting lift. After only a few cycles of database updates, the Pareto front visibly forms, and this result is achieved at a very moderate computational cost [68]. This variant has been extended successfully to an internal flow optimization problem related to an automobile air-conditioning system and governed by the Navier-Stokes equations. This more difficult problem has been proposed by Renault within the OMD2 ANR project. These studies have been reported in A. Zerbinati's doctoral thesis [38].

6.2.1.3. Exact shape gradients

MGDA has successfully been tested over a two-objective optimization problem governed by two-dimensional elasticity. The deformation of a plate is calculated using an isogeometric approximation (see 6.3) and compliance derived from it. The exact parametric shape gradient is calculated, yielding the gradient of the objective function in two antagonistic situations differing by the loading. Pareto-fronts are thus identified.

6.2.1.4. Perspectives

MGDA offers the possibility to handle in a rational way several objective-functions for which gradients are known or approximated concurrently. This potential opens methodological paths to several themes of interest in high-fidelity simulation-based optimization: optimization of complex systems whose performance is evaluated w.r.t. several criteria originating from different, coupled disciplines; optimization under uncertainties, by introducing sensitivities as additional objectives; optimization of time-dependent systems, such as optimization of flow-control devices that generate a periodic flow (see next subsection), by converting the problem into a multi-point problem by time-discretization of the time and parameter-dependent functional; etc.

6.2.2. *Flow control*

Participants: Régis Duvigneau, Jérémie Labroquère, Emmanuel Guilmineau [Ecole Centrale de Nantes].

Shape optimization methods are not efficient to improve the performance of fluid systems, when the flow is characterized by a strong unsteadiness related to a massive detachment. This is typically the case for the flow around an automotive body or a wing in stall condition. To overcome this difficulty, flow control strategies are developed, that aim at manipulating vortex dynamics by introducing some active actuators, such as periodic blowing/suction jets. In this context, the choice of the control parameters (location, amplitude, frequency) is critical and not straightforward. Therefore, we develop a methodology to determine optimal control parameters by coupling the simulation of unsteady actuated flows with optimization algorithms. Two research axes have been considered :

- the resolution of the unsteady sensitivity equations derived from the state equations, to exhibit the dependency of the flow dynamics with respect to the control ;
- the optimization of control parameters using a statistical metamodel-based strategy.

In this perspective, unsteady Reynolds Averaged Navier-Stokes equations are solved, with some turbulence closures. Different models for synthetic jet have been implemented to simulate the actuation, and then validated for different turbulence closures [70].

Specific developments have be carried out in the metamodel-based optimizer to include a noise term into Gaussian Process model, which is used to filter errors arising from unsteady simulations. A systematic assessment of modeling and numerical errors has been archived [57], for a backward facing step test-case, with the objective of controlling the re-attachment point location.

This activity is conducted in collaboration with the CFD team of Ecole Centrale de Nantes.

6.2.3. *Robust design*
Participants: Jean-Antoine Désidéri, Régis Duvigneau, Daïgo Maruyama.

This work aims at developing robust design tools for aircraft w.r.t. aerodynamic performance subject to uncertainties, arising from geometrical features and fluctuations of inflow conditions. The robust design process is considered as a multi-objective optimization problem, which consists in minimizing or maximizing statistical moments of the cost function.

In the context of airfoil design, MGDA is used to improve simultaneously the mean and variance of the lift and drag coefficients, yielding a four-objective optimization problem [71].

6.2.4. *Sonic boom reduction*
Participants: Gérald Carrier [Research Engineer, ONERA/DAAP], Jean-Antoine Désidéri, Andrea Minelli, Itham Salah El Din [Research Engineer, ONERA/DAAP].

When an aircraft flies at supersonic speed, it generates at ground level an N-shaped shock structure which can cause serious environmental damage ("sonic boom"). Thus a problem of interest in aerodynamic optimization is to design such an aircraft to reduce the intensity of the sonic boom while maintaining the aerodynamic performance (drag minimization under lift constraint). Andrea Minelli aimed at contributing to this two-discipline optimization problem. In the first part of his work, an inverse problem has been formulated and solved for "shaped sonic boom" and found in excellent agreement with the George-Seebass-Darden theory [82] for the calculation of the Whitham function corresponding to the lowest-boom (axisymmetric) shape. Method and results have been generalized to more general geometries and have been presented internationally in [58].

Besides, aero-acoustic optimizations have been realized successfully by coupling the aerodynamic optimizer (based on Euler calculations by the elsA software) with the sonic-boom computation in a Nash game formulation. These experiments, conducted with our optimization platform FAMOSA, have demonstrated that starting from the shape optimized aerodynamically, one could retrieve smoothly a shape corresponding to nearly-optimal sonic-boom reduction [36]. and [54].

6.2.5. *Helicopter rotor blade optimization in both situations of hovering and forward flight*
Participants: Michel Costes [Research Engineer, ONERA/DAAP], Jean-Antoine Désidéri, Arnaud Le Pape [Research Engineer, ONERA/DAAP], Enric Roca Leon.

E. Roca Leon is conducting a CIFRE thesis supported by EUROCOPTER (Marignane) at ONERA DAAP. This thesis follows the doctoral thesis of A. Dumont in which the adjoint-equation approach was used to optimize a rotor blade in hovering flight. The goal of this new thesis is to solve a two-objective optimization problem in which the hovering-flight criterion is considered preponderant, but a new criterion that takes into account the forward-flight situation is also introduced, concurrently. The second criterion is the power necessary to maintain the forward motion. The first phase of thesis work has been devoted to the set up of a hierarchy of models from low to high fidelity, in order to calibrate appropriate functional criteria. Then, actual two-objective optimizations are conducted via our Nash game approach to competitive optimization with territory splitting based on reduced Hessian diagonalization. A first successful experiment has been realized in which 16 geometrical parameters have been optimized to reduce the power in forward motion while maintaining sub-optimality of the drag in hover. These results have been accepted for presentation at the American Helicopter Society Forum [62], and [53].

6.2.6. *Optimum design in naval hydrodynamics*
Participants: Régis Duvigneau, Louis Blanchard, Elisa Berini [K-Epsilon company].

Naval hydrodynamics field has recently shown a growing interest for optimum design methods. The computational context is especially complex because it implies unsteady two-phase turbulent flows, with possibly very high Reynolds number (up to 10^9). The use of automated design optimization methods for such problems requires new developments to take into account the large CPU time necessary for each simulation and the specificity of the geometries considered.

Some developments have been initiated on the geometrical modelling of hull shapes by parametric surfaces. The objective was to be able to modify existing hull shapes by controlling a small number of parameters, that are meaningful for naval architects. We have considered as test-case the bow shape for trawler ships. As a second step, an optimum shape procedure has been set up, based on a metamodel-based optimizer, the developed CAD model and the simulation tool for free-surface flows provided by K-Epsilon company. The objective was to reduce the wave drag of a trawler ship by adding a bow, whose parameters are optimized [50].

6.3. Isogeometric analysis and design

Participants: Régis Duvigneau, Bernard Mourrain [Galaad project-team], Alexandros Ginnis [Nat. Tech. Univ. of Athens], Bernd Simeon [Tech. Univ. of Kaiserslautern], Gang Xu [Hangzhou Dianzi Univ.].

Design optimization stands at the crossroad of different scientific fields (and related software): Computer-Aided Design (CAD), Computational Fluid Dynamics (CFD) or Computational Structural Dynamics (CSM), parametric optimization. However, these different fields are usually not based on the same geometrical representations. CAD software relies on Splines or NURBS representations, CFD and CSM software uses grid-based geometric descriptions (structured or unstructured), optimization algorithms handle specific shape parameters. Therefore, in conventional approaches, several information transfers occur during the design phase, yielding approximations that can significantly deteriorate the overall efficiency of the design optimization procedure. Moreover, software coupling is often cumbersome in this context.

The isogeometric approach proposes to definitely overcome this difficulty by using CAD standards as a unique representation for all disciplines. The isogeometric analysis consists in developing methods that use NURBS representations for geometric modeling, computational domain description and solution basis functions. Using such a unique data structure allows to compute the solution on the exact geometry (not a discretized geometry), obtain a more accurate solution (high-order approximation), reduce spurious numerical sources of noise that deteriorate convergence, avoid data transfers between the software. Moreover, NURBS representations are naturally hierarchical and allows to define multi-level algorithms for solvers as well as optimizers.

In this context, some studies on elliptic problems have been conducted in collaboration with the Galaad project-team and Hangzhou Dianzi University, such as the development of methods for adaptive parameterization including an a posteriori error estimate [46], [47], [48]. A collaborative work has also been carried out with the Technical University of Kaiserslautern, concerning the computation of shape gradients for linear elasticity problems, and with the National Technical University of Athens for hull shape optimization [55].

6.4. Optimum design in structural mechanics

6.4.1. *Shape Optimization in Multidisciplinary Non-Linear Mechanics*

Participants: Aalae Benki, Jean-Antoine Désidéri, Abderrahmane Habbal, Gael Mathis [ArcelorMittal, CRAA].

In collaboration with the ArcelorMittal's Center for Research in Automotive and Applications (CRAA), we study the multidisciplinary shape and parameter design of highly non linear mechanical 2D and 3D structures. We have developed methods adapted to the approximation of Pareto Fronts such as Normal Boundary Intersection NBI and Normalized Normal Constraint Method NNCM. Due to the time consuming cost evaluation, the use of cheap to evaluate surrogate models is mandatory. We have studied the consistency of the approach NBI or NNCM plus surrogates, which turned out to be successful for a broad panel of standard mathematical benchmarks. The coupling is successfully applied to a small scale industrial case, namely the shape optimization of a can bottom vis à vis dome reversal pressure and dome growth criteria. We have

then defined a Nash game between criteria where the latter are approximated by the RBF metamodels. First, we validated the computation of a Nash equilibrium for mathematical functions, then we computed Nash equilibria for the small scale industrial case of the shape optimization of the can bottom.

Then, we considered the 3D problem of an automotive twist beam. In this 3D case, we aim to Pareto-optimal shapes for two objectives, the first being to minimize the Von-Mises strain to guarantee the formability of the twist beam, and the second being to maximize the stiffness. For solution with higher stiffness than the initial one, we could decrease the thickness to obtain a mass reduction with the same end-user properties.

We also introduced, to our knowledge for the first time in the structural optimization area, the notion of Kalai-Smorodinky equilibria which is aimed at the selection of equilibria among Pareto-optimal solutions. We applied this notion of equilibria to both industrial cases, and compared the results to Nash equilibria. [56] [64]

../../../../projets/opale/IMG/DRPDG2D.png ../../../projets/opale/IMG/TwistBeam3D.png

Figure 3. Concurrent design in industrial applications. A packaging problem of commercial cans (left). Automotive twist beam (right)

6.4.2. *Optimization of Addendum Surfaces in Stamping*

Participants: Fatima Zahra Oujebbour, Rachid Ellaia, Abderrahmane Habbal, Ziheng Zhao.

Within the OASIS Consortium (ArcelorMittal, ErDF, Inria, UTC, EURODECISION, ESILV, NECS, Delta-CAD, SCILAB-DIGITEO), the Opale project-team leads the Optimization task. Our aim is to develop decentralized decision-making algorithms dedicated to find efficient solutions (Pareto optimal) in a complex multidisciplinary framework (forming, stamping, welding non-linear processes, spring-back, vibration, in-function linear processes, crash and fatigue non linear and non differentiable processes) for several (between three and five) criteria. An important difficulty when trying to identify the Pareto Front, even when using adapted methods such the Normal Boundary Intersection, is that the criteria involved (thanks to the high nonlinearity in the mechanical models) exhibit many local optima. So one must use global optimization methods. We have studied the hybrid approach Simulated Annealing with Simultaneous Perturbation SASP for a suite of mathematical test-cases. To envisage the application of our method to the complex CPU time consuming stamping process, we lead an intermediate phase dedicated to the validation of the SASP method for the minimization of the spring-back that follows the stamping of a metal sheet, the design variable being the process parameters (two then four parameters). Then, we considered the more complex shape design of the initial blank. The initial

blank design is a critical step in stamping design procedure, therefore it should be optimally designed. Our aim is to find the optimal initial blank shape that avoids or at least minimizes the springback and failure flaws. For this study, the geometry of the blank contour is described by parametric spline curves. Seven control points (P1,...,P7) are used to define the spline curves in order to have a wide variety of geometries. The exact computational evaluation of our criteria, springback and failure, is very expensive (the FE model request around 45 min to predict these two criteria) and the design space is of quite high dimension. Therefore, we considered the recourse to the sparse grid interpolation. Optimization process based on sparse grid interpolation is an optimal alternative in which criteria can be approximated with a suitable interpolation formula that needs significantly less points than the full grid. the obtained metamodel using sparse grid interpolation needs less than 1s to predict springback and failure on the same computation machine. To find the optimal initial blank shape, it was decided to perform the optimization process using the obtained metamodel. The construction of the sparse grid interpolant was based on the Chebyshev Gauss-Lobatto grid type and using the polynomial basis functions. This technique achieves a good accuracy with a competitive number of grid points. The comparison of the obtained fronts shows that we can capture Pareto solutions by NBI and NNCM with fewer points than NSGAII which requires a large number of populations and several generations to obtain the Pareto front. [60] [61] [63] [77]

../../../../projets/opale/IMG/embout/1.png ../../../../projets/opale/IMG/embout/2.png ../../../../projets/opale/IMG/BlankS

../../../../projets/opale/IMG/NBI/.png ../../../../projets/opale/IMG/NNCM.png ../../../../projets/opale/IMG/NSGAII
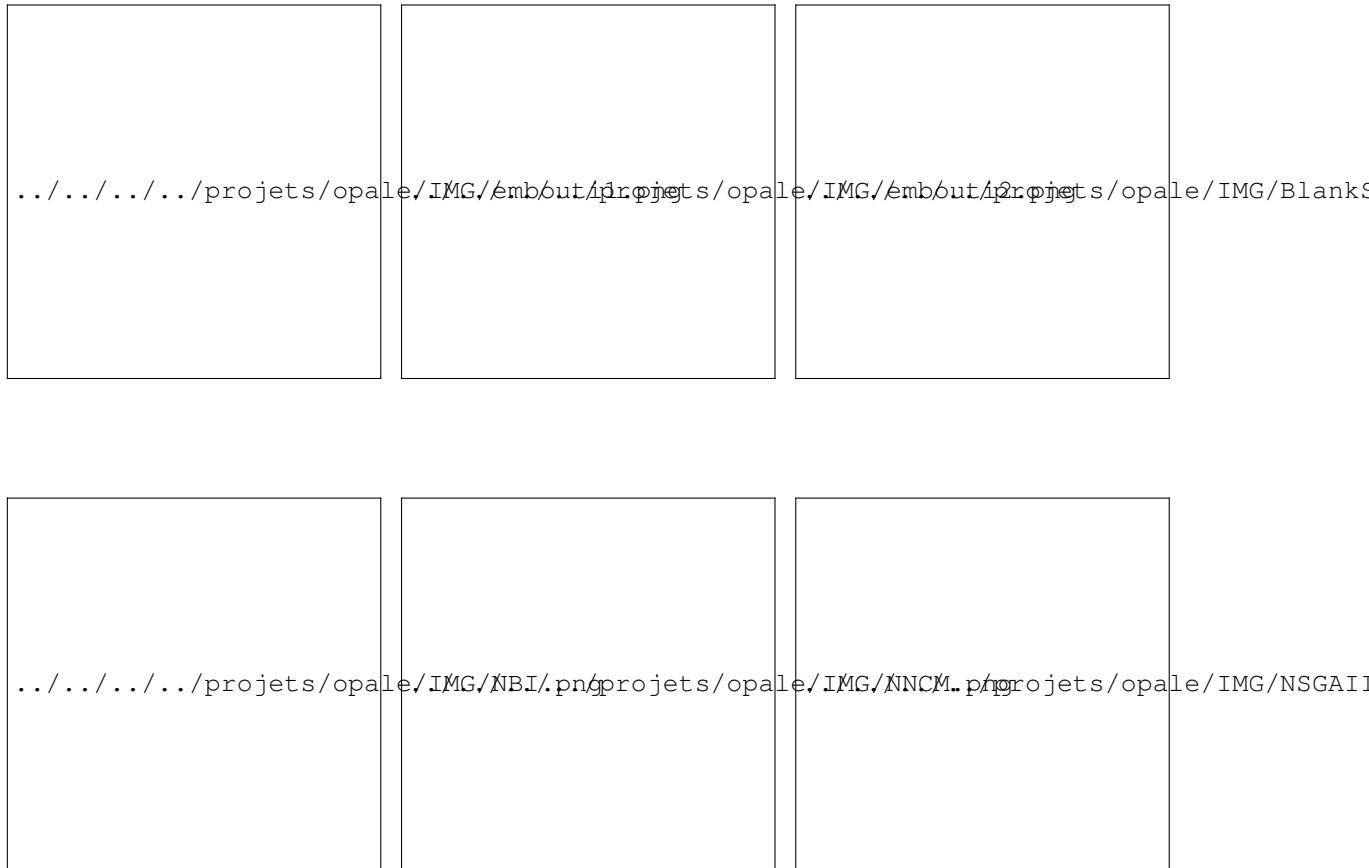
Figure 4. Multiobjective design of the stamping process of a high performance steel sheet. The design variable is the initial blank shape, and the costs are elastic spring-back and failure. Sparse grid approximation of the costs is used. The Pareto front obtained by NBI and NNCM (lower-left) are compared to a NSGA-II one (lower-right).

6.5. Application of shape and topology design to biology and medicine

6.5.1. *Assessing the ability of the 2D Fisher-KPP equation to model cell-sheet wound closure*

Participants: Abderrahmane Habbal, Hélène Barelli [Univ. Nice Sophia Antipolis, CNRS, IPMC], Grégoire Malandain [Inria, EPI Morpheme].

We address in this joint collaboration the ability of the widely used Fisher-KPP equations to render some of the dynamical features of epithelial cell-sheets during wound closure.

Our approach is based on nonlinear parameter identification, in a two-dimensional setting, and using advanced 2D image processing of the video acquired sequences. As original contribution, we lead a detailed study of the profiles of the classically used cost functions, and we address the "wound constant speed" assumption, showing that it should be handled with care.
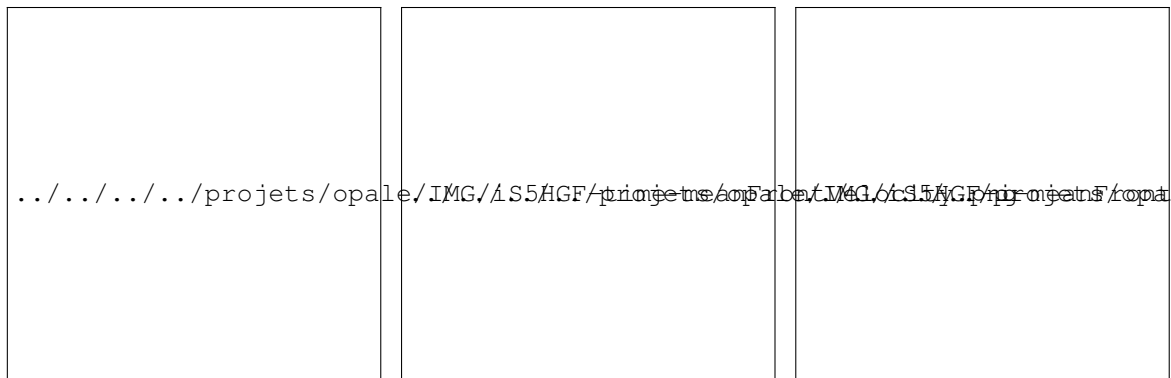
We study five MDCK cell monolayer assays in a reference, activated and inhibited migration conditions. Modulo the inherent variability of biological assays, we show that in the assay where migration is not exogeneously activated or inhibited, the wound velocity is constant. The Fisher-KPP equation is able to accurately predict, until the final closure of the wound, the evolution of the wound area, the mean velocity of the cell front, and the time at which the closure occurred. We also show that for activated as well as for inhibited migration assays, many of the cell-sheet dynamics cannot be well captured by the Fisher-KPP model. Original unexplored utilizations of the model such as wound assays classification based on the calibrated diffusion and proliferation rate parameters is ongoing.[49] [76]

../../../../projets/opale/IMG/iS5HGF-areaetvariparlievIMGyiS5HGF-wooinet-edopabngIMG/S5HGFr

(a)   (b)   (c)

../../../../projets/opale/IMG/iS5HGF-prioietne-corabetMG/cS15AGEphirmeetnTropaVel1MG/S5pGEn

(d)   (e)   (f)

Figure 5. **A regular wound assay** *(a) Time evolution of wound area (in pixel). (b) Time evolution of the leading-edge length (in pixel). (c) 3D XT view at first and mid-rows. (d) Mean (in time) velocity of pixels located at the leading edge (in pixel/min). (e) Averaged (in space) leading-edge velocity (in pixel/min). (f) 2D XT view at first and mid-rows.*

../../../../projets/opale/IMG/S5exp_pdeo-area.opale/IMG/S5exp_pdeo-Mitgr-dpiahRaIMG/S5HGFp

(a)   (b)   (c)
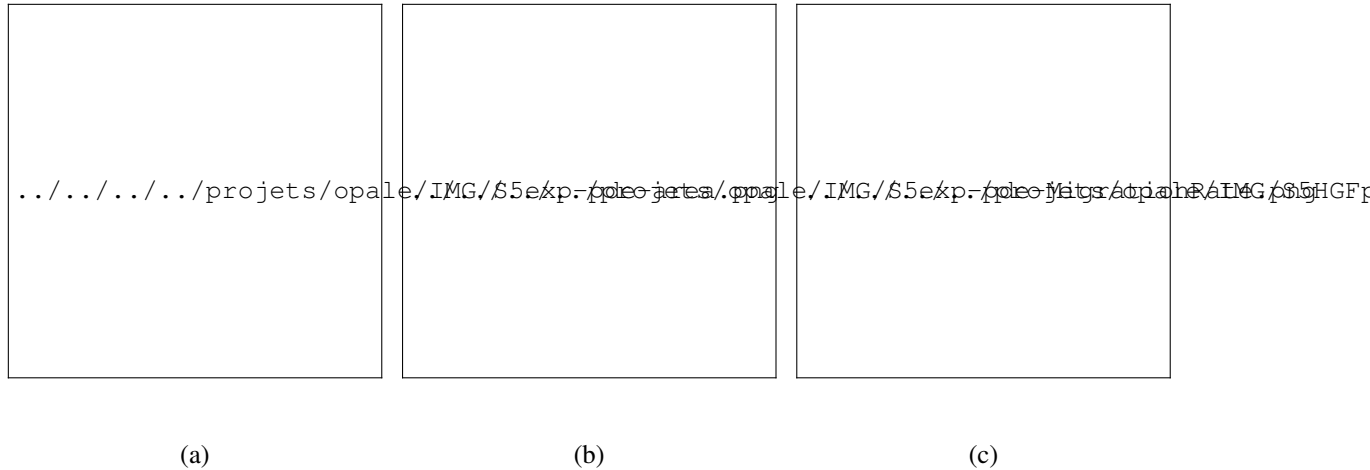
*Figure 6. **A regular wound assay. Computational vs experimental wound evolution.** (a) Time variation of experimental (blue) versus computed (red) wound area (in pixel). (b) Time variation of the experimental (blue-dot) versus computed (red) migration rate (in pixel/min). (c) 3D XT view at first and mid-rows.*

../../../../projets/opale/IMG/S4exp_pdeo-area.opale/IMG/S4exp_pdeo-Mitgr-dpiahRaIMG/S4HGFp

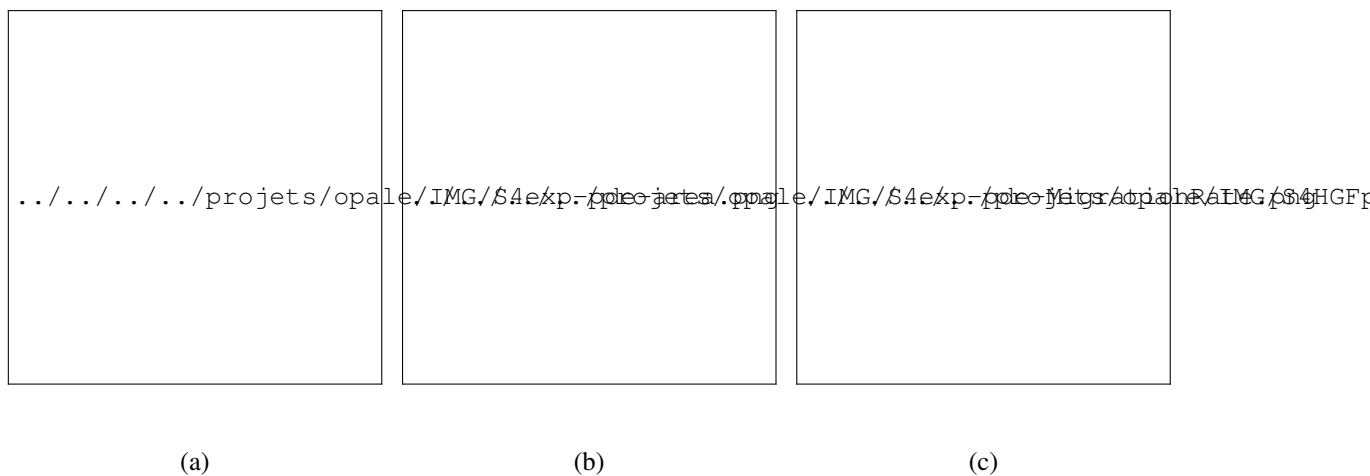(a)   (b)   (c)

*Figure 7. **An accelerating activated wound assay. Computational vs experimental wound evolution.** (a) Time variation of experimental (blue) versus computed (red) wound area (in pixel). (b) Time variation of the experimental (blue-dot) versus computed (red) migration rate (in pixel/min). (c) 3D XT view at first and mid-rows.*

PERCEPTION Team

6. New Results

6.1. High-resolution depth maps based on TOF-stereo fusion

The combination of range sensors with color cameras can be very useful for a wide range of applications, e.g., robot navigation, semantic perception, manipulation, and telepresence. Several methods of combining range- and color-data have been investigated and successfully used in various robotic applications. Most of these systems suffer from the problems of noise in the range-data and resolution mismatch between the range sensor and the color cameras, since the resolution of current range sensors is much less than the resolution of color cameras. High-resolution depth maps can be obtained using stereo matching, but this often fails to construct accurate depth maps of weakly/repetitively textured scenes, or if the scene exhibits complex self-occlusions. Range sensors provide coarse depth information regardless of presence/absence of texture. The use of a calibrated system, composed of a time-of-flight (TOF) camera and of a stereoscopic camera pair, allows data fusion thus overcoming the weaknesses of both individual sensors. We propose a novel TOF-stereo fusion method based on an efficient seed-growing algorithm which uses the TOF data projected onto the stereo image pair as an initial set of correspondences. These initial "seeds" are then propagated based on a Bayesian model which combines an image similarity score with rough depth priors computed from the low-resolution range data. The overall result is a dense and accurate depth map at the resolution of the color cameras at hand. We show that the proposed algorithm outperforms 2D image-based stereo algorithms and that the results are of higher resolution than off-the-shelf color-range sensors, e.g., Kinect. Moreover, the algorithm potentially exhibits real-time performance on a single CPU. Two journal papers were submitted in 2013 and currently they are under review.

6.2. Continuous action recognition

Continuous action recognition is more challenging than isolated recognition because classification and segmentation must be simultaneously carried out. We build on the well known dynamic time warping (DTW) framework and devise a novel video alignment technique, dynamic *frame* warping (DFW), which performs isolated recognition based on a per-frame representation of videos and on aligning a test sequence with a model sequence. Next we devise two extensions which are able to perform action recognition and video segmentation in a concomitant manner, namely one-pass DFW and two-pass DFW. Both these algorithms have their roots in the continuous speech recognition domain but, to the best of our knowledge, their extension to visual recognition of actions and activities has been overlooked. We test and illustrate the proposed methods with several public-domain datasets and we compare both the isolated and continuous recognition algorithms with several recently published methods. One journal paper was submitted in 2013 and currently is under review.

6.3. High-dimensional regression

We addressed the problem of approximating high-dimensional data with a low-dimensional representation. We make the following contributions. We propose an inverse regression method which exchanges the roles of input and response, such that the low-dimensional variable becomes the regressor, and which is tractable. We introduce a mixture of locally-linear probabilistic mapping model that starts with estimating the parameters of inverse regression, and follows with inferring closed-form solutions for the forward parameters of the high-dimensional regression problem of interest. Moreover, we introduce a partially-latent paradigm, such that the vector-valued response variable is composed of both observed and latent entries, thus being able to deal with data contaminated by experimental artifacts that cannot be explained with noise models. The proposed probabilistic formulation could be viewed as a latent-variable augmentation of regression. We devise expectation-maximization (EM) procedures based on a data augmentation strategy which facilitates the maximum-likelihood search over the model parameters. We propose two augmentation schemes and we

describe in detail the associated EM inference procedures that may well be viewed as generalizations of a number of EM regression, dimension reduction, and factor analysis algorithms. The proposed framework is validated with both synthetic and real data. We provide experimental evidence that our method outperforms several existing regression techniques. See [26], [12].

6.4. Simultaneous sound-source separation and localization

Human-robot communication is often faced with the difficult problem of interpreting ambiguous auditory data. For example, the acoustic signals perceived by a humanoid with its on-board microphones contain a mix of sounds such as speech, music, electronic devices, all in the presence of attenuation and reverberations. We proposed a novel method, based on a generative probabilistic model and on active binaural hearing, allowing a robot to robustly perform sound-source separation and localization. We show how interaural spectral cues can be used within a constrained mixture model specifically designed to capture the richness of the data gathered with two microphones mounted onto a human-like artificial head. We describe in detail a novel expectation-maximization (EM) algorithm that alternates between separation and localization, we analyze its initialization, speed of convergence and complexity, and we assess its performance with both simulated and real data. Subsequently, we studied the *binaural manifold*, i.e., the low-dimensional space of sound-source locations embedded in the high-dimensional space of perceived interaural spectral features, and we provided a method for mapping interaural cues onto source locations. See [21], [12]. A journal paper was submitted in 2013 and accepted with minor revisions.

6.5. The geometry of non-coplanar microphone arrays

We addressed the problem of sound-source localization from time-delay estimates using arbitrarily-shaped non-coplanar microphone arrays. A novel geometric formulation is proposed, together with a thorough algebraic analysis and a global optimization solver. The proposed model is thoroughly described and evaluated. The geometric analysis, stemming from the direct acoustic propagation model, leads to necessary and sufficient conditions for a set of time delays to correspond to a unique position in the source space. Such sets of time delays are referred to as *feasible sets*. We formally prove that every feasible set corresponds to exactly one position in the source space, whose value can be recovered using a closed-form localization mapping. Therefore we seek for the optimal feasible set of time delays given, as input, the received microphone signals. This time delay estimation problem is naturally cast into a programming task, constrained by the feasibility conditions derived from the geometric analysis. A global branch-and-bound optimization technique is proposed to solve the problem at hand, hence estimating the best set of feasible time delays and, subsequently, localizing the sound source. Extensive experiments with both simulated and real data are reported; we compare our methodology to four state-of-the-art techniques. This comparison clearly shows that the proposed method combined with the branch-and-bound algorithm outperforms existing methods. These in-depth geometric understanding, practical algorithms, and encouraging results, open several opportunities for future work. See [18], [25], [11].

6.6. Audiovisual calibration and alignment

We addressed the problem of aligning visual (V) and auditory (A) data using a sensor that is composed of a camera-pair and a microphone-pair. The original contribution of the paper is a method for AV data aligning through estimation of the 3D positions of the microphones in the visual-centred coordinate frame defined by the stereo camera-pair. We exploit the fact that these two distinct data sets are conditioned by a common set of parameters, namely the (unknown) 3D trajectory of an AV object, and derive an EM-like algorithm that alternates between the estimation of the microphone-pair position and the estimation of the AV object trajectory. The proposed algorithm has a number of built-in features: it can deal with A and V observations that are misaligned in time, it estimates the reliability of the data, it is robust to outliers in both modalities, and it has proven theoretical convergence. We report experiments with both simulated and real data. See [24] (this work received the best paper award).

6.7. Audiovisual fusion for human-robot interaction

Natural human-robot interaction in complex and unpredictable environments is one of the main research lines in robotics. In typical real-world scenarios, humans are at some distance from the robot and the acquired signals are strongly impaired by noise, reverberations and other interfering sources. In this context, the detection and localisation of speakers plays a key role since it is the pillar on which several tasks (e.g.: speech recognition and speaker tracking) rely. We address the problem of how to detect and localize people that are both seen and heard by a humanoid robot. We introduce a hybrid deterministic/probabilistic model. Indeed, the deterministic component allows us to map the visual information into the auditory space. By means of the probabilistic component, the visual features guide the grouping of the auditory features in order to form AV objects. The proposed model and the associated algorithm are implemented in real-time (17 FPS) using a stereoscopic camera pair and two microphones embedded into the head of the humanoid robot NAO. We performed experiments on (i) synthetic data, (ii) a publicly available data set and (iii) data acquired using the robot. The results we obtained validate the approach and encourage us to further investigate how vision can help robot hearing. See [19], [20], [27], [11], [13]

Prima Project-Team

5. New Results

5.1. Attention-Based Navigation

Participants: Adrian Bourgaud, Carlos Di Pietro, Thierry Fraichard, Rémi Paulin, Patrick Reignier, Andre Van Den Berg.

Assistant robots and robot companions are designed to share the human living space, to navigate among and interact with human beings. From the mobility point of view, roboticists have recently striven to develop navigation scheme geared towards achieving so-called "socially acceptable motions". To that end, various concepts borrowed from environmental psychology and anthropology have been used, the "personal space" concept from Proxemics being perhaps the most widely used. The purpose of our work here is to further the research in this area by taking into account other factors such as human activities, interaction configurations and intentions. An attentional model derived from cognitive psychology is used to dynamically determine the "focus of attention" of the persons involved in a given task. Depending on the task at hand, the robot uses the attention information in order to decide its future course of action so as, for instance, to attract one person's attention or, on the contrary, to minimize the disturbance caused. In 2013, a paper describing the first results obtained was presented during the Israeli Conf. on Robotics [14].

5.2. 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 Plateform datasets.

5.3. Ikio, a sociable kiosk

Participants: Rémi Barraquand [correspondant], Jiří Pytela, Johan Girod.

In the Personal Assisted Living project we investigate the design of iKio: a sociable kiosk. A simple sketch of the iKio is illustrated in figure 6 . The general idea is to enhance the interaction ability of tablet and smartphone. What motivates the choice of this type of devices is the observation that people have come to treat these gadgets as their own body appendage. As pointed out by the recent study conducted by the Pew Research Center [2], people are starting to use their phones and tablets for more sensitive activities that were almost considered taboo in the past, also these devices are becoming substitute for other traditional devices like photo and video cameras.

---

[2]http://pewinternet.org/

The design of iKio is therefore influenced by this emerging form of symbiosis and aims to enhance both user-experience and human-technologies interaction. As follow, iKio does not have a fixed body per se, instead it is embodied in a tablet which can be carried along with people in their daily activities but which can also be docked into any mechanical structure that will provide it with enhanced abilities. Using such mechanical structure iKio can express emotion and interact more easily in the physical space of people. The core of iKio is specifically designed to handle and to support ostensive-inferential communication which is characteristic of human communication in contrast with the code model of communication argued to be the main reason of unadapted and autistic interaction between technologies and human. An early prototype of iKio is illustrated in figure 6 . It was constructed using the Bioloid [3] construction kit.

5.4. Limits and performances of embedded RGBD sensors on mobile robots for social interaction

Participants: Amaury Nègre, Dominique Vaufreydaz [correspondant].

While working on sociably acceptable companion robots, we highlighted some problems of embedding RGBD sensors on mobile robots. Performances of our algorithms can be severely decreased by intrinsic parameters of the robot: linear and angle speeds, height and angle of view of the mounted RGBD sensor, etc. We are currently conducting experiments on influence of these parameters on our perception of humans within a home-like environment. As an extra expected results, we will provide to the research community a corpus that can be used as benchmark for several tasks in mobility: 2D and 3D face detection, body and skeleton detection, fall detection and engagement detection.

---

[3] http://www.robotis.com/xe/bioloid_en
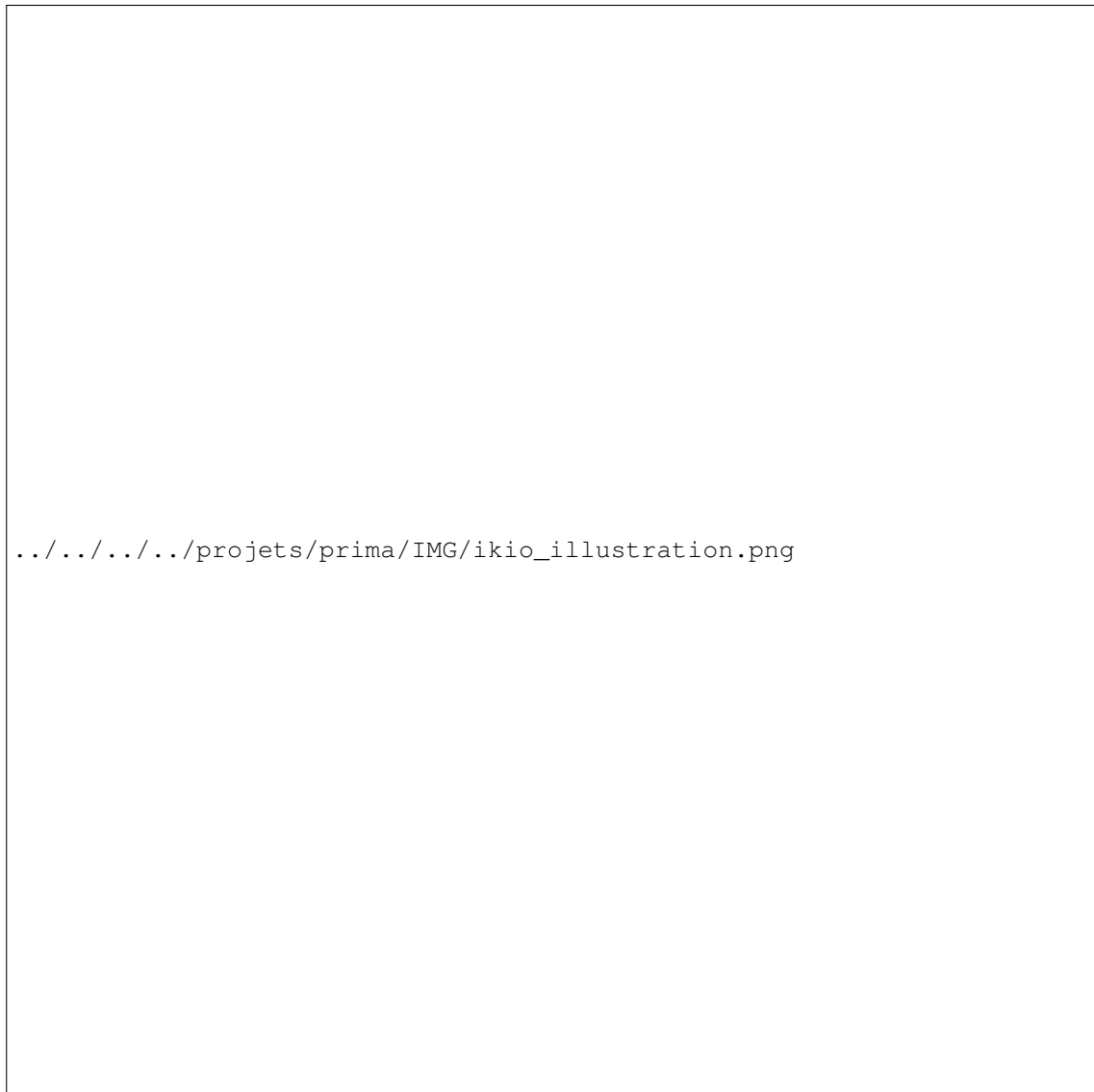
../../../../projets/prima/IMG/ikio_illustration.png

Figure 6. Preliminary sketch of the iKio together with an early prototype, both the 3d model and its realization using the Bioloid Kit

../../../../projets/prima/IMG/RobulabExperiment.png

Figure 7. Robotic platform within the home-like environment for mobile RGBD experiments.

PRIVATICS Team

5. New Results

5.1. Online Social Networks Tracking

Participants: Mohamed Ali Kaafar, Abdelberi Chaabane.

Behavioural advertisement, profiling, adver-gaming and social advertisement illustrate how user personal information and social relations have been integrated to the market model. In other words, user information is *commodified*: the user identity becomes a commodity to be sold and bought. This radical change raises several privacy questions and leads to clamouring for better understanding, regulation and protection of user privacy. Within this context, there is both a long-term and a short-term dimension to our work. For the sort term, I showed that OSN can present a real threat to user privacy as the *full* control of user data – both access and dissemination – is hard to achieve. For the long term, our work calls for a better educational approach to privacy as well as a stricter regulation.

5.2. Behavioural advertisement

Participants: Mohamed Ali Kaafar, Abdelberi Chaabane.

Online Social Networks Tracking. I examined web user tracking capabilities of the three major global OSNs. I studied the mechanisms which enable these services to persistently and accurately follow users web activity, and evaluate to which extent this phenomena is spread across the web. Through a study of the top 10K websites, our findings indicate that OSN tracking is diffused among almost all website categories, independently from the content and the audience. I also evaluated the tracking capabilities in practice and demonstrated – by analysing a real traffic traces – that OSNs can reconstruct a significant portion of users web profile and browsing history. I finally provided insights into the relation between the browsing history characteristics and the OSN tracking potential, highlighting the high risk properties. This work shows that web tracking in combination with personal information from social networks represents a serious privacy violation that shifts the tracking from a virtual tracking (i.e. the user is virtual) to a real "physical" tracking (i.e. based on user personal information).

5.3. Selling Off Privacy at Auction

Participants: Claude Castelluccia, Lukasz Olejnik, Cédric Lauradoux, Minh-Dung Tran.

The first one is a privacy analysis of Real-Time Bidding (RTB) and Cookie Matching (CM). RTB is a technology that allows ad buyers (advertisers) and ad sellers (publishers) to buy and sell ad spaces at real-time auctions through ad exchanges. In RTB, when user visits a publisher page, the ad impression (i.e. one ad display in an ad space) and the user information are immediately broadcast by the ad exchange to a number of bidders (i.e. advertisers or their representatives) for them to bid for the chance to serve ads to this user. CM protocol allows the ad exchange and the bidder to synchronize their cookies of the same user, thus facilitating their exchange of user data.

In [41], we characterize and quantify the potential user web history leakage from ad exchanges to bidders in RTB as a result of exchanging user data. We also discuss and quantify the extent to which companies can potentially collude to increase their tracked user profiles using CM. In addition, we leverage a design characteristic of RTB to observe the winning price of each RTB auction. By analyzing these prices, we show how advertisers evaluate the value of user privacy. This work (titled Selling Off Privacy at Auction) will be presented in NDSS 2014, San Diego, USA in February, 2014.

5.4. Wi-Fi and privacy

Participants: Cédric Lauradoux, Mathieu Cunche, Levent Demir.

Active service discovery in Wi-Fi involves wireless stations broadcasting their Wi-Fi fingerprint, i.e. the SSIDs of their preferred wireless networks. The content of those Wi-Fi fingerprints can reveal different types of information about the owner. In [5], we focus on the relation between the fingerprints and the links between the owners. Our hypothesis is that social links between devices owners can be identified by exploiting the information contained in the fingerprint. More specifically we propose to consider the similarity between fingerprints as a metric, with the underlying idea: similar fingerprints are likely to be linked. We have studied the performances of several similarity metrics on a controlled dataset and then apply the designed classifier to a dataset collected in the wild. Our study is based on a dataset collected in Sydney, Australia, composed of fingerprints belonging to more than 8000 devices.

Extending this problem, we present a set of attacks that allow an attacker to link a Wi-Fi device to its owner identity. We present two methods that, given an individual of interest, allow identifying the MAC address of its Wi-Fi enabled portable device. Those methods do not require a physical access to the device and can be performed remotely, reducing the risks of being noticed. We present in [4], [35] scenarios in which the knowledge of an individual MAC address could be used for mischief.

5.5. Sensor security and privacy

Participants: Claude Castelluccia, Marine Minier, Cédric Lauradoux, Mathieu Cunche.

Wireless sensor networks (WSNs) are composed of a large number of low-cost, low-power, and multi-functional sensor nodes that communicate at short distance through wireless links. They are usually deployed in an open and uncontrolled environment where attackers may be present. Due to the use of low-cost materials, hardware components are not tamper-resistant and an adversary could access to a sensor's internal state.

In [7], we consider packet pollution attack. Packet pollution attack is considered as the most threatening attack model against network coding based sensor networks. A widely held belief says that, in a single source multi-destination dissemination scenario, the total number of polluted packets in the network will grow with the length of the transmission path, and the decoding failure (DF) rate at the further destination nodes are relatively lower. In this work, we first obtain an opposite result by analyzing the pollution attack in multicast scenarios, and find out a convergence trend of pollution attack by network coding system, and quantify the network resiliency against the pollution attacks which happen at any place along the source-destination paths. Then, the analysis result is proved by our simulations on two most widely deployed buffer strategies, Random-In Random-Out (RIRO) and First-in First-Out (FIFO). Finally, it is proved that RIRO has a much advanced security feature than FIFO in constraining the pollution attack gradually, and almost vanished in the end.

An adversary can easily capture even a single node and inserts duplicated nodes at any location in the network. If no specific detection mechanisms are established, the attacker could lead many insidious attacks such as subverting data aggregation protocols by injecting false data, revoking legitimate nodes and disconnecting the network if the replicated nodes are judiciously placed in the network. In [8], we first introduce the algorithm already published in PIMRC 2009 that describes a new hierarchical distributed algorithm for detecting node replication attacks using a Bloom filter mechanism and a cluster head selection. This mechanism could be efficiently used in a WSN as soon as the network is built with a clustering algorithm creating a three tiers hierarchy. We extend the results of our previous results by a theoretical discussion on the bounds of our algorithm. We also perform extensive simulations of our algorithm for random topologies and we compare those results with other proposals of the literature. Finally we show the effectiveness of our algorithm and its energy efficiency.

Finding entropy sources is a major issue to design non-deterministic random generators for headless devices. Our goal in [22] is to evaluate a collection of sensors (e.g. thermometer, accelerometer, magnetometer) as potential sources of entropy. A challenge in the analysis of these sources is the estimation of min-entropy. We have followed the NIST recommendations to obtain pessimistic estimations from the dataset collected during our campaign of experiments. The most interesting sensors of our study are: the accelerometer, the magnetometer, the vibration sensor and the internal clock. Contrary to previous results, we observe far less entropy than it was expected before. Other sensors which measures phenomena with high inertia such as the temperature or air pressure provide very little entropy.

In [12], we propose a key certification protocol for wireless sensor networks that allows nodes to autonomously exchange their public keys and verify their authenticity using one-way accumulators. We examine and compare different accumulator implementations for our pro- tocol on the Sun SPOT platform. We observe that our protocol performs best with accumulators based on Elliptic Curve Cryptography (ECC): ECC-based accumulators have roughly the same speed as Secure Bloom filters, but they have a smaller memory footprint.

5.6. Buidling blocks

Participant: Marine Minier.

In [17], we develop a complete library of lightweight block ciphers dedicated to 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 must be autonomous. We study on a dedicated platform of sensors most of the recent lightweight block ciphers as well as some conventional block ciphers. First, we describe the design of the chosen block ciphers with a security summary and we then present some implementation tests performed on our platform. The library is available online: http://bloc.project.citi-lab.fr/library.html.

In [23], we present two related key impossible differential attacks against 14 rounds of Piccolo-80 and 21 rounds of Piccolo-128 without the whitening layers. Piccolo is a new lightweight block cipher proposed by SONY at CHES 2011. The attack against Piccolo-80 has a time and data complexity of $2^{68.19}$ whereas the time/data complexity of the attack against Piccolo-128 is $2^{117.77}$.

While Generalized Feistel Networks have been widely studied in the literature as a building block of a block cipher, we propose in [13] a unified vision to easily represent them through a matrix representation. We then propose a new class of such schemes called Extended Generalized Feistel Networks well suited for cryptographic applications. We instantiate those proposals into two particular constructions and we finally analyze their security.

We also obtain, in [24] a result concerning an integral distinguisher on the SHA-3 finalist Grøstl-512 v3.

5.7. Formal and legal issues of privacy

Participants: Thibaud Antignac, Denis Butin, Daniel Le Métayer.

- **Privacy by design** The privacy by design approach is often praised by lawyers as well as computer scientists as an essential step towards a better privacy protection. The general philosophy of privacy by design is that privacy should not be treated as an afterthought but rather as a first-class requirement during the design of a system. The approach has been applied in different areas such as smart metering, electronic traffic pricing, ubiquitous computing or location based services. More generally, it is possible to identify a number of core principles that are widely accepted and can form a basis for privacy by design. For example, the Organization for Economic Co-operation and Development (OECD) has put forward principles such as the consent, limitation of use, data quality, security and accountability. One must admit however that the take-up of privacy by design in the industry is still rather limited. This situation is partly due to legal and economic reasons: as long as the law does not impose binding commitments, ICT providers and data collectors do not have sufficient incentives to invest into privacy by design. The situation on the legal side might change in Europe though because the regulation proposed by the European Commission in January 2012 (to replace the European Directive 95/46/EC), which is currently under discussion, includes binding commitments on privacy by design.

  But the reasons for the lack of adoption of privacy by design are not only legal and economic: even though computer scientists have devised a wide range of privacy enhancing tools, no general methodology is available to integrate them in a consistent way to meet a set of privacy requirements. The next challenge in this area is thus to go beyond individual cases and to establish sound foundations and methodologies for privacy by design. As a first step in this direction, we have focused on the data minimization principle which stipulates that the collection should be limited

to the pieces of data strictly necessary for the purpose, and we have proposed a framework to reason about the choices of architecture and their impact in terms of privacy. The first strategic choices are the allocation of the computation tasks to the nodes of the architecture and the types of communications between the nodes. For example, data can be encrypted or hashed, either to protect their confidentiality or to provide guarantees with respect to their correctness or origin. The main benefit of a centralized architecture for the "central" actor is that he can trust the result because he keeps full control over its computation. However, the loss of control by a single actor in decentralized architectures can be offset by extra requirements ensuring that errors (or frauds) can be detected *a posteriori*. In order to help the designer grasp the combination of possible options, our framework provides means to express the parameters to be taken into account (the service to be performed, the actors involved, their respective requirements, etc.) and an inference system to derive properties such as the possibility for an actor to detect potential errors (or frauds) in the computation of a variable. This inference system can be used in the design phase to check if an architecture meets the requirements of the parties or to point out conflicting requirements.

- **Accountability**

  The principle of accountability, which was introduced three decades ago in the OECD guidelines, has been enjoying growing popularity over the last few years as a solution to mitigate the loss of control by increasing transparency of data processing. At the European level, the Article 29 Working Group published an opinion dedicated to the matter two years ago and the principle is expected to be enshrined in the upcoming European data protection regulation. But the term "accountability" is used with different meanings by different actors and the principle itself has been questioned by some authors as providing deceptive protections and also possibly introducing new risks in terms of privacy. We have studied the different interpretations of the notion of accountability following a multidisciplinary approach and we have argued that *strong accountability* should be a cornerstone of future data protection regulations. By *strong accountability* we mean a principle of accountability which

  - applies not only to policies and procedures, but also to practices, thus providing means to oversee the effective processing of the personal data, not only the promises of the data controller and its organisational measures to meet them;
  - is supported by precise binding commitments enshrined in law;
  - involves audits by independent entities.

  Strong accountability should benefit all stakeholders: data subjects, data controllers, and even data protection authorities whose workload should be considerably streamlined.

  But accountability is a requirement to be taken into account from the initial design phase of a system because of its strong impact on the implementation of the log architecture. Using real-world scenarios, we have shown that decisions about log architectures are actually nontrivial. We have addressed the question of what information should be included in logs to make their a posteriori compliance analysis meaningful. We have shown how log content choices and accountability definitions mutually affect each other and incites service providers to rethink up to what extent they can be held responsible. These different aspects are synthesized into guidelines to avoid common pitfalls in accountable log design. This analysis is based on case studies performed on our implementation of the PPL policy language.

- **Verification of privacy properties** The increasing official use of security protocols for electronic voting deepens the need for their trustworthiness, hence for their formal verification. The impossibility of linking a voter to her vote, often called voter privacy or ballot secrecy, is the core property of many such protocols. Most existing work relies on equivalence statements in cryptographic extensions of process calculi. We have proposed the first theorem-proving based verification of voter privacy which overcomes some of the limitations inherent to process calculi-based analysis. Unlinkability between two pieces of information is specified as an extension to the Inductive Method for security protocol verification in Isabelle/HOL. New message operators for association extraction and

synthesis are defined. Proving voter privacy demanded substantial effort and provided novel insights into both electronic voting protocols themselves and the analysed security goals. The central proof elements have been shown to be reusable for different protocols with minimal interaction.

- **Privacy and discrimination**

  The interactions between personal data protection, privacy and protection against discriminations are increasingly numerous and complex. For example, there is no doubt that misuses of personal data can adversely affect privacy and self-development (for example, resulting in the unwanted disclosure of personal data to third parties, in identity theft, or harassment through email or phone calls), or lead to a loss of choices or opportunities (for example, enabling a recruiter to obtain information over the Internet about political opinions or religious beliefs of a candidate and to use this information against him). It could even be suggested that privacy breaches and discriminations based on data processing are probably the two most frequent and the most serious types of consequences of personal data breaches. We have studied these interactions from a multidisciplinary (legal and technical) perspective and argued that an extended application of the application of non-discrimination regulations could help strengthening data protection. We have have analysed and compared personal data protection, privacy and protection against discriminations considering both the types of data concerned and the *modus operandi* (*a priori* versus *a posteriori* controls, actors in charge of the control, etc.). From this comparison, we have drawn some conclusions with respect to their relative effectiveness and argued that *a posteriori* controls on the use of personal data should be strengthened and the victims of data misuse should get compensations which are significant enough to represent a deterrence for data controllers. We have also advocated the establishment of stronger connections between anti-discrimination and data protection laws, in particular to ensure that any data processing leading to unfair differences of treatments between individuals is prohibited and can be effectively punished.

<p style="text-align:center">ROMA Team</p>

6. New Results

6.1. Scheduling tree-shaped task graphs to minimize memory and makespan

In this work [37], we investigate the execution of tree-shaped task graphs using multiple processors. Each edge of such a tree represents a large IO file. A task can only be executed if all input and output files fit into memory, and a file can only be removed from memory after it has been consumed. Such trees arise, for instance, in the multifrontal method of sparse matrix factorization. The maximum amount of memory needed depends on the execution order of the tasks. With one processor the objective of the tree traversal is to minimize the required memory. This problem was well studied and optimal polynomial algorithms were proposed. Here, we extend the problem by considering multiple processors, which is of obvious interest in the application area of matrix factorization. With the multiple processors comes the additional objective to minimize the time needed to traverse the tree, i.e., to minimize the makespan. Not surprisingly, this problem proves to be much harder than the sequential one. We study the computational complexity of this problem and provide an inapproximability result even for unit weight trees. Several heuristics are proposed, each with a different optimization focus, and they are analyzed in an extensive experimental evaluation using realistic trees.

6.2. Model and complexity results for tree traversals on hybrid platforms

In this work [35], we study the complexity of traversing tree-shaped workflows whose tasks require large I/O files. We target a heterogeneous architecture with two resources of different types, where each resource has its own memory, such as a multicore node equipped with a dedicated accelerator (FPGA or GPU). Tasks in the workflow are tagged with the type of resource needed for their processing. Besides, a task can be processed on a given resource only if all its input files and output files can be stored in the corresponding memory. At a given execution step, the amount of data stored in each memory strongly depends upon the ordering in which the tasks are executed, and upon when communications between both memories are scheduled. The objective is to determine an efficient traversal that minimizes the maximum amount of memory of each type needed to traverse the whole tree. In this work, we establish the complexity of this two-memory scheduling problem, provide inapproximability results, and show how to determine the optimal depth-first traversal. Altogether, these results lay the foundations for memory-aware scheduling algorithms on heterogeneous platforms.

6.3. On the combination of silent error detection and checkpointing

In this work [19], we revisit traditional checkpointing and rollback recovery strategies, with a focus on silent data corruption errors. Contrarily to fail-stop failures, such latent errors cannot be detected immediately, and a mechanism to detect them must be provided. We consider two models: (i) errors are detected after some delays following a probability distribution (typically, an Exponential distribution); (ii) errors are detected through some verification mechanism. In both cases, we compute the optimal period in order to minimize the waste, i.e., the fraction of time where nodes do not perform useful computations. In practice, only a fixed number of checkpoints can be kept in memory, and the first model may lead to an irrecoverable failure. In this case, we compute the minimum period required for an acceptable risk. For the second model, there is no risk of irrecoverable failure, owing to the verification mechanism, but the corresponding overhead is included in the waste. Finally, both models are instantiated using realistic scenarios and application/architecture parameters.

6.4. Checkpointing algorithms and fault prediction

In this series of work [22], [49], we deal with the impact of fault prediction techniques on checkpointing strategies, when the fault-prediction system provides either prediction windows or exact predictions. We extend the classical first-order analysis of Young and Daly in the presence of a fault prediction system, characterized by its recall and its precision. In this framework, we provide optimal algorithms to decide whether and when to take predictions into account, and we derive the optimal value of the checkpointing period. These results allow us to analytically assess the key parameters that impact the performance of fault predictors at very large scale.

6.5. Mapping applications on volatile resources

In this series of work [12], [27], [28], we study the execution of iterative applications on volatile processors such as those found on desktop grids. We envision two models, one where all tasks are assumed to be independent, and another where all tasks are tightly coupled and keep exchanging information throughout the iteration. These two models cover the two extreme points of the parallelization spectrum. We develop master-worker scheduling schemes that attempt to achieve good trade-offs between worker speed and worker availability. Any iteration entails the execution of a fixed number of independent tasks or of tightly-coupled tasks. A key feature of our approach is that we consider a communication model where the bandwidth capacity of the master for sending application data to workers is limited. This limitation makes the scheduling problem more difficult both in a theoretical sense and in a practical sense. Furthermore, we consider that a processor can be in one of three states: available, down, or temporarily preempted by its owner. This preempted state also complicates the scheduling problem. In practical settings, e.g., desktop grids, master bandwidth is limited and processors are temporarily reclaimed. Consequently, addressing the aforementioned difficulties is necessary for successfully deploying master-worker applications on volatile platforms. Our first contribution is to determine the complexity of the scheduling problems in their offline versions, i.e., when processor availability behaviors are known in advance. Even with this knowledge, the problems are NP-hard. Our second contribution is an evaluation of the expectation of the time needed by a worker to complete a set of tasks. We obtain a close formula for independent tasks and an analytical approximation for tightly-coupled tasks. Those evaluations rely on a Markovian assumption for the temporal availability of processors, and are at the heart of some heuristics that aim at favoring "reliable" processors in a sensible manner. Our third contribution is a set of heuristics for both models, which we evaluate in simulation. Our results provide guidance to selecting the best strategy as a function of processor state availability versus average task duration.

6.6. Using group replication for resilience on exascale systems

High performance computing applications must be resilient to faults. The traditional fault-tolerance solution is checkpoint-recovery, by which application state is saved to and recovered from secondary storage throughout execution. It has been shown that, even when using an optimal checkpointing strategy, the checkpointing overhead precludes high parallel efficiency at large scale. Additional fault-tolerance mechanisms must thus be used. Such a mechanism is replication, i.e., multiple processors performing the same computation so that a processor failure does not necessarily imply an application failure. In spite of resource waste, replication can lead to higher parallel efficiency when compared to using only checkpoint-recovery at large scale. In this work [11], we propose to execute and checkpoint multiple application instances concurrently, an approach we term group replication. For Exponential failures we give an upper bound on the expected application execution time. This bound corresponds to a particular checkpointing period that we derive. For general failures, we propose a dynamic programming algorithm to determine non-periodic checkpoint dates as well as an empirical periodic checkpointing solution whose period is found via a numerical search. Using simulation we evaluate our proposed approaches, including comparison to the non-replication case, for both Exponential and Weibull failure distributions. Our broad finding is that group replication is useful in a range of realistic application and checkpointing overhead scenarios for future exascale platforms.

6.7. Unified model for assessing checkpointing protocols at extreme-scale

In this work [10], we present a unified model for several well-known checkpoint/restart protocols. The proposed model is generic enough to encompass both extremes of the checkpoint/restart space, from coordinated approaches to a variety of uncoordinated checkpoint strategies (with message logging). We identify a set of crucial parameters, instantiate them and compare the expected efficiency of the fault tolerant protocols, for a given application/platform pair. We then propose a detailed analysis of several scenarios, including some of the most powerful currently available HPC platforms, as well as anticipated Exascale designs. The results of this analytical comparison are corroborated by a comprehensive set of simulations. Altogether, they outline comparative behaviors of checkpoint strategies at very large scale, thereby providing insight that is hardly accessible to direct experimentation.

6.8. Revisiting the double checkpointing algorithm

In this work [33], we study fast checkpointing algorithms which require distributed access to stable storage. This work revisits the approach base upon double checkpointing, and compares the blocking algorithm of Zheng, Shi, and Kalé, with the non-blocking algorithm of Ni, Meneses, and Kalé in terms of both performance and risk. We also extend the model that they have proposed to assess the impact of the overhead associated to non-blocking communications. We then provide a new peer-to-peer checkpointing algorithm, called the triple checkpointing algorithm, that can work at constant memory, and achieves both higher efficiency and better risk handling than the double checkpointing algorithm. We provide performance and risk models for all the evaluated protocols, and compare them through comprehensive simulations.

6.9. Multi-criteria checkpointing strategies: Optimizing response-time versus resource utilization

Failures are increasingly threatening the efficiency of HPC systems, and current projections of Exascale platforms indicate that rollback recovery, the most convenient method for providing fault tolerance to general-purpose applications, reaches its own limits at such scales. One of the reasons explaining this unnerving situation comes from the focus that has been given to per-application completion time, rather than to platform efficiency. In this work [26], we discuss the case of uncoordinated rollback recovery where the idle time spent waiting recovering processors is used to progress a different, independent application from the system batch queue. We then propose an extended model of uncoordinated checkpointing that can discriminate between idle time and wasted computation. We instantiate this model in a simulator to demonstrate that, with this strategy, uncoordinated checkpointing per application completion time is unchanged, while it delivers near-perfect platform efficiency.

6.10. Optimal checkpointing period: Time vs. energy

In this work [18], we deal with parallel scientific applications using non-blocking and periodic coordinated checkpointing to enforce resilience. We provide a model and detailed formulas for total execution time and consumed energy. We characterize the optimal period for both objectives, and we assess the range of time/energy trade-offs to be made by instantiating the model with a set of realistic scenarios for Exascale systems. We give a particular emphasis to I/O transfers, because the relative cost of communication is expected to dramatically increase, both in terms of latency and consumed energy, for future Exascale platforms.

6.11. Energy-aware checkpointing of divisible tasks with soft or hard deadlines

In this work [20], we aim at minimizing the energy consumption when executing a divisible workload under a bound on the total execution time, while resilience is provided through checkpointing. We discuss several variants of this multi-criteria problem. Given the workload, we need to decide how many chunks to use, what are the sizes of these chunks, and at which speed each chunk is executed. Furthermore, since a failure may occur during the execution of a chunk, we also need to decide at which speed a chunk should be re-executed in

the event of a failure. The goal is to minimize the expectation of the total energy consumption, while enforcing a deadline on the execution time, that should be met either in expectation (soft deadline), or in the worst case (hard deadline). For each problem instance, we propose either an exact solution, or a function that can be optimized numerically. The different models are then compared through an extensive set of experiments.

6.12. Assessing the performance of energy-aware mappings

In this work [8], we aim at mapping streaming applications that can be modeled by a series-parallel graph onto a 2-dimensional tiled chip multiprocessor (CMP) architecture. The objective of the mapping is to minimize the energy consumption, using dynamic voltage and frequency scaling (DVFS) techniques, while maintaining a given level of performance, reflected by the rate of processing the data streams. This mapping problem turns out to be NP-hard, and several heuristics are proposed. We assess their performance through comprehensive simulations using the StreamIt workflow suite and randomly generated series-parallel graphs, and various CMP grid sizes.

6.13. Computing the throughput of probabilistic and replicated streaming applications

In this work [7], we investigate how to compute the throughput of probabilistic and replicated streaming applications. We are given (i) a streaming application whose dependence graph is a linear chain; (ii) a one-to-many mapping of the application onto a fully heterogeneous target platform, where a processor is assigned at most one application stage, but where a stage can be replicated onto a set of processors; and (iii) a set of random variables modeling the computation and communication times in the mapping. We show how to compute the throughput of the application, i.e., the rate at which data sets can be processed, under two execution models, the Strict model where the actions of each processor are sequentialized, and the Overlap model where a processor can compute and communicate in parallel. The problem is easy when application stages are not replicated, i.e., assigned to a single processor: in that case the throughput is dictated by the critical hardware resource. However, when stages are replicated, i.e., assigned to several processors, the problem becomes surprisingly complicated: even in the deterministic case, the optimal throughput may be lower than the smallest internal resource throughput. The first contribution of this work is to provide a general method to compute the throughput when mapping parameters are constant or follow I.I.D. exponential laws. The second contribution is to provide bounds for the throughput when stage parameters (computation and communication times) form associated random sequences, and are N.B.U.E. (New Better than Used in Expectation) variables: the throughput is bounded from below by the exponential case and bounded from above by the deterministic case. An extensive set of simulation allows us to assess the quality of the model, and to observe the actual behavior of several distributions.

6.14. Reliability and performance optimization of pipelined real-time systems

In this work [6], we consider pipelined real-time systems that consist of a chain of tasks executing on a distributed platform. The processing of the tasks is pipelined: each processor executes only one interval of consecutive tasks. We are interested in minimizing both the input-output latency and the period of application mapping. For dependability reasons, we are also interested in maximizing the reliability of the system. We therefore assign several processors to each interval of tasks, so as to increase the reliability of the system. Both processors and communication links are unreliable and subject to transient failures. We assume that the arrival of the failures follows a constant parameter Poisson law, and that the failures are statistically independent events. We study several variants of this multiprocessor mapping problem, with several hypotheses on the target platform (homogeneous/heterogeneous speeds and/or failure rates). We provide NP-hardness complexity results, and optimal mapping algorithms for polynomial problem instances. Efficient heuristics are presented to solve the general case, and experimental results are provided.

6.15. Scheduling linear chain streaming applications on heterogeneous systems with failures

In this work [5], we study the problem of optimizing the throughput of streaming applications for heterogeneous platforms subject to failures. Applications are linear graphs of tasks (pipelines), with a type associated to each task. The challenge is to map each task onto one machine of a target platform, each machine having to be specialized to process only one task type, given that every machine is able to process all the types before being specialized in order to avoid costly setups. The objective is to maximize the throughput, i.e., the rate at which jobs can be processed when accounting for failures. Each instance can thus be performed by any machine specialized in its type and the workload of the system can be shared among a set of specialized machines. For identical machines, we prove that an optimal solution can be computed in polynomial time. However, the problem becomes NP-hard when two machines may compute the same task type at different speeds. Several polynomial time heuristics are designed for the most realistic specialized settings. Simulation results assess their efficiency, showing that the best heuristics obtain a good throughput, much better than the throughput obtained with a random mapping. Moreover, the throughput is close to the optimal solution in the particular cases where the optimal throughput can be computed.

6.16. A survey of pipelined workflow scheduling: Models and algorithms

In this survey [4], we consider a large class of applications that need to execute the same workflow on different data sets of identical size. Efficient execution of such applications necessitates intelligent distribution of the application components and tasks on a parallel machine, and the execution can be orchestrated by utilizing task-, data-, pipelined-, and/or replicated-parallelism. The scheduling problem that encompasses all of these techniques is called pipelined workflow scheduling, and it has been widely studied in the last decade. Multiple models and algorithms have flourished to tackle various programming paradigms, constraints, machine behaviors or optimization goals. This work surveys the field by summing up and structuring known results and approaches.

6.17. Reclaiming the energy of a schedule: Models and algorithms

In this work [1], we consider a task graph to be executed on a set of processors. We assume that the mapping is given, say by an ordered list of tasks to execute on each processor, and we aim at optimizing the energy consumption while enforcing a prescribed bound on the execution time. Although it is not possible to change the allocation of a task, it is possible to change its execution speed. Rather than using a local approach such as backfilling, we consider the problem as a whole and study the impact of several speed variation models on its complexity. For continuous speeds, we give a closed-form formula for trees and series-parallel graphs, and we cast the problem into a geometric programming problem for general directed acyclic graphs. We show that the classical dynamic voltage and frequency scaling (DVFS) model with discrete modes leads to an NP-complete problem, even if the modes are regularly distributed (an important particular case in practice, which we analyze as the incremental model). On the contrary, the Vdd-hopping model that allows to switch between different supply voltages (VDD) while executing a task leads to a polynomial solution. Finally, we provide an approximation algorithm for the incremental model, which we extend for the general DVFS model.

6.18. Non-clairvoyant reduction algorithms for heterogeneous platforms

In this work [24], we revisit the classical problem of the reduction collective operation in a heterogeneous environment. We discuss and evaluate four algorithms that are non-clairvoyant, i.e., they do not know in advance the computation and communication costs. On the one hand, Binomial-stat and Fibonacci-stat are static algorithms that decide in advance which operations will be reduced, without adapting to the environment; they were originally defined for homogeneous settings. On the other hand, Tree-dyn and Non-Commut-Tree-dyn are fully dynamic algorithms, for commutative or non-commutative reductions. With identical computation costs, we show that these algorithms are approximation algorithms with constant or asymptotic ratios. When costs are exponentially distributed, we perform an analysis of Tree-dyn based on Markov chains.

Finally, we assess the relative performance of all four non-clairvoyant algorithms with heterogeneous costs through a set of simulations.

6.19. Non-linear divisible loads: There is no free lunch

Divisible Load Theory (DLT) has received a lot of attention in the past decade. A divisible load is a perfect parallel task, that can be split arbitrarily and executed in parallel on a set of possibly heterogeneous resources. The success of DLT is strongly related to the existence of many optimal resource allocation and scheduling algorithms, what strongly differs from general scheduling theory. Moreover, recently, close relationships have been underlined between DLT, that provides a fruitful theoretical framework for scheduling jobs on heterogeneous platforms, and MapReduce, that provides a simple and efficient programming framework to deploy applications on large scale distributed platforms.

The success of both have suggested to extend their framework to non-linear complexity tasks. In this work [23], we show that both DLT and MapReduce are better suited to workloads with linear complexity. In particular, we prove that divisible load theory cannot directly be applied to quadratic workloads, such as it has been proposed recently. We precisely state the limits for classical DLT studies and we review and propose solutions based on a careful preparation of the dataset and clever data partitioning algorithms. In particular, through simulations, we show the possible impact of this approach on the volume of communications generated by MapReduce, in the context of Matrix Multiplication and Outer Product algorithms.

6.20. Direct solvers for sparse linear systems

This work is closely related to the MUMPS solver (see Section 5.1) and was performed in close collaboration with INPT (Toulouse). First, we have pursued the study of low-rank representations to speed-up sparse direct solvers using the so called BLR (Block Low Rank) format [44]. This work was done in collaboration with LSTC (Livermore Software Technology Corp., USA) and in the context of a contract with EDF which funded the PhD thesis of Clément Weisbecker at INPT. We also worked on shared-memory parallelism [61] in the context of the PhD thesis of Wissam M. Sid-Lakhdar. Concerning low-rank approximations, they were experimented on geophysics applications [38] (Helmholtz equations) in the context of a collaboration with members of the ISTerre and Geoazur laboratories. The impact of both low-rank compression and shared-memory parallelism was also studied on electromagnetism problems [17], in collaboration with University of Padova (Italy) and CEDRAT.

We have started the design and implementation of a distributed-memory low-rank multifrontal solver. When computations are faster (thanks to low-rank compression or multithreading within each node), we observed that communications become critical; we are therefore currently studying the limits of the communication schemes from the MUMPS approach and their possible improvements.

On numerical and industrial aspects, we worked on rank detection and null space basis computations (in collaboration with CERFACS and Total/Hutchinson) as well as on improved parallel pivoting strategies for symmetric indefinite systems, in collaboration with ESI-Group (see Section 7.1).

6.21. Push-relabel based algorithms for the maximum transversal problem

In this work [14], we investigate the push-relabel algorithm for solving the problem of finding a maximum cardinality matching in a bipartite graph in the context of the maximum transversal problem. We describe in detail an optimized yet easy-to-implement version of the algorithm and fine-tune its parameters. We also introduce new performance-enhancing techniques. On a wide range of real-world instances, we compare the push-relabel algorithm with state-of-the-art algorithms based on augmenting paths and pseudoflows. We conclude that a carefully tuned push-relabel algorithm is competitive with all known augmenting path-based algorithms, and superior to the pseudoflow-based ones.

6.22. Constructing elimination trees for sparse unsymmetric matrices

The elimination tree model for sparse unsymmetric matrices and an algorithm for constructing it have been recently proposed [82], [83]. The construction algorithm has a worst-case time complexity of $\Theta(mn)$ for an $n \times n$ unsymmetric matrix having m off-diagonal nonzeros. In this work [15], we propose another algorithm that has a worst-case time complexity of $\mathcal{O}(m \log n)$. We compare the two algorithms experimentally and show that both algorithms are efficient in general. The algorithm of Eisenstat and Liu is faster in many practical cases, yet there are instances in which there is a significant difference between the running time of the two algorithms in favor of the proposed one.

6.23. Semi-matching algorithms for scheduling parallel tasks under resource constraints

In this work [25], we study the problem of minimum makespan scheduling when tasks are restricted to subsets of the processors (resource constraints), and require either one or multiple distinct processors to be executed (parallel tasks). This problem is related to the minimum makespan scheduling problem on unrelated machines, as well as to the concurrent job shop problem, and it amounts to finding a semi-matching in bipartite graphs or hypergraphs. The problem is known to be NP-complete for bipartite graphs with general vertex (task) weights, and solvable in polynomial time for unweighted graphs with unit weights (i.e., unit-weight tasks). We prove that the problem is NP-complete for hypergraphs even in the unweighted case. We design several greedy algorithms of low complexity to solve two versions of the problem, and assess their performance through a set of exhaustive simulations. Even though there is no approximation guarantee for these low-complexity algorithms, they return solutions close to the optimal (or a known lower bound) in average.

6.24. Maximum cardinality bipartite matching algorithms on GPUs

In two studies [30], [31], we propose, develop, and evaluate maximum cardinality matching algorithms from two different families (called push-relabel and augmenting-path based) on GPUs. The problem of finding a maximum cardinality matching in bipartite graphs has applications in computer science, scientific computing, bioinformatics, and other areas. To the best of our knowledge, the proposed algorithms are the first investigation of the push-relabel and augmenting-path based on GPUs/ We compare the proposed algorithms with serial and multicore implementations from the literature on a large set of real-life problems where in majority of the cases one of our GPU-accelerated algorithms is demonstrated to be faster than both the sequential and multicore implementations.

6.25. Analysis of partitioning models and metrics in parallel sparse matrix-vector multiplication

Graph/hypergraph partitioning models and methods have been successfully used to minimize the communication among processors in several parallel computing applications. Parallel sparse matrix-vector multiplication (SpMxV) is one of the representative applications that renders these models and methods indispensable in many scientific computing contexts. In this work [36], [55], we investigate the interplay of the partitioning metrics and execution times of SpMxV implementations in three libraries: Trilinos, PETSc, and an in-house one. We carry out experiments with up to 512 processors and investigate the results with regression analysis. Our experiments show that the partitioning metrics influence the performance greatly in a distributed memory setting. The regression analyses demonstrate which metric is the most influential for the execution time of the libraries.

6.26. On partitioning and reordering problems in a hierarchically parallel hybrid linear solver

PDSLin is a general-purpose algebraic parallel hybrid (direct/iterative) linear solver based on the Schur complement method. The most challenging step of the solver is the computation of a preconditioner based

on the global Schur complement. Efficient parallel computation of the preconditioner gives rise to partitioning problems with sophisticated constraints and objectives. In this work [39], we identify two such problems and propose hypergraph partitioning methods to address them. The first problem is to balance the workloads associated with different subdomains to compute the preconditioner. We first formulate an objective function and a set of constraints to model the preconditioner computation time. Then, to address these complex constraints, we propose a recursive hypergraph bisection method. The second problem is to improve the data locality during the parallel solution of a sparse triangular system with multiple sparse right-hand sides. We carefully analyze the objective function and show that it can be well approximated by a standard hypergraph partitioning method. Moreover, an ordering compatible with a post ordering of the subdomain elimination tree is shown to be very effective in preserving locality. To evaluate the two proposed methods in practice, we present experimental results using linear systems arising from some applications of our interest. First, we show that in comparison to a commonly-used nested graph dissection method, the proposed recursive hypergraph partitioning method reduces the preconditioner construction time, especially when the number of subdomains is moderate. This is the desired result since PDSLin is based on a two-level parallelization to keep the number of subdomains small by assigning multiple processors to each subdomain. We also show that our second proposed hypergraph method improves the data locality during the sparse triangular solution and reduces the solution time. Moreover, we show that partitioning time can be greatly reduced while maintaining its quality by removing quasi-dense rows from the solution vectors.

6.27. UMPA: A Multi-objective, multi-level partitioner for communication minimization

In this work [42], we propose a directed hypergraph model and a refinement heuristic to distribute communicating tasks among the processing units in a distributed memory setting. The aim is to achieve load balance and minimize the maximum data sent by a processing unit. We also take two other communication metrics into account with a tie-breaking scheme. With this approach, task distributions causing an excessive use of network or a bottleneck processor which participates to almost all of the communication are avoided. We show on a large number of problem instances that our model improves the maximum data sent by a processor up to 34% for parallel environments with 4, 16, 64, and 256 processing units compared to the state of the art which only minimizes the total communication volume.

6.28. A Partitioning-based divisive clustering technique for maximizing the modularity

In this work [43], we present a new graph clustering algorithm aimed at obtaining clusterings of high modularity. The algorithm pursues a divisive clustering approach and uses established graph partitioning algorithms and techniques to compute recursive bipartitions of the input as well as to refine clusters. Experimental evaluation shows that the modularity scores obtained compare favorably to many previous approaches. In the majority of test cases, the algorithm outperformed the best known alternatives. In particular, among 13 problem instances common in the literature, the proposed algorithm improves the best known modularity in 9 cases.

6.29. Randomized matching heuristics with quality guarantees on shared memory parallel computers

In this work [56], we propose two heuristics for the bipartite matching problem that are amenable to shared-memory parallelization. The first heuristic is very intriguing from parallelization perspective. It has no significant algorithmic synchronization overhead and no conflict resolution is needed across threads. We show that this heuristic has an approximation ratio of around 0.632. The second heuristic is designed to obtain a larger matching by employing the well-known Karp-Sipser heuristic on a judiciously chosen subgraph of the original graph. We show 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 exploit the

structure of the selected subgraphs to propose a specialized implementation which demonstrates a very good scalability. Based on our experiments and theoretical evidence, we conjecture that this second heuristic obtains matchings with cardinality of at least 0.866 of the maximum cardinality. We discuss parallel implementations of the proposed heuristics on shared memory systems. Experimental results, for demonstrating speed-ups and verifying the theoretical results in practice, are provided.

6.30. On the minimum edge cover and vertex partition by quasi-cliques problems

A γ-quasi-clique in a simple undirected graph is a set of vertices which induces a subgraph with the edge density of at least γ for $0 < \gamma < 1$. A cover of a graph by γ-quasi-cliques is a set of γ-quasi-cliques where each edge of the graph is contained in at least one quasi-clique. The minimum cover by γ-quasi-cliques problem asks for a γ-quasi-clique cover with the minimum number of quasi-cliques. A partition of a graph by γ-quasi-cliques is a set of γ-quasi-cliques where each vertex of the graph belongs to exactly one quasi-clique. The minimum partition by γ-quasi-cliques problem asks for a vertex partition by γ-quasi-cliques with the minimum number of quasi-cliques. In this work [60], we show that the decision versions of the minimum cover and partition by γ-quasi-cliques problems are NP-complete for any fixed γ satisfying $0 < \gamma < 1$.

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

6. New Results

6.1. Flexible Radio Front-End

The contributions on hardware design are twofold. First, the development of a Full-Duplex architecture for OFDM systems. Second, a proposal of a Wake-Up scheme for home networking with reduced power consumption.

6.1.1. *Full-Duplex systems*

Zhan et al. [23] focused on the study of active analog self-interference cancellation (AASIC) techniques in full-duplex OFDM systems. This original approach aims at proposing a cancellation technique at RF level for wideband systems. A theoretical study confronted to simulations was proposed with a particular emphasis on the channel estimation of the interfering signal. This study was completed with an analysis on the phase noise and the thermal noise impact.

6.1.2. *Wake-Up Architectures*

Khoumeri et al. [28] proposed radio architectures for allowing energy savings by letting devices to switch off part of the transmission components when they are not in use. Based on classical WiFi systems, the proposed architecture offers the ability to use a conventional emitter, using only a particular subcarrier fingerprint to identify the node to wake-up, hence avoiding a high level of false wake-up.

6.2. Agile Radio Resource Sharing

The contributions of the axis in *agile radio resource sharing* can be gathered in three groups: (a) green communications; (b) performance analysis; and (c) scheduling and power allocation techniques.

6.2.1. *Green Communications*

The main contributions in the subject of green communications focus on the problem of increasing the energy efficiency of Orthogonal Frequency-Division Multiple Access (OFDMA) wireless networks. In particular, Tsilimantos et al. in [21] and Hasan et al. in [15] studied different techniques to strategically switch off some of the base stations in cellular systems while guaranteeing a given quality of service (QoS). In [21], the authors use methods from stochastic geometry to determine the number of active cells that can be switched off while the outage probability, or equivalently the signal to interference plus noise ratio (SINR), remains the same. In [15], this problem is studied from a decentralized point of view using methods from coalitional game theory.

6.2.2. *Performance Analysis*

The contributions in performance analysis are mainly oriented to the field of body area networks (BANs), [17] indoor adaptive OFDM wireless networks and relay channels.

In [17], Lauzier et al. presented the results of a measurement campaign whose primary objective was to characterize the complete mesh of a BAN and analyze the quality of every radio link between the different nodes. In [18], [19], the Multi-Resolution Frequency Domain ParFlow (MRFDPF) model is used to calculate the bit error rate (BER) and study the feasibility of adaptive modulation in OFDM systems.

In the context of relay channels, Ferrand et al. [32] studied the asymptotic *coding gain* of the packet error rate of relay channels in which the radio links are subject to both fading and log-normal shadowing effects simultaneously.

6.2.3. Scheduling and Power Allocation Techniques

Power allocation techniques and scheduling were studied by Ferrand et al. [33] and Wang et al. [9]. More specifically, advances in the study of the achievable rate region of relay channels in the case of global power constraints were reported in [33]. Cooperative scheduling techniques in the context of BAN were proposed in [9] to reduce inter-BAN interference using tools from game theory.

6.3. Software Radio Programming Model

Software defined radio (SDR) technology has evolved rapidly and is now reaching market maturity. Still, a lot of issues have yet to be studied. Mickaël Dardaillon, Kevin Marquet, Tanguy Risset and others highlighted the constraints imposed by recent radio protocols and presented current architectures, solutions, and challenges for programming SDR [31].

6.3.1. Dataflow programming

To enable dynamic adaptation of computation intensive multimedia dataflow applications, Lionel Morel, Kevin Marquet and others have studied language extensions, together with the corresponding run-time support. They show that this approach can be used to monitor and control throughput [20] and offer quality of service [29], with a low impact on the overall performance.

6.3.2. Energy-efficient Localization

Guillaume Salagnac and others address the tradeoff between energy consumption and localization performance in a mobile sensor network application [7]. The focus is on augmenting GPS location with more energy-efficient location sensors to bound position estimate uncertainty while GPS is off. Such combined strategies can cut node energy consumption by one third while still meeting application-specific positioning criteria.

6.3.3. Swap Fairness for Thrashing Mitigation

In the context of shared hosting or virtualization, where multiple users run uncoordinated and selfish workloads, François Goichon, Guillaume Salagnac and Stéphane Frénot introduced an accounting layer that forces swap fairness among processes competing for main memory [13]. It ensures that a process cannot monopolize the swap subsystem by delaying the swap operations of abusive processes, reducing the number of system-wide page faults while maximizing memory utilization.

6. New Results

6.1. Components and Contracts

Participants: Gregor Goessler, Quentin Sabah, Jean-Bernard Stefani.

6.1.1. Analysis of logical causality

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". Surprisingly, the study of logical causality has so far received little attention in computer science, with the notable exception of [69] and its instantiations. However, this approach relies on a causal model that may not be known, for instance in presence of black-box components.

Improving on previous results, we have proposed in [21] an approach to enhance the fault diagnosis in black-box component-based systems, in which only events on component interfaces are observable. For such systems, we have described a causality analysis framework 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, and applicable to real-time systems. We have illustrated the analysis with a case study from the medical device domain.

In [5] we have proposed a formal framework for reasoning about causality, and blaming system-level failures on the component(s) that caused them. The framework is general in the sense that it applies to many different models of computation and communication (MoC), such as synchronous and asynchronous computation, and communication by messages or shared variables. We are currently instantiating the framework to specific MoC, in particular, to timed automata, and developing a refinement of our original approach that reduces the number of false positives.

6.1.2. Supporting isolation for actors in shared memory

The actor model of concurrency, as supported *e.g.*, by the Erlang programming language, is an appealing programming model for the construction of concurrent and distributed systems, and multicore programming in particular. Although much work has taken place in particular during the past ten years on efficient implementations of the actor model, the design space is far from being completely understood.

As part of Quentin Sabah's thesis [10], we have developed a variant of the actor model that, in contrast to previous works, ensures a strict isolation between actors while imposing no restriction on the form of data exchanged in messages. We have formally specified an abstract machine, called SIAAM (see Sec.5.4.5), for an extension of the Java language with our actor model, and implemented it as a modified Jikes virtual machine, a state of the art Java virtual machine. A combination of points-to and live variable analyses has been implemented using the Soot framework, that can be used to remove unnecessary read and write checks for isolation. A diagnosis tool built on top of the analyses helps programmers to pinpoint potential problems (exceptions raised indicating a potential violation of isolation). We have shown with artificial and small applicative benchmarks that, using our analyses to improve performance, our implementation is reasonably efficient and imposes low overhead for the benefit of strict isolation.

In addition, we have developed a Coq proof of the isolation property enforced by SIAAM, namely that no information between actors can take place outside of message exchanges, despite the presence of a shared heap between actors.

6.2. Real-Time multicore programming

Participants: Vagelis Bebelis, Gwenaël Delaval, Pascal Fradet, Alain Girault, Gregor Goessler, Bertrand Jeannet, Gideon Smeding, 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 [57]. 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 [90]. 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 [39].

Following our past results on the PRET-C programming language [35], 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 [22] (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.

This work has been conducted within the RIPPES associated team.

6.2.2. *WCET analysis*

Our past work on the WCET analysis of PRET-C programs has led us to design static analyses, for instance to prune unfeasible paths in the control flow graph [36]. In 2013, we have worked on how to take into account direct mapped instruction caches in WCET analysis. Instruction caches are essential to address if one wants to analyze large embedded programs. Our cache analysis technique offers the same precision as the most precise techniques [80], while improving analysis time by up to 240 times. This improvement is achieved by analyzing individual blocks of the control flow graph separately, and by proposing a tailored abstract domain to represent efficiently the cache state [14], [25]. In contrast with previous abstract analysis methods [88], [85], our analysis is able to offer the same precision as the concrete approaches [80].

6.2.3. *Tradeoff exploration between reliability, power consumption, and execution time*

For autonomous critical real-time embedded systems (*e.g.*, satellites), guaranteeing a very high level of reliability is as important as keeping the power consumption as low as possible. We have designed an off-line ready list scheduling heuristics which, from a given software application graph and a given multiprocessor architecture (homogeneous and fully connected), produces a static multiprocessor schedule that optimizes three criteria: its *length* (crucial for real-time systems), its *reliability* (crucial for dependable systems), and its *power consumption* (crucial for autonomous systems). Our tri-criteria scheduling heuristics, *TSH*, uses the *active replication* of the operations and the data-dependencies to increase the reliability, and uses *dynamic voltage and frequency scaling* to lower the power consumption [37], [38]. TSH implements a ready list scheduling heuristics, and we have formulated a new multi-criteria cost function such that we are able to prove rigorously that the static schedules we generate meet both the reliability constraint and the power consumption constraint [12].

By running TSH on a single problem instance, we are able to provide the Pareto front for this instance in 3D, therefore exposing the user to several tradeoffs between the power consumption, the reliability and the execution time. Thanks to extensive simulation results, we have shown how TSH behaves in practice. Firstly, we have compared TSH versus an optimal Mixed Linear Integer Program on small instances; the experimental results show that TSH behaves very well compared to the ILP. Secondly, we have compared TSH with the ECS heuristic (Energy-Conscious Scheduling [77]); the experimental results show that TSH performs systematically better than ECS.

This is a joint work with Ismail Assayad (U. Casablanca, Morocco) and Hamoudi Kalla (U. Batna, Algeria), who both visit the team regularly.

6.2.4. *Modular distribution*

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 Xavier Nicollin (Grenoble INP, VERIMAG lab).

We are working on type systems to formalize, in a uniform way, both the clock calculus and the location calculus of a synchronous data-flow programming language (the HEPTAGON language, inspired from LUCID SYNCHRONE [49]). On one hand, the clock calculus infers the clock of each variable in the program and checks the clock consistency: *e.g.*, a time-homogeneous function, like +, should be applied to variables with identical clocks. On the other hand, the location calculus infers the spatial distribution of computations and checks the spatial consistency: *e.g.*, a centralized operator, like +, should be applied to variables located at the same location. Compared to the PhD of Gwenaël Delaval [55], [56], the goal is to achieve *modular* distribution. 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. We make use of our uniform type system to express the computing locations as first-class abstract types, exactly like clocks. It allows us to compile a typed variable (typed by both the clock and the location calculi) into `if ... then ... else ...` structures, whose conditions will be valuations of the clock and location variables.

We currently work on an example of software-defined radio. We have shown on this example how to use a modified clock calculus to describe the localisation of values as clocks, and the architecture as clocks (for the computing resources) and their relations (for communication links).

6.2.5. *Distribution of synchronous programs under real-time constraints*

The goal of Gideon Smeding's PhD thesis [11] was to propose a quasi-synchronous framework encompassing constraints on the relative speed of clocks, together with a formalism for reasoning about clock-dependent properties within the model. This framework should provide a seamless link between synchronous models and their asynchronous implementation.

The quasi-synchronous approach developed in [11] considers independently clocked, synchronous components that interact via communication-by-sampling or FIFO channels. We have defined relative drift bounds on pairs of recurring events such as clock ticks or the arrival of a message. Drift bounds express constraints on the stability of clocks, *e.g.*, at least two ticks of one per three consecutive ticks of the other. We can thus move from total synchrony, where all clocks tick simultaneously, to global asynchrony by relaxing the drift bounds. As constraints are more relaxed, behavior diverges more and more from synchronous system behavior. In many systems, such as distributed control systems, occasional deviations of input and output signals of the controller from their behavior in the synchronous model may be acceptable as long as the frequency of such deviations is bounded. The approach of [11] takes as inputs a program written in a Lustre-like language extended with asynchronous communication by sampling, application requirements on the distribution in the form of weakly-hard constraints [45] bounding *e.g.*, the tolerated loss of data tokens, and platform assertions (*e.g.*, relative clock speeds, available communication resources), and verifies whether the program meets the requirements under the platform assertions.

6.2.6. 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.

Last year, we have introduced the *schedulable parametric data-flow (SPDF)* MoC for dynamic streaming applications [60]. SPDF extends the standard dataflow model by allowing rates to be parametric. SPDF was designed to be statically analyzable while retaining sufficient expressive power.

Following the same lines, we have recently proposed the *Boolean Parametric Data Flow (BPDF)* MoC which combines integer parameters (to express dynamic rates) and boolean parameters (to express the activation and deactivation of communication channels) [15], [26], [24]. High dynamism 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 which ensure statically the liveness and the boundedness of BDPF graphs. Our case studies are video decoders for high definition video streaming such as VC-1.

We have proposed a generic and flexible framework to generate parallel ASAP schedules targeted to the new STHORM many-core platform designed by STMicroelectronics [29], [23]. The parametric dataflow graph is associated with generic or user-defined specific constraints aimed at minimizing, timing, buffer sizes, power consumption, or other criteria. The scheduling algorithm executes with minimal overhead and can be adapted to different scheduling policies just by changing some constraints. The safety of both the dataflow graph and constraints can be checked statically and all schedules are guaranteed to be bounded and deadlock free. This parallel scheduling framework has been developed for a parametric MoC without booleans. We are now focusing on extending it to BPDF applications.

This research is the central topic of Vagelis Bebelis' PhD thesis. It is conducted in collaboration with STMicroelectronics.

6.2.7. Abstract Acceleration of general linear loops

We have investigated abstract acceleration techniques for computing loop invariants for numerical programs with linear assignments and conditionals. Whereas abstract interpretation techniques typically over-approximate the set of reachable states iteratively, abstract acceleration captures the effect of the loop with a single, non-iterative transfer function applied to the initial states at the loop head.

In contrast to previous acceleration techniques, our approach applies to any linear loop without restrictions. Its novelty lies in the use of the Jordan normal form decomposition of the loop body to derive symbolic expressions for the entries of the matrix modeling the effect of $n >= 0$ iterations of the loop. The entries of such a matrix depend on n through complex polynomial, exponential and trigonometric functions. Therefore, we introduced an abstract domain for matrices that captures the linear inequality relations between these complex expressions. This results in an abstract matrix for describing the fixpoint semantics of the loop. We also developed a technique to take into account the guard of the loop by bounding the number of loop iterations, which relies again on the Jordan normal form decomposition.

Our approach integrates smoothly into standard abstract interpreters and can handle programs with nested loops and loops containing conditional branches. We evaluate it over small but complex loops that are commonly found in control software, comparing it with other tools for computing linear loop invariants. The loops in our benchmarks typically exhibit polynomial, exponential and oscillatory behaviors that present challenges to existing approaches, that are either too unprecise (classical abstract interpretation) or limited to a restricted class of loops (*e.g.*, translation with resets in the case of abstract acceleration, or stable loops, in the sense of control theory, for ellipsoid methods). Our approach finds non-trivial invariants to prove useful bounds on the values of variables for such loops, clearly outperforming the existing approaches in terms of precision while exhibiting good performance.

A paper presenting this technique has been accepted to POPL'2014. An extended version has been published in arXiv [30].

6.2.8. *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.*, [87] 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 [82]. 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 [61].

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.

In [52], we have proposed a technique for the synthesis of safety controllers for switched systems using multi-scale abstractions that allow us to deal with fast switching while keeping the number of states in the abstraction at a reasonable level. The finest scales of the abstraction are effectively explored only when fast switching is needed, that is when the system approaches the unsafe set. We have implemented these results in the tool COSYMA (COntroller SYnthesis using Multi-scale Abstractions, see Sec. 5.4.2) [20]. The tool accepts a description of a switched system represented by a set of differential equations and the sampling parameters used to define an approximation of the state-space on which discrete abstractions are computed. The tool generates a controller — if it exists — for the system that enforces a given safety or time-bounded reachability specification.

In [19], we have presented an approach using mode sequences of given length as symbolic states for our abstractions. We have shown that the resulting symbolic models are approximately bisimilar to the original switched system and that an arbitrary precision can be achieved by considering sufficiently long mode sequences. The advantage of this approach over existing ones is double: first, the transition relation of the symbolic model admits a very compact representation under the form of a shift operator; second, our approach does not use lattices over the state-space and can potentially be used for higher dimensional systems. We have provided a theoretical comparison with the lattice-based approach and presented a simple criterion enabling to choose the most appropriate approach for a given switched system. We have applied the approach to a model of road traffic for which we have synthesized a schedule for the coordination of traffic lights under constraints of safety and fairness.

6.3. Language Based Fault-Tolerance

Participants: Dmitry Burlyaev, Pascal Fradet, Alain Girault, Jean-Bernard Stefani.

6.3.1. *Automatic Transformations for Fault tolerant Circuits*

In the recent 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 have designed a simple hardware description language inspired from LUSTRE and Lucid Synchrone. It is a core functional language manipulating synchronous boolean streams. We consider both single-event upsets (SEU) and single-event transients (SET) and all fault models of the form "at most 1 SEU or SET within n clock signals". The language's semantics as well as fault modes have been formalized in Coq and many basic (library) properties have been shown on that language.

We have expressed several variants of triple modular redundancy (TMR) as program transformations. We have proposed a verification-based approach to minimize the number of voters in TMR [16]. Our technique guarantees that the resulting circuit *(i)* is fault tolerant to the soft-errors defined by the fault model and *(ii)* is functionally equivalent to the initial one. Our approach operates at the logic level and takes into account the input and output interface specifications of the circuit. Its implementation makes use of graph traversal algorithms, fixed-point iterations, and BDDs. Experimental results on the ITC'99 benchmark suite indicate that our method significantly decreases the number of inserted voters which entails a hardware reduction of up to 55% and a clock frequency increase of up to 35% compared to full TMR. We address scalability issues arising from formal verification with approximations and assess their efficiency and precision.

We are currently studying the definition of other fault-tolerant techniques (*e.g.*, time redundancy, mixed time/space redundancy) as program transformations. We are also considering the use of the Coq proof assistant to certify that the transformations make the programs fault tolerant *w.r.t.* specific fault models. Our long term goal is to design an aspect-like language allowing users to specify and tune a wide range of fault tolerance techniques, while ensuring that the corresponding transformations ensure well-defined fault-tolerance properties. The advantage would be to produce fault-tolerant circuits by specifying fault-tolerant properties/strategies separately from their functional specifications.

6.3.2. *Concurrent flexible reversibility*

In the recent years, we have been investigating reversible concurrent computation, and investigated various reversible concurrent programming models, with the hope that reversibility can shed some light on the common semantic features underlying various forms of fault recovery techniques (including, exceptions, transactions, and checkpoint/rollback schemes).

As part of this research program, we have devised a reversible variant of the higher-order π-calculus, equipped with an imperative rollback operation that allows a concurrent program to be rolled back to a past execution state, and a primitive form of compensation to control (forward execution) after a rollback operation [18]. We have shown that these two extensions provide very powerful primitives for programming different forms of rollback/compensation schemes. We have shown in particular that they are powerful enough to provide a faithful encoding of a notion of communicating transaction proposed in the literature. We have started the development of a behavioral theory for this crollπ calculus, and proved in particular a context lemma, similar to that of the π-calculus, although the reversible machinery makes its proof more involved.

This work was done in collaboration with Inria teams FOCUS in Bologna and CELTIQUE in Rennes, and as part of the ANR REVER project.

6. New Results

6.1. Data Mining for Material Flow Analysis: Application in the Territorial Breakdown of French Regions

One of the major issues for assessment of the long-term sustainability of urban areas is related to the concept of "imported sustainability". In order to produce such an assessment for a given territory, one must first identify and quantify the types of materials used, and the impacts associated to these uses. Material Flow Analysis (MFA) is directly related to how the material circulates and how it is transformed within a territory. In most cases this analysis is performed at national and regional levels, where the statistical data is available. The challenge is to establish such an analysis at smaller scales, e.g. in the case of France, at the department or city level.

We have explored the possibility of applying data analysis at the regional level by generating a mathematical model that can fit well the data at regional scale and estimate well the departmental one. The downscaling procedure relies on the assumption that the obtained model at level 'n' (for example region) will be also true at level 'n+1' (for example department), such that it could properly estimate the unknown data based on a set of chosen drivers (socio-economic data). We have designed and implemented techniques based on parameter optimization and model selection as well as robust estimation, in order to estimate the best drivers for a given set of territories, i.e. the socio-economic data (e.g. employees per type of manufacturing industry, population data, etc.) that best correlate with the production of various types of agricultural or other products [19].

6.2. Calculating indices for urban sprawl

Urban sprawl is a complex concept, that is generally associated with auto-oriented, low-density development. It is the subject of a wide range of research efforts, aiming at understanding and characterizing the underlying driving factors. We have followed up on an effort by Burchfield et al. who proposed a simple measure for urban sprawl, a so-called sprawl index. We proposed several variants of this index with the aim of achieving richer and/or more flexible characterizations of urban sprawl [16]. We also proposed ways of determining metropolitan areas that have similar patterns of urban sprawl, using clustering techniques.

6.3. Computing environmental accounts from the consumer's viewpoint using Input Output Analysis

The Russian-American economist Wassily Leontief introduced Input-Output Analysis (IOA) in the 30's and was awarded the Nobel prize in economy in 1973 for this contribution. IOA is a macro-economic tool which investigates the links and retroactions between the sectors of an economy. It makes it possible to allocate production factors (labor, capital etc.) and environmental externalities of production processes (depletion on resources, emission of pollutants etc.) to final consumption. Our first task was to reproduce the results from various studies on the carbon footprint of France. We couldn't reproduce the results from *Analyse des impacts environnementaux de la consommation des ménages et des marges de manœuvre pour réduire ces impacts* (Ademe, 2012). We underlined a mistake in the mathematical formulas presented in the annex of the paper but couldn't concur it was indeed the source of discrepancy because we were not granted access to the raw data of the study. Our results are however in line with the papers originating from the statistical service of the ministry of ecology (SOeS, J-L Pasquier).

Because working on excel sheets, although widespread in a large number of agencies, proves very inefficient, we started to work on the development of a software called Wassily (see "new results" section) that would automatize the critical calculations. We prepared databases on input-output tables and air emissions of several European countries based on Eurostat data and started to work on the architecture of the software with Julien Alapetite. In parallel, we reviewed the downscaling and regionalizing techniques in the IO literature and looked for the necessary information concerning the Rhône-Alpes region. We concluded that enough data was available in order to carry out downscaling but that data was still too scarce for the finer levels of regionalizing.

6.4. Mapping and land use and land cover change for the ESNET project

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 2030/2040 horizon under various constraints of urban policy planning, changes in agricultural and forest management, and climate change impact on ecosystems. A preliminary task in this research program was the elaboration of very detailed maps (both in terms of land use and of resolution) of the study area at three different dates (1998, 2003 and 2009) based on available satellite and IGN data, in order to characterize past land use patterns as well as agricultural rotation patterns. These have been made and completed at Inria with the hiring of specialized engineers in these tasks, funded by the ESNET program. This exercice informs the next task (land use and land cover change – LUCC – modelling). Hosting this work at Inria was not only logical in terms of the available computer environment, but also useful in terms of visibility of Inria from outside planning agencies.

The LUCC model itself is developed partly at Inria (for modelling expertise) and partly at LECA (for expertise on ecological change drivers). The model development is still underway but in a rather advanced stage. Relevant drivers for urban development have been identified and statistically characterized. The so-called "transition potentials" (which characterize change of land use over a given period of time) are in the process of being calibrated. The next steps involve the completion of this calibration task, the development of relevant scenarios (underway by the whole ESNET collaboration) and projections of land use into the future. Some sensitivity analysis will also be performed in order to characterize the robustness of the model.

6.5. Benchmarking tools for the climate negotiation of GHG emission reduction trajectories

Climate negotiations related to global warming are another important issue of sustainable development. In this framework that is place at international scale we propose a benchmarking tool that is designed to avoid the main limitations of actual negotiation schemes. Our approach is based on the original Soft Landing proposition, made by Criqui and Kouvaritakis in the early 2000. We develop an up to date solution which improves the original idea mainly by introducing common but differentiated emission reduction profiles and by developing a dedicated algorithm for that purpose (called REDEM). To be compatible with global objectives, it is commonly accepted that for most developing regions, the national emission curves should admit a maximum and then should progressively decline. Similarly, we emphasize the fact that, in order to achieve the global objectives, all states will have to entail mitigation efforts, the intensity which may be measured by the rate of variation of the national emissions. At one point, the effort will reach a maximum, when the rate of variation in absolute value is at its maximum, and then decrease. In other words, there will also be a peak in the effort. Then we propose to base the benchmark on this peak of effort. This work has been done in collaboration with EDDEN Laboratory, in particular Patrick Criqui and Constantin Ilasca.

TYREX Team

6. New Results

6.1. Multimedia Models and Formats

Modeling and authoring web content including rich media and interactions is still an open problem. We have drawn a reference state of the art of this area in [17]. We have also contributed to the cultural heritage domain through the experimentation of a multimedia production chain for digitized theatre performances based on semantic annotations [12].

In the context of the Claire project (see section 7.1.1), the results we obtained in 2013 in this area are:

- providing a data model which is format agnostic to cope with existing and future rendering systems. More specifically, we specified a chaptering component that includes the structuration and navigation features for continuous media such as video.
- prototyping a web environment for authoring such rich media (see section 5.2 . This authoring services are built as a contribution of the Mozilla Popcorn Maker project (Popcorn Maker).
- experimenting this environment for the production of the multimedia part of the first MOOC developed by OpenClassrooms, our main partner in this project.

6.2. XML Processing

In the area of XML processing, we obtained new results in several directions:

- We showed how to translate Schematron descriptions into the tree logic [15];
- We built the first IDE equipped with path reasoning capabilities [13];
- We showed that a whole class of logical combinators (or "macros") can be used as an intermediate language between the query language and the logical language [20]. This provides a gain in terms of succinctness for the logical formalism.
- We continued our work on a novel technique and a tool for the static type-checking of XQuery programs, using backward type inference.
- We made preliminary investigations on how to support backward navigation axes in the static type checking for XQuery [18].
- In a joint work with the Exmo team, we benchmarked solvers for deciding the problem of query containment for fragments of SPARQL [14].

We briefly review these results below.

6.2.1. Rule-Based Validation à la Schematron

One major concept in web development using XML is validation: checking whether some document instance fulfills structural constraints described by some schema. Over the last few years, there has been a growing debate about XML validation, and two main schools of thought emerged about the way it should be done. On the one hand, some advocate the use of validation with respect to complete grammar-based descriptions such as DTDs and XML Schemas. On the other hand, motivated by a need for greater flexibility, others argue for no validation at all, or prefer the use of lightweight constraint languages such as Schematron with the aim of validating only required constraints, while making schema descriptions more compositional and more reusable.

We built a compiler for Schematron [15]. This compiler takes a Schematron description as input and generates the corresponding constraints as a logical formula. We showed that validators used in each of these approaches share the same theoretical foundations, meaning that the two approaches are far from being incompatible. Our findings include that modal logic can be seen as a unifying formal ground for the construction of robust and efficient validators and static analyzers using any of these schema description techniques. This reconciles the two approaches from both a theoretical and a practical perspective, therefore facilitating any combination of them.

6.2.2. Integrated Development Environments with Path Reasoning Capabilities

One of the challenges in web development is to help achieving a good level of quality in terms of code size and runtime performance, for popular domain-specific languages such as XQuery, XSLT, and XML Schema. We presented the first IDE augmented with static detection of inconsistent XPath expressions that assists the programmer for simplifying the development and debugging of any application involving XPath expressions [13]. The tool is based on newly developed formal verification techniques based on expressive modal logics, which are now mature enough to be introduced in the process of software development. We further develop this idea in the context of XQuery for which we introduce an analysis for identifying and eliminating dead code automatically. This proof of concept aims at illustrating the benefits of equipping modern IDEs with reasoning capabilities.

6.2.3. Logical Combinators for Rich Type Systems

A popular technique in the static analysis for query languages relies on the construction of compilers that effectively translate 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.

We showed that a whole class of logical combinators (or "macros") can be used as an intermediate language between the query language and the logical language [20]. Those logical combinators provide an exponential gain in succinctness over the corresponding explicit logical representation, yet preserve the typical exponential time complexity of the subsequent logical decision procedure. 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 illustrated this from a very practical point of view on a few examples such as numerical occurrence constraints and tree frontier properties, which are concrete problems found in the XML world.

6.2.4. Backward type inference for XQuery

We have continued our work on the design of a novel technique for static type-checking of XQuery programs based on backward type inference. The tool looks for errors in the program by jointly analyzing the source code of the program, input and output schemas that respectively describe the sets of documents admissible as input and as output of the program. The crux and the novelty of our results reside in the joint use of backward type inference and a two-way logic to represent inferred tree type portions. This allowed us to design and implement a type-checker for XQuery which is more precise and supports a larger fragment of XQuery than the approaches previously proposed in the literature; in particular compared to the only few actually implemented static type-checkers such as the one in Galax. The whole system uses compilers and a satisfiability solver for deciding containment for two-way regular tree expressions. Our tool takes an XQuery program and two schemas S_{in} and S_{out} as input. If the program is found incorrect, then it automatically generates a counter-example valid w.r.t. S_{in} and such that the program produces an invalid output w.r.t S_{out}. This counter-example can be used by the programmer to fix the program.

6.2.5. XQuery and Static Typing: Tackling the Problem of Backward Axes

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. We argue that this is due to a discrepancy between the semantics of the language and its type algebra: namely, the values of the language are (possibly inner) tree nodes, which may have siblings and ancestors in the data. The types on the other hand are regular tree types, as usual in the XML world: they describe sets of trees. The type associated to a node then corresponds to the subtree whose root is that node and contains no information about the rest of the data. This makes navigational expressions using 'backward axes,' which return e.g. the siblings of a node, impossible to type.

In [18], we discussed how to solve this discrepancy and proposed a compromise: to use extended types representing possibly inner tree nodes in some key parts of a program, and to cut out the subtrees from their original context in the rest.

6.2.6. Semantic Web queries and μ-calculus

Querying the semantic web is mainly done through the SPARQL language or its extensions through paths and entailment regimes. Query containment is the problem of deciding if the answers to a query are included in those of another query for any queried database [4], [3]. This problem is very important for query optimization purposes. In the SPARQL context, it can be equally useful for distributing federated queries or for implementing schema-based access control. In order to experimentally assess implementation strengths and limitations, we provided a first SPARQL containment test benchmark. We studied the query demographics on DBPEDIA logs to design benchmarks for relevant query containment solvers. We tested available solvers on their domain of applicability on three different benchmark suites [14]. (i) tested solutions are overall functionally correct, (ii) in spite of its complexity, SPARQL query containment is practicable for acyclic queries, (iii) state-of-the-art solvers are at an early stage both in terms of capability and implementation.

This work has been developed in collaboration with the EXMO team. The benchmarks, results and software are available at http://sparql-qc-bench.inrialpes.fr.

6.3. Mixed Reality Environment

The concept of Mixed Reality comes from the fact that the real-virtual dichotomy is not sharp. Augmented Reality (AR) mode refers to all cases in which the auditory or visual display of a real environment is augmented by virtual sound or graphic objects. Pedestrian navigation is one of the numerous applications that fit into this field. Depending on the real speed of the user and on the real environment in which he moves (inside a building, ...), the system is augmented with synthetic audio instructions and points of interest. OpenStreetMap format has been extended to support navigation authoring and information related to the various passive or active location providers supported by IXE such as PDR, GPS and NFC.

6.3.1. Navigation Authoring

We defined a cue-based XML language (A2ML, for Advanced Audio Markup Language) using SMIL for internal and external synchronization of sound objects. A2ML is specified by a RELAX-NG grammar. A rule-based selector mechanism allows defining style sheets for OpenStreetMap (OSM) elements. This auditory display together with TTS makes our IXE browser accessible to visually impaired people. Format and Delivery for Mixed Reality Content IXE is based on an extended OSM data format with triggering zones, relations or groups with specific semantics and nodes or POIs whose URIs refer to content expressed in HTML5 and A2ML. Content delivery can be of two types, push or pull. Push content is coming from POIs which trigger when the user enters a new zone. This kind of content is very useful for navigation. We support it through a triggering specification that is inserted in the OpenStreetMap document. We use style sheets with rules to specify both the audio and visual rendering of the various types of OSM nodes. Pull content allows users to search detailed information about the artifacts that are located in the content referenced by the POI. Most of the time, this content is described using HTML5 and A2ML.

6.3.2. Location Provider Fusion

Pedestrian navigation can be done with several sensors. GPS locations are better for outside locations, PDR is useful to guide people indoor, but we can also use NFC tags, user proprioception, wifi... Our researches focus on a smart fusion of providers depending on sensor accuracy and on the context in which the person moves. We start by using Kalman Filter to smooth locations and disable jumps during the walk. These algorithms have been successfully tested during Venturi Y2 demo.

6.3.3. Map Rendering

We worked on offline map rendering around two solutions. The first one is based on an open source Android project called Mapsforge; it provides a tile generation mechanism from a given OpenStreetMap file and a tile caching system for fast rendering on mobile devices. We mainly enhanced the open source project by increasing the zoom level limitation (21 by default) to 24 for displaying indoor maps. The other solution on which we worked is SVG-oriented and based on OpenLayers (dedicated to web browsers). As the rendering uses SVG we are no longer limited by a maximum zoom level. On the other hand, the SVG drawing has to be fully designed by the author, as we don't support, for the time being, SVG file generation from an OpenStreetMap document. These two approaches are different and their uses depend on the desired level of customization of the rendering (generated automatically or manually).

URBANET Team

6. New Results

6.1. Characterizing and measuring urban networks

Participants: Marco Fiore, Diala Naboulsi, Razvan Stanica, Sandesh Uppoor

6.1.1. Properties of urban vehicular traffic and implications on mobile networking.

The goal of Sandesh Uppoor's PhD thesis [4] was to model and understand the mobility dynamics of high-speed vehicular users and their effect on wireless network architectures in an urban environment. Given the importance of developing the study on a realistic representation of vehicular mobility, we first survey the most popular approaches for the generation of synthetic road traffic and discuss the features of publicly available vehicular mobility datasets. Using original travel demand information of the population of a metropolitan area (Cologne area, Germany), detailed road network data and realistic microscopic driving models, we propose a novel state-of-art vehicular mobility dataset that closely mimics the real-world road traffic dynamics in both time and space [25]. We then study the impact of such mobility dynamics from the perspective of wireless cellular network architecture in presence of a real-world base station deployment. In addition, by discussing the effects of vehicular mobility on autonomous network architecture, we hint at the opportunities for future heterogeneous network paradigms and demonstrate how incomplete representations of vehicular mobility may result in over-optimistic network connectivity and protocol performance [8].

Motivated by the time-evolving mobility dynamics observed in our original dataset, we also propose an on line approach to predict near-future macroscopic traffic flows. We analyze the parameters affecting the mobility prediction in an urban environment and unveil when and where network resource management is more crucial to accommodate the traffic generated by users on-board. Such studies unveil multiple opportunities in transportation management either for building new roads, installing electric charging points, or for designing intelligent traffic light systems, thereby contributing to urban planning.

6.1.2. Feasibility of multi-hop vehicular communications in an urban environment.

Despite the growing interest in a real-world deployment of vehicle- to-vehicle communication, many topological features of the resulting vehicular network remain largely unknown. We still lack a clear understanding of the level of connectivity achievable in large-scale urban scenarios, of the availability and reliability of connected multi-hop paths, and of the evolution of such features over daytime. In [14], we investigate how the instantaneous topology of the vehicular network would look like in the case of a typical middle-sized European city, using the example of the Cologne mobility trace. Through a complex network analysis, we unveil the low connectivity, availability, reliability and navigability of the network, and exploit our findings to derive network design and usage guidelines.

6.1.3. Investigating the accuracy of mobile urban sensing.

Community urban sensing is one of the emerging applications enabled by the growing popularity of mobile user devices, like smartphones and in-vehicle monitoring systems. Such devices feature sensing and wireless communication capabilities, which enable them to sample large-scale phenomena, like air pollution and vehicular traffic congestion, and upload these data to the Internet. In [10], we focus on the above scenario and investigate the level of accuracy that can be achieved in estimating the phenomenon of interest through a mobile crowdsourcing application. Specifically, we take a signal processing-based approach and leverage results on signal reconstruction from sets of irregularly spaced samples. We apply such results to a realistic scenario where samples are collected by vehicular and pedestrian users, and study the accuracy level of the phenomenon estimation as the penetration rate of the sensing application varies.

6.1.4. Analysis of mobile network call detail records.

The growing ubiquity of mobile communications has offered researchers new possibilities to understand human mobility over the last few years. In [22], we analyze Call Detail Records (CDR) made available within the context of the Orange D4D Challenge, focusing on calls of individuals in the city of Abidjan, Ivory Coast, over a period of five months. Our results illustrate how aggregated CDR can be used to tell apart typical and special mobility behaviors, and demonstrate how macroscopic mobility flows extracted from these cellular network data reflect the daily dynamics of a highly populated city. We discuss how these macroscopic mobility flows can help solve problems in developing urban areas.

6.2. Scalable solutions for capillary networks

Participants: Isabelle Augé-Blum, Jin Cui, Marco Fiore, Ochirkhand Erdene-Ochir, Alexandre Mouradian, Hervé Rivano, Razvan Stanica, Fabrice Valois

6.2.1. Real-time wireless sensor networks.

Critical applications for WSNs are emerging, with real-time and reliability requirements. Critical applications are applications on which depend human lives and the environment: a failure of a critical application can thus have dramatic consequences. We are especially interested in anomaly detection applications (forest fire detection, landslide detection, intrusion detection, etc), which require bounded end to end delays and high delivery ratio. Few WSNs protocols of the literature allow to bound end to end delays. Among the proposed solutions, some allow to effectively bound the end to end delays, but do not take into account the characteristics of WSNs (limited energy, large scale, etc). Others take into account those aspects, but do not give strict guaranties on the end to end delays. In this sense, the PhD thesis of Alexandre Mouradian [2] proposes a real-time anomaly detection solution composed of:

- A virtual coordinate system which allows to discriminate nodes in a 2-hop neighborhood and to bound the number of hops between any source and the sink.

- A cross-layer protocol for WSNs (named RTXP) based on the proposed virtual coordinate system. Thanks to these coordinates it is possible to introduce determinism in the accesses to the medium and to bound the hop-count, this allows to bound the end to end delay. RTXP adapts its duty-cycle to the traffic loads and uses an opportunistic routing scheme to increase its delivery ratio. We show, by simulation, that RTXP outperforms real-time protocols of the literature for anomaly detection in WSNs under harsh radio conditions.

- A real-time aggregation scheme to mitigate the alarm storm problem which causes collisions and congestion and thus limit the network lifetime. This scheme is also based on the virtual coordinate system and is used before RTXP in order to reduce the number of similar alarms converging toward the sink.

6.2.2. Formal verification of wireless sensor networks protocols.

WSN protocols used by critical applications must be formally verified in order to provide the strongest possible guaranties: simulations and tests are not sufficient in this context, formal proofs of compliance with the specifications of the application have to be provided.. Unfortunately the radio link is unreliable and it is thus difficult to give hard guarantees on the temporal behavior of the protocols. Indeed, a message may experience a very high number of retransmissions and the temporal guarantee can only be given with a certain probability. This probability must meet the requirements of the application. Network protocols have been successfully verified on a given network topology without taking into account unreliable links. Nevertheless, the probabilistic nature of radio links may change the topology (links which appear and disappear). Thus instead of a single topology we have a set of possible topologies, each topology having a probability to exist. In [12], we propose a method that produces the set of topologies, checks the property on every topology, and gives the probability that the property is verified. This technique is independent from the verification technique, i.e. each topology can be verified using any formal method which can give a "yes" or "no" answer to the question: "Does the model of the protocol respect the property?". We apply this method on the f-MAC protocol. We use

UPPAAL model checker as verification tool. We implement a tool that automatizes the process and thus show the feasibility of our proposition. We compare the results of the verification with simulation results. It appears that the verification is, as expected, conservative but not overly pessimistic compared to the simulated worst case. Besides we show that f-MAC is a reliable real-time protocol for WSNs (for up to 6 nodes), as we were not able to detect faults.

Moreover, in [2], a verification technique which mixes Network Calculus and Model Checking is proposed, in order to be both scalable and exhaustive. This technique consists in modeling the interaction of each node with the rest of the network with arrival curves and then to verify with UPPAAL that each node is capable of handling these interactions while meeting the deadlines. We apply this methodology in order to formally verify our pervious proposition, RTXP.

6.2.3. *Reliability in wireless sensor networks.*

WSN critical applications require the respect of time and reliability constraints. In [13], we provide a theoretical study of the reliability in WSNs. We define the reliability as the probability of success of an end-to-end transmission in the WSN. In this work, we use two radio propagation models : a basic model where the nodes have a set of neighbors they can communicate with, with a given probability, and the log-normal shadowing model, where probability of reception depends on the emitter-receiver distance. We determine the reliability of two routing schemes : unicast-based routing (classical routing) and broadcast-based routing (opportunistic routing). We conclude that the broadcast-based routing allows to reach a higher reliability than the unicast case. The main result is that we show the existence of a reliability bottleneck at the sink node in the case of the broadcast-based routing. We show that the addition of another sink improves the reliability of the network in this case.

6.2.4. *Resiliency in wireless sensor networks.*

Because of their open and unattended deployment, in possibly hostile environments, powerful adversaries can easily launch Denial-of-Service (Dos) attacks on wireless sensor networks, cause physical damage to sensors, or even capture them to extract sensitive information (encryption keys, identities, addresses, etc.). Consequently, the compromised node poses severe security and reliability concerns, since it allows an adversary to be considered as a legitimate node inside the network. To cope with these "insider" attacks, stemming from node compromise, "beyond cryptography" algorithmic solutions must be envisaged to complement the traditional cryptographic solutions. In this sense, in [1], we first propose the resiliency concept. Our goal is to propose a definition of the resiliency in our context (security of WSNs routing protocols) and a new metric to compare routing protocols. The originality of this metric is that we combine the graphical representation (qualitative information) with the aggregation method (quantitative information). We introduce a two dimensional graphical representation with multiple axes forming an equiangular polygon surface. This method allows to aggregate meaningfully several parameters and makes it easier to visually discern various trade-offs, thus greatly simplifying the process of protocol comparison.Secondly, we propose the protocol behaviors enhancing resiliency. Our proposition consists in three elements: (i) introduce random behaviors (ii) limit route length (iii) introduce data replication. Random behaviors increase uncertainty for an adversary, making the protocols unpredictable. Data replication allows route diversification between the sources and the sink, thus improving the delivery success and fairness. Limitation of the route length is necessary to reduce the probability of a data packet to meet a malicious insider along the route. The quantitative metric enables to propose a new resiliency taxonomy of WSNs routing protocols. According to this taxonomy, the gradient based routing is the most resilient when it is combined with the proposed behaviors. Thirdly, several variants of the gradient-based routing (classical and randomized) under more complex and realistic adversary model (several combined attacks) are considered to extend our simulations. Several values of bias are introduced to the randomized variants and two data replication methods (uniform and adaptive) are considered. Without attacks, the most biased variants without replications are the most efficient. However, under moderate attacks, the replication uniform is the most adapted, while under intense attacks, the replication adaptive is the most suitable. Finally, a theoretical study of the resiliency is introduced. We present an analytical study of the biased random walk routing under attacks. The influence of bias is evaluated and two replication methods that previously evaluated by simulations are considered. After presenting the delivery success and the energy consumption of all scenarios, we

evaluate them with our resiliency metric. This study permits to confirm the results obtained with simulations and it shows that the bias is essential to enhance the resiliency of random routing.

6.2.5. *Data aggregation in wireless sensor networks.*

Data aggregation is a crucial problem in wireless sensor networks due to their constrained-energy and constrained-bandwidth nature. In [26], we highlight the aggregation benefits at the Network layer and MAC layer by modeling the energy consumption for some energy-efficient routing protocols and MAC protocols. Besides, we define two parameters, the aggregation ratio and the packet size coefficient to evaluate the efficiency of an aggregation method, and to discuss the trade-off. Additionally, we investigate the differences between time series and compressive sensing, which are representative state-of-the-art solutions for forecasting aggregation and compressing aggregation respectively.

6.2.6. *Routing in delay-tolerant networks.*

Delay-Tolerant Networks (DTN) model systems that are characterized by intermittent connectivity and frequent partitioning. Routing in DTNs has drawn much research effort recently. Since very different kinds of networks fall in the DTN category, many routing approaches have been proposed. In particular, the routing layer in some DTNs has information about the schedules of contacts between nodes and about data traffic demand. Such systems can benefit from a previously proposed routing algorithm based on linear programming that minimizes the average message delay. This algorithm, however, is known to have performance issues that limit its applicability to very simple scenarios. In [9], we propose an alternative linear programming approach for routing in Delay-Tolerant Networks. We show that our formulation is equivalent to that presented in a seminal work in this area, but it contains fewer LP constraints and has a structure suitable to the application of Column Generation (CG). Simulation shows that our CG implementation arrives at an optimal solution up to three orders of magnitude faster than the original linear program in the considered DTN examples.

6.2.7. *Performance evaluation of vehicular communications.*

Wireless vehicular networks face different problems and challenges, especially in a dense urban environment. In [23], we first characterize the different types of loss in vehicular networks: radio propagation problems, expired security messages, collision with one hop neighbor and collisions with hidden terminals. In a second step, we give the architecture of the wireless vehicular network and describe the Medium Access Control (MAC) quality of service mechanisms proposed by vehicular environment standards that aim at meeting the road drivers' expectation and increasing road safety. To complete this image, in [24], we provide a literature survey that covers the solutions proposed in order to enable critical dissemination of urgent messages and surpass the challenging vehicular dynamic topology. More particularly, we detail the following techniques: beaconing frequency reduction, transmit rate control, power control, adaptation of the contention window and adaptation of the carrier sense threshold.

6.2.8. *Secure node localization in mobile ad-hoc networks.*

A growing number of ad hoc networking protocols and location-aware services require that mobile nodes learn the position of their neighbors. However, such a process can be easily abused or disrupted by adversarial nodes. In absence of a priori trusted nodes, the discovery and verification of neighbor positions presents challenges that have been scarcely investigated in the literature. In [6], we address this open issue by proposing a fully distributed cooperative solution that is robust against independent and colluding adversaries, and can be impaired only by an overwhelming presence of adversaries. Results show that our protocol can thwart more than 99% of the attacks under the best possible conditions for the adversaries, with minimal false positive rates.

In a vehicular context, knowledge of the location of vehicles and tracking of the routes they follow are a requirement for a number of applications. However, public disclosure of the identity and position of drivers jeopardizes user privacy, and securing the tracking through asymmetric cryptography may have an exceedingly high computational cost. In [11], we address all of the issues above by introducing A-VIP, a lightweight privacy-preserving framework for tracking of vehicles. A-VIP leverages anonymous position beacons from

vehicles, and the cooperation of nearby cars collecting and reporting the beacons they hear. Such information allows an authority to verify the locations announced by vehicles, or to infer the actual ones if needed. We assess the effectiveness of A-VIP through testbed implementation results.

6.3. Cellular network solutions

Participants: Marco Fiore, Anis Ouni, Hervé Rivano, Razvan Stanica, Fabrice Valois

6.3.1. *Optimizing capacity and energy consumption in wireless mesh networks.*

Wireless mesh networks (WMN) are a promising solution to support high data rate and increase the capacity provided to users, e.g. for meeting the requirements of mobile multimedia applications. However, the rapid growth of traffic load generated by the terminals is accompanied by an unsustainable increase of energy consumption, which becomes a hot societal and economical challenges. This thesis relates to the problem of the optimization of network capacity and energy consumption of wireless mesh networks. The network capacity is defined as the maximum achievable total traffic in the network per unit time.

The thesis of Anis Ouni [3] addresses this issue and is divided into four main parts. First, we address the problem of improvement of the capacity of 802.11 wireless mesh networks. We highlight some insensible properties and deterministic factors of the capacity, while it is directly related to a bottleneck problem. Then, we propose a joint TDMA/CSMA scheduling strategy for solving the bottleneck issue in the network.

Second, we focus on broadband wireless mesh networks based on time-frequency resource management. In order to get theoretical bounds on the network performances, we formulate optimization models based on linear programming and column generation algorithm. These models lead to compute an optimal offline configuration which maximizes the network capacity with low energy consumption. A realistic SINR model of the physical layer allows the nodes to perform continuous power control and use a discrete set of data rates.

Third, we use the optimization models to provide practical engineering insights on WMN. We briefly study the tradeoff between network capacity and energy consumption using a realistic physical layer and SINR interference model [27]. In particular, we show that power control and multi-rate functionalities allow to reach optimal throughput with lower energy consumption using a mix of single hop and multi- hop routes.

Finally, we focus on capacity and energy optimization for heterogeneous cellular networks. We develop optimization tools to calculate an optimal configuration of the network that maximizes the network capacity with low energy consumption. We then propose a heuristic algorithm that calculates a scheduling and partial sleeping of base stations in two different strategies, called LAFS and MAFS.

6.3.2. *Sleep protocols for heterogeneous LTE networks.*

The tremendous increase of the traffic demand in cellular networks imposes a massive densification of the traditional cellular infrastructure. The network architecture becomes heterogeneous, in particular 4G networks where LTE micro-eNodeBs are deployed to strengthen the coverage of macro-eNodeBs. This densification yields major issues related to the energy consumption of the infrastructure. Indeed, there is fixed and significant amount of energy required to run each additional node, whatever the traffic load of the network. Mitigating this fixed energy consumption is therefore a major challenge from a societal and economical viewpoint. Extensive researches about energy-saving highlight that to save energy the better strategy is to switch off the radio part of nodes. This is the heart of wireless sensor networks energy-saving strategies, even though the objective for WSN is to maximize the battery life of each individual nodes. In [18], we develop a parallel between the principles of WSN protocols and the requirements of cellular infrastructures. We then propose a distributed and localized algorithm to dynamically switch off and on the micro-eNodeBs of an LTE heterogeneous network following the traffic demand evolution in time and analyze it in terms of energy savings. We show that one can expect energy savings of approximately 12% when implementing sleep modes whereas the energy cost for sending the traffic decreases by 24%.

6.3.3. *Content downloading through a vehicular network.*

The focus of the work we present in [7] is twofold: information dissemination from infrastructure nodes deployed along the roads, the so-called Road-Side Units (RSUs), to passing-by vehicles, and content downloading by vehicular users through nearby RSUs. In particular, in order to ensure good performance for both content dissemination and downloading, the presented study addresses the problem of RSU deployment and reviews previous work that has dealt with such an issue. The RSU deployment problem is then formulated as an optimization problem, where the number of vehicles that come in contact with any RSU is maximized, possibly considering a minimum contact time to be guaranteed. Since such optimization problems turn out to be NP-hard, heuristics are proposed to efficiently approximate the optimal solution. The RSU deployment obtained through such heuristics is then used to investigate the performance of content dissemination and downloading through ns2 simulations. Simulation tests are carried out under various real-world vehicular environments, including a realistic mobility model, and considering that the IEEE 802.11p standard is used at the physical and medium access control layers. The performance obtained in realistic conditions is discussed with respect to the results obtained under the same RSU deployment, but in ideal conditions and protocol message exchange. Based on the obtained results, some useful hints on the network system design are provided.

6.3.4. *Offloading Floating Car Data.*

Floating Car Data (FCD) is currently collected by moving vehicles and uploaded to Internet-based processing centers through the cellular access infrastructure. As FCD is foreseen to rapidly become a pervasive technology, the present network paradigm risks not to scale well in the future, when a vast majority of automobiles will be constantly sensing their operation as well as the external environment and transmitting such information towards the Internet. In order to relieve the cellular network from the additional load that widespread FCD can induce, we study [16] a local gathering and fusion paradigm, based on vehicle-to-vehicle (V2V) communication. We show how this approach can lead to significant gain, especially when and where the cellular network is stressed the most. Moreover, we propose several distributed schemes to FCD offloading based on the principle above that, despite their simplicity, are extremely efficient and can reduce the FCD capacity demand at the access network by up to 95%.

6.3.5. *Mobile malware propagation in vehicular networks.*

The large-scale adoption of vehicle-to-vehicle (V2V) communication technologies risks to significantly widen the attack surface available to mobile malware targeting critical automobile operations. Given that outbreaks of vehicular computer worms self-propagating through V2V links could pose a significant threat to road traffic safety, it is important to understand the dynamics of such epidemics and to prepare adequate countermeasures. In [17], we perform a comprehensive characterization of the infection process of variously behaving vehicular worms on a road traffic scenario of unprecedented scale and heterogeneity. We then propose a simple yet effective data-driven model of the worm epidemics, and we show how it can be leveraged for smart patching infected vehicles through the cellular network in presence of a vehicular worm outbreak.