



RESEARCH CENTER
Grenoble - Rhône-Alpes

FIELD

Activity Report 2012

Section New Results

Edition: 2013-04-24

ALGORITHMICS, PROGRAMMING, SOFTWARE AND ARCHITECTURE

1. ARIC Team	4
2. COMPSYS Project-Team	10
3. CONVECS Team	15
4. POP ART Project-Team	23

APPLIED MATHEMATICS, COMPUTATION AND SIMULATION

5. BIPOP Project-Team	30
6. MISTIS Project-Team	35
7. NANO-D Team	46
8. NECS Project-Team	53
9. OPALE Project-Team	63

COMPUTATIONAL SCIENCES FOR BIOLOGY, MEDICINE AND THE ENVIRONMENT

10. BAMBOO Project-Team	75
11. BEAGLE Team	83
12. DRACULA Project-Team	86
13. IBIS Project-Team	89
14. MOISE Project-Team	94
15. NUMED Project-Team	107
16. STEEP Exploratory Action	108

NETWORKS, SYSTEMS AND SERVICES, DISTRIBUTED COMPUTING

17. AVALON Team	111
18. DANTE Team	119
19. MESCAL Project-Team	123
20. MOAIS Project-Team	127
21. PLANETE Project-Team	129
22. ROMA Team	149
23. SARDES Project-Team	155
24. SOCRATE Team	157
25. URBANET Team	163

PERCEPTION, COGNITION, INTERACTION

26. E-MOTION Project-Team	168
27. EXMO Project-Team	195
28. IMAGINE Team	197
29. LEAR Project-Team	210
30. MAVERICK Team	219
31. MORPHEO Team	229
32. PERCEPTION Team	234
33. PRIMA Project-Team	237
34. WAM Project-Team	241

ARIC Team

6. New Results

6.1. Applications

Florent de Dinechin contributed high-performance signal processing on an FPGA to a prototype of high-throughput receiver for optical fiber transmission developed by Alcatel [33]. He also wrote a book chapter exposing the potential of FPGA-specific arithmetic for high-performance computing [49].

6.2. Hardware and FPGA Arithmetic

6.2.1. Mixed-precision fused multiply-and-add

With B. de Dinechin, from Kalray, N. Brunie and F. de Dinechin proposed to extend the classical fused-multiply-and-add operator with a larger addend and result. This enables higher-precision computation of sums of products at a cost that remains close to that of the classical FMA [29].

6.2.2. Multiplication by rational constants versus division by a constant

Motivated by the division by 3 or by 9 appearing in some stencil kernels, F. de Dinechin investigated how the periodicity of the binary representation of a rational constant could be exploited to design an architecture multiplying by this constant [18]. With L. S. Didier, this approach was then compared to a specialisation of divider architectures to the division by small integer constants, which is shown to match well the fine structure of FPGAs [32].

6.2.3. Floating-point exponentiation on FPGA

F. de Dinechin, with P. Echeverria and M. Lopez-Vallejo (U. Madrid) and B. Pasca (Altera), implemented the first floating-point unit for the pow and powr functions of the IEEE-754-2008 standard [50]. These functions compute x^y , and differ only in the specification of special cases. The implementation, parameterized in exponent and significand size, combines suitably modified exponential and logarithm units.

6.2.4. Arithmetic around the bit heap

F. de Dinechin, M. Istoan, G. Sergent, K. Illyes, B. Popa, and N. Brunie extended FloPoCo with a versatile framework for manipulating sums of weighted bits [51], [44]. This is a relevant way of implementing polynomials, filters and other coarse arithmetic cores.

6.2.5. Improving computing architectures

To improve High-Level Synthesis (HLS) for FPGAs, B. Pasca (former PhD student in AriC), with Ch. Alias (Inria Compsys) and A. Plesco (Zettice) developed tiling and scheduling algorithms that exploit the deeply pipelined operator at the core of a computing kernel [14].

With S. Collange and G. Diamos, N. Brunie proposed improvements in the architecture of general-purpose graphical processing units [28].

N. Brunie and F. de Dinechin, with Kalray's B. de Dinechin, are investigating embedding a reconfigurable core in the Kalray MPPA architecture. For this purpose, N. Brunie developed an environment for the design exploration of such an accelerator. This environment produces the hardware on one side, and its programming tools on the other side [43].

6.3. Elementary Functions

6.3.1. (M,p,k) -friendly points: a table-based method for trigonometric function evaluation

N. Brisebarre, M. Ercegovac (U. California at Los Angeles) and J.-M. Muller [25] present a new way of approximating the sine and cosine functions by a few table look-ups and additions. It consists in first reducing the input range to a very small interval by using rotations with “ (M, p, k) friendly angles”, proposed in this work, and then by using a bipartite table method in a small interval. An implementation of the method for 24-bit case is described and compared with CORDIC. Roughly, the proposed scheme offers a speedup of 2 compared with an unfolded double-rotation radix-2 CORDIC.

6.3.2. On Ziv’s rounding test

With Ch. Lauter (LIP6), F. de Dinechin, J.-M. Muller and S. Torres proved and generalized a code sequence due to Ziv, which is used to round correctly a real value approximated (with a known error bound) as the unevaluated sum of two floating-point numbers [52].

6.4. Arithmetic Algorithms

6.4.1. Binary floating-point operators for VLIW integer processors

C.-P. Jeannerod and J. Jourdan-Lu [35] proposed software implementations of \sin , \cos and sincos over $[-\pi/4, \pi/4]$ that have proven 1-ulp accuracy and whose respective latencies on STMicroelectronics’ ST231 VLIW integer processor are 19, 18 and 19 cycles. To get such performances they introduced a novel algorithm for simultaneous sine and cosine that combines univariate and bivariate polynomial evaluation schemes.

In the same context, C.-P. Jeannerod, J. Jourdan-Lu and C. Monat (STMicroelectronics Compilation Expertise Center, Grenoble) [36] studied the implementation of *custom* (i.e., specialized, fused, or simultaneous) operators, and provided qualitative evidence of the benefits of supporting such operators in addition to the five basic ones: this allows to be up to 4.2x faster on individual calls, and up to 1.59x faster on DSP kernels and benchmarks.

6.4.2. Error bounds for complex floating-point division with an FMA

Assuming that a fused multiply-add (FMA) instruction is available, C.-P. Jeannerod, N. Louvet and J.-M. Muller [37] obtained sharp error bounds for various alternatives to Kahan’s 2 by 2 determinant algorithm. Combining such alternatives with Kahan’s original scheme leads to componentwise-accurate algorithms for complex floating-point division, and for these algorithms sharp or reasonably sharp error bounds were also obtained.

6.4.3. Computation of correctly-rounded sums

P. Kornerup (U. of Southern Denmark), V. Lefèvre and J.-M. Muller [19] have shown that among the set of the algorithms with no comparisons performing only floating-point additions/subtractions, the 2Sum algorithm introduced by Knuth is minimal, both in terms of number of operations and depth of the dependency graph. They also prove that under reasonable conditions, an algorithm performing only round-to-nearest additions/subtractions cannot compute the round-to-nearest sum of at least three floating-point numbers. They also present new results about the computation of the correctly-rounded sum of three floating-point numbers.

6.4.4. Comparison between binary64 and decimal64 floating-point numbers

N. Brisebarre, C. Lauter (U. Paris 6), M. Mezzarobba and J.-M. Muller [27] introduce an algorithm that allows one to quickly compare a binary64 floating-point (FP) number and a decimal64 FP number, assuming the “binary encoding” of the decimal formats specified by the IEEE 754-2008 standard for FP arithmetic 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.

6.5. Computer Algebra

6.5.1. *Faster multivariate interpolation with multiplicities*

M. Chowdhury (U. Western Ontario), C.-P. Jeannerod, V. Neiger (ENS de Lyon), É. Schost (U. Western Ontario) and G. Villard proposed fast randomized algorithms for interpolating multivariate polynomials with multiplicities. In the special bivariate case, this allows to accelerate the interpolation step of Guruswami and Sudan's list-decoding by a factor (list size)/(multiplicity).

6.5.2. *On the complexity of solving quadratic boolean systems*

M. Bardet (U. Rouen), J.-Ch. Faugère (PolSys), B. Salvy, and P.-J. Spaenlehauer (PolSys) [16] dealt with the fundamental problem in computer science of finding all the common zeroes of polynomials systems of quadratic polynomials over the field with 2 elements. The cryptanalysis of several modern ciphers reduces to this problem. Up to now, the best complexity bound was reached by an exhaustive search. They 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, their complexity breaks the 2^n barrier. Experiments on random systems show that the algebraic assumptions are satisfied with probability very close to 1.

6.5.3. *Power series solutions of singular (q)-differential equations*

A. Bostan (Algorithms), M. F. I. Chowdhury (U. Western Ontario), R. Lebreton (Lix), B. Salvy, and É. Schost (U. Western Ontario) provided in [23] algorithms computing power series solutions of a large class of differential or q-differential equations or systems. Their number of arithmetic operations grows linearly with the precision, up to logarithmic terms.

6.5.4. *Fast computation of common left multiples of linear ordinary differential operators*

A. Bostan (Algorithms), F. Chyzak (Algorithms), Ziming Li (Chinese Academy of Sciences), and B. Salvy studied in [24] tight bounds and fast algorithms for LCLMs of several linear differential operators with polynomial coefficients. They analyzed the arithmetic complexity of existing algorithms for LCLMs, as well as the size of their outputs. They proposed a new algorithm that recasts the LCLM computation in a linear algebra problem on a polynomial matrix. This algorithm yields sharp bounds on the coefficient degrees of the LCLM, improving by one order of magnitude the best bounds obtained using previous algorithms. The complexity of the new algorithm is almost optimal, in the sense that it nearly matches the arithmetic size of the output.

6.5.5. *Space complexity of fast D-finite function evaluation*

M. Mezzarobba [41] showed that D-finite functions, i.e., solutions of linear differential equations with polynomial coefficients, can be evaluated in quasi-linear time and linear space with respect to the precision. In comparison, existing fast algorithms due to Chudnovsky and Chudnovsky and to van der Hoeven achieved the same time complexity with an overhead of a logarithmic factor in terms of memory usage.

6.5.6. *Multiple precision evaluation of the Airy function with reduced cancellation*

The series expansion at the origin of the Airy function $\text{Ai}(x)$ is alternating and hence problematic to evaluate for $x > 0$ due to cancellation. Based on a method recently proposed by Gawronski, Müller, and Reinhard, Sylvain Chevillard and Marc Mezzarobba [31] exhibit two functions F and G , both with nonnegative Taylor expansions at the origin, such that $\text{Ai}(x) = G(x)/F(x)$. The sums are now well-conditioned, but the Taylor coefficients of G turn out to obey an ill-conditioned three-term recurrence. They use the classical Miller algorithm to overcome this issue. They bound all errors and their implementation allows an arbitrary and certified accuracy, that can be used, e.g., for providing correct rounding in arbitrary precision.

6.5.7. Algorithms for combinatorial structures: well-founded systems and Newton iterations

C. Pivoteau (U. Marne-la-Vallée), B. Salvy, and M. Soria (UPMC) [21] considered systems of recursively defined combinatorial structures. They gave algorithms checking that these systems are well founded, computing generating series and providing numerical values. Their framework is an articulation of the constructible classes of Flajolet and Sedgewick with Joyal's species theory. They extend the implicit species theorem to structures of size zero. A quadratic iterative Newton method was shown to solve well-founded systems combinatorially. From there, truncations of the corresponding generating series were obtained in quasi-optimal complexity. This iteration transfers to a numerical scheme that converges unconditionally to the values of the generating series inside their disk of convergence. These results provide important subroutines in random generation. Finally, the approach was extended to combinatorial differential systems.

6.6. Euclidean Lattice Reduction and Applications

6.6.1. Lattice algorithms and hardness proofs

X.-W. Chang (McGill), D. Stehlé and G. Villard [17] proposed the first fully rigorous perturbation analysis of the R-factor of LLL-reduced matrices under column-wise perturbations. This study is very useful to devise LLL-type algorithms relying on floating-point approximations.

L. Luzzi (ENSEA), C. Ling (Imperial College) and D. Stehlé improved [20] the analyses of efficient Bounded Distance Decoding algorithms for lattices, and investigated the consequences for lattice-coded multiple-input multiple-output (MIMO) systems.

A. Langlois and D. Stehlé [54] introduced the Module-SIS and Module-LWE average-case lattice problems and reduced worst-case lattice problems to them. This provides a progressive transformation from the non-structured average-case lattices problems SIS and LWE, to the quite restricted but efficient average-case lattices problems Ring-SIS and Ring-LWE.

6.6.2. Cryptography

S. Ling (Nanyang Technological University, Singapore) and D. Stehlé [55] described the first public-key traitor tracing encryption scheme with security relying on the hardness of standard worst-case problems on Euclidean lattices.

J.-C. Belfiore (Telecom Paritech), L. Luzzi (ENSEA), C. Ling (Imperial College) and D. Stehlé [53] proved that nested lattice codes can achieve semantic security and strong secrecy over the Gaussian wiretap channel.

S. Ling (Nanyang Technological University, Singapore), K. Nguyen (NTU), H. Wang (NTU) and D. Stehlé [40] generalized Stern's zero-knowledge proof of knowledge protocol to obtain a statistical zero-knowledge proof of knowledge for the Inhomogeneous Small Integer Solution ISIS problem (in the infinity norm). This scheme is the first one that comes with no norm loss in the knowledge extraction procedure, leading to cryptographic constructions with tighter security proofs.

N. Attrapadung (AIST, Japan), J. Herranz (UPC, Spain), F. Laguillaumie, B. Libert (UCL, Belgium), E. de Panafieu (ENS Cachan), C. Ràfols (UPC, Spain) [15] proposed the first attribute-based encryption (ABE) schemes allowing for truly expressive access structures and with constant ciphertext size.

G. Castagnos (IMB) and F. Laguillaumie [38] gave a generic approach to design homomorphic encryption schemes, which extends Gjøsteen's framework. A specific scheme allows an arbitrary number of multiplications in the groups, as well as a pairing evaluation on the underlying plaintexts.

J. Herranz (UPC, Spain), F. Laguillaumie, B. Libert (UCL, Belgium) and C. Ràfols (URV, Catalonia) [34] proposed the first two attribute-based (for threshold predicates) signature schemes with constant size signatures. Their security is proven in the selective-predicate and adaptive-message setting, in the standard model, under chosen message attacks.

S. Canard (Orange Labs), G. Fuchsbauer (University of Bristol, UK), A. Gouget (Gemalto), F. Laguillaumie [30] defined a new cryptographic primitive called plaintext-checkable encryption, which extends public-key encryption by the following functionality: given a plaintext, a ciphertext and a public key, it is universally possible to check whether the ciphertext encrypts the plaintext under the key. They provide efficient generic random-oracle constructions based on any probabilistic or deterministic encryption scheme as well as a practical construction in the standard model.

6.7. Reliability and Accuracy

6.7.1. Standardization of interval arithmetic

We contributed to the creation in 2008 and N. Revol chairs the IEEE 1788 working group on the standardization of interval arithmetic <http://grouper.ieee.org/groups/1788/>. More than 140 persons from over 20 countries take part in the discussions, around 1500 messages were exchanged in 2012. We are currently voting on portions of the text of the standard and have good hope that the group will reach a final version of the standard within the allotted time. An extension has been granted for 2 more years, until December 2014.

The annual in-person meeting, chaired by N. Revol, took place at the end of the SCAN 2012 conference in Novosibirsk, Russia, the 28th of September. It was broadcasted via the Web and feedback was possible through e-mails. More than 20 persons attended the meeting.

V. Lefèvre participated in various discussions, either in the mailing-list or in small subgroups (he sent around 390 mail messages in 2012). He proposed a motion, which passed, on properties needed by number formats for operations between intervals and numbers (constructors, midpoint, etc.).

The latest discussions dealt with:

- flavors: even if there continues to be a give-and-take between proponents of a “small” standard involving just basic interval arithmetic and those who also want to also include the less common “modal arithmetic”, this motion about “flavors” intends to allow inclusion of modal interval arithmetic consistently and simply, possibly at a later stage or revision of the standard;
- expressions: what is regarded as an expression by P1788, the relation with the programming languages, what this implies concerning the allowed optimizations, etc.;
- decorations: what are the properties of functions we want to track along a computation, how the empty interval is handled, etc.;
- reproducibility: across several runs of a translated (e.g., compiled) program or across platforms, representation-independent behavior, reproducibility for parallel programs, etc.

A personal view of the current status of the work of the IEEE P1788 group and of directions for future work has been presented in [46], [45].

6.7.2. Interval matrix multiplication

Several formulas exist for the product of two intervals using the midpoint-radius representation: they trade off accuracy for efficiency. The use of these formulas for the product of matrices with interval coefficients allows to use BLAS3 routines and to benefit from their performances in terms of execution time [48]. The accuracy of these methods are studied in [42]. As it can be difficult to ensure that a prescribed rounding mode is actually in use, formulas that are oblivious to the rounding mode are developed [22]. The implementations of these variants on multicores are compared in [47].

6.7.3. Rigorous polynomial approximation using Taylor models in Coq

One of the most common and practical ways of representing a real function on machines is by using a polynomial approximation. It is then important to properly handle the error introduced by such an approximation. N. Brisebarre, M. Joldes (Uppsala Univ., Sweden), E. Martin-Dorel, M. Mayero, J.-M. Muller, I. Pasca, L. Rideau (Marelle), and L. Théry (Marelle) have worked on the problem of offering guaranteed error bounds for a specific kind of rigorous polynomial approximation called Taylor model [26]. They carry out this work in

the Coq proof assistant, with a special focus on genericity and efficiency for our implementation. They give an abstract interface for rigorous polynomial approximations, parameterized by the type of coefficients and the implementation of polynomials, and they instantiate this interface to the case of Taylor models with interval coefficients, while providing all the machinery for computing them.

COMPSYS Project-Team

6. New Results

6.1. Enhancing the Compilation of Synchronous Data-Flow Languages

Participants: Paul Feautrier, Abdoulaye Gamatié [LIFL], Laure Gonnord [Compsys/LIFL].

In [25] a new (light) numerical-Boolean abstraction was proposed for an efficient static analysis of synchronous programs that describe multi-clock embedded systems in the language Signal. In this abstraction, relations between clocks and numerical variables are modeled by Boolean-affine formulas. These formulas can easily be extracted from the program text. From the results of a satisfiability test of these formulas, clock properties can be deduced, which, when submitted to the Signal compiler, may improve the resulting target program.

In collaboration with Abdoulaye Gamatié, we proposed an extension of the previous approach to modular programs. This extension necessitates the use of an extended SMT (satisfiability modulo theory) solver – able for instance to deal with quantifier elimination – which has been implemented by Paul Feautrier by reusing some of the Syntol tools. This work is still unpublished but will be soon submitted to a journal.

6.2. Dataflow Analysis of Polyhedral X10 Programs

Participants: Paul Feautrier, Sanjay Rajopadhye [Colorado State Universty], Vijay Saraswat [IBM Research], Tomofumi Yuki [Colorado State University].

X10 is a recent parallel language, developed by IBM Research, whose aim is to increase programmers productivity. It is a descendant of Java and it includes several new parallel constructs, such as `async` and `finish`, which generalize `fork` and `join`, `clocks`, which generalize barriers, and `at`, which enables the remote execution of program fragments. X10 programs are guaranteed to be deadlock-free, but may exhibit non-deterministic behaviors or *races*.

We have devised a verifier for the `async/finish` fragment of X10 and polyhedral programs. The approach consists in computing the source of each array access. In the presence of parallel constructs, the sequencing predicate is no longer a total order and a read may have several sources, which indicates a race. A proof-of-concept tool has been implemented. This work will be presented at the next *Principles and Practice of Parallel Programming* conference (PPoPP'13 in Shenzhen, China) [13].

6.3. Data-Aware Process Networks

Participants: Christophe Alias, Alexandru Plesco [Compsys/Zettice].

New techniques were introduced to generate and compile optimized data-aware process networks, from a C program annotated with pragmas (see Section 5.9 and the software tool Dcc). These techniques are essential for the Zettice start-up and are not made publicly available for the moment.

6.4. Optimizing Remote Accesses for HLS

Participants: Christophe Alias, Alain Darte, Alexandru Plesco [Compsys/Zettice].

Some data- and compute-intensive applications can be accelerated by offloading portions of codes to platforms such as GPGPUs or FPGAs. However, to get high performance for these kernels, it is mandatory to restructure the application, to generate adequate communication mechanisms for the transfer of remote data, and to make good usage of the memory bandwidth. In the context of the high-level synthesis (HLS), from a C program, of hardware accelerators on FPGA, we showed how to automatically generate optimized remote accesses for an accelerator communicating to an external DDR memory. Loop tiling is used to enable block communications, suitable for DDR memories. Pipelined communication processes are generated to overlap communications and computations, thereby hiding some latencies, in a way similar to double buffering. Finally, not only intra-tile but also inter-tile data reuse is exploited to avoid remote accesses when data are already available in the local memory.

We showed how to generate the sets of data to be read from (resp. written to) the external memory just before (resp. after) each tile so as to reduce communications and reuse data as much as possible in the accelerator. The main difficulty arises when some data may be (re)defined in the accelerator and should be kept locally. We proposed an automatic optimized code generation scheme, entirely at source-level, i.e., in C, that allows us to compile all the necessary glue (the communication processes) with the same HLS tool as for the computation kernel. Our method, implemented in the tool Chuba (see Section 5.7) uses advanced polyhedral techniques for program analysis and transformation. Experiments with Altera HLS tools demonstrate how to use our techniques to efficiently map C kernels to FPGA.

This work, astride two different fields (compilation for high-performance computing and high-level synthesis) turned out to be very difficult to publish. It was finally accepted at PPOPP'12 [6], but only as a short paper (2 pages). We requested to retain the copyright of this work to be able to publish a longer version. It was accepted at the IMPACT'12 workshop [7], which makes paper available on the web, but with no copyright. It was finally accepted as a full publication at the DATE'13 conference [8].

6.5. Parametric Inter-Tile Reuse for Kernel Offloading

Participants: Alain Darte, Alexandre Isoard.

The method described in Section 6.4 is not parametric in terms of the tile size, i.e., the tile size needs to be fixed before compiling the program. Furthermore, the size of the required local memory depends on the tile size and is available only after program analysis. As a result, to select the tile size with respect to the size of the local memory, the program first needs to be compiled (actually analyzed) for all tile sizes. A parametric program analysis would be much more convenient. The situation is even worse to get the runtime performances in terms of the tile size. Indeed, so far, Chuba generates a C code that is generic and cannot be immediately compiled by C2H. A few modifications by hand are still needed, such as inserting the adequate pragmas for C2H, transforming array accesses to linearized addresses with the right base addresses, changing some arrays into non-aliasing pointers so that C2H, whose dependence analyzer and software pipeliner are weak, can generate codes with the right initiation intervals, etc. These changes are minor and systematic and take time when performed by hand. A fully-parametric compilation scheme would be a plus.

The formulation proposed in Section 6.4 is unfortunately quadratic in terms of the tile size, which prevents to parameterize it. Indeed, it relies on parametric linear programming, which works only with a linear use of parameters. As part of the Master internship of Alexandre Isoard, we nevertheless succeeded to design a fully-parametric scheme for inter-tile reuse and buffer size computation. The method is much more involved but is still compatible with approximations. These results have still to be implemented and submitted for publication.

6.6. Semantic Program Transformations

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

Traditionally, a program transformation is considered to be *correct* if each data dependence of the original program is respected. In that case, both original and transformed programs perform *exactly* the same computation. We can relax this condition by expecting both programs to perform the same computation, modulo the *semantic* properties of the operators (e.g., associativity, commutativity). Semantic program transformations extend the traditional corpus of program transformations and can reveal new optimization opportunities.

More specifically, we are interested in *semantic loop tiling*, a special case of loop tiling, where the input arrays are tiled, and the program is restructured to use high-level matrix operations between data tiles, instead of the original scalar operations. Surprisingly, it turns out that in most cases, the semantic tiling is simply obtained by substituting the scalar variables by the tiles (matrices), and the original operators by the corresponding matrix operators (e.g., a/b by $\text{MatMul}(A, \text{Inv}(B))$). The approach currently investigated consists in two steps: (i) guess the semantic tiling, and (ii) prove the (semantic) equivalence with the original program.

Our current contribution is an heuristic to check the equivalence of two programs modulo associativity/commutativity so as to achieve the step (ii). The two programs should fit in the polyhedral model but can involve explicit reductions. This work is currently under submission, and is part of the PhD thesis of Guillaume Iooss.

6.7. Modular Termination of Large Programs

Participants: Christophe Alias, Guillaume Andrieu [LIFL], Laure Gonnord [Compsys/LIFL].

Program termination is an essential step in program verification. In [16], we showed how to check the termination of programs whose control can be summarized by an integer interpreted automaton. This was done by computing a *ranking function* (kind of schedule) by means of integer linear programming techniques. This approach, though powerful, clearly lacks scalability and cannot handle large programs.

We overcame this limitation by proceeding into two steps. First, we extract, from the program to be analyzed, the part useful for termination, i.e., the smaller program slice with the same control behavior. Then, we show that proving the termination of the whole program (slice) boils down to prove the termination of small programs, which can be handled by the technique of [16]. Experimental results show that many large programs can be handled this way.

This work was part of the engineer internship of Guillaume Andrieu. Our technique has been implemented in a tool called SToP (see Section 5.12) and presented at the workshop TAPAS'12 [9].

6.8. Lower Bounds for the Inherent Data Locality Properties of Computations

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

Data movement will account for most of the energy as well as execution time on upcoming exascale architectures, including data movement between processors as well as data movement across the memory hierarchy within each processor. Therefore a fundamental characterization of the data access complexity of algorithms is increasingly important.

We addressed the problem of data access or I/O complexity in a two-level memory hierarchy, as studied in the seminal work of Hong and Kung [26]. We developed a novel approach based on graph min-cut for deriving lower bounds on I/O complexity with two significant advantages over the S-partitioning model of Hong and Kung: (1) the approach can be used to develop analytical expressions with tighter lower bounds for I/O, and (2) unlike any previous model, our new lower bound approach can be automated for analyzing an arbitrary computational directed acyclic graph. We show tighter analytically-derived lower bounds as well as very promising experimental results thanks to a prototype tool that implements our fully-automated analysis.

This work has been submitted and is part of an informal collaboration with P. Sadayappan from the University of Columbus (CSU).

6.9. A Polynomial Spilling Heuristic: Layered Allocation

Participants: Albert Cohen [Inria, Parkas], Boubacar Diouf [Université Paris Sud, Parkas], Fabrice Rastello.

Register allocation is subdivided into two sub-problems: first, the *allocation* (or its dual problem the *spilling*) selects the set of variables that will reside in registers (resp. in memory) at each point of the program. Then, the *assignment* or *coloring* picks a specific register where a variable will reside. Building on some properties of the static single assignment form (SSA), it is now possible to decouple the allocation from the assignment. Indeed, the interference graph of a program in SSA form is a chordal graph. In this context, MAXLIVE, the maximal number of variables simultaneously live at a program point, is used during the spilling phase as a criterion to guarantee that the forthcoming assignment will be performed without any spill. If MAXLIVE is lower than or equal to R , the number of available registers, then all the variables will be assigned without any spill. This *decoupled* approach was advocated by Fabri, Appel and George, Darté et al., and others.

Existing spilling heuristics rely on a sufficient condition to guarantee register assignment, and incrementally spill until the condition holds. As we just mentioned, for programs under SSA, the condition is necessary and sufficient: `MAXLIVE` has to be lower than or equal to R . Incremental spilling decisions to satisfy this condition tend to be overly local and suboptimal. Indeed, incremental spilling itself is NP-complete, and heuristics based upon it trade too much their optimality for polynomiality. In contrast to incremental spilling, we proposed to adopt the symmetric approach: incremental allocation. Intuition for it emerges from two observations allowing for more global spilling decisions:

1. Register allocation is pseudo-polynomial in the number of registers, suggesting a heuristic that solves (optimally) roughly $R/step$ allocation problems on $step$ registers each. The final allocation is the layered composition of the stepwise allocations.
2. Stepwise optimality does not guarantee an overall optimal allocation, but experiments show that it comes very close to optimal, even with $step = 1$. Intuition for this comes from recent work by Diouf et al., observing that allocation decisions tend to be a monotonic function of the number of registers.

This work, which will be presented at CGO'13 [11], proposes a new graph-based allocation heuristic, based on a maximum clique cover formulation to define the profitability of spilling variables. It exploits the pseudo-polynomial complexity in the number of registers of the allocation problem under SSA — as opposed to the symmetric, spilling problem, which remains strongly NP-complete. It addresses the spill-everywhere problem in a decoupled context and also proposes an extension to non-decoupled approaches. It introduces *layered allocation* a new strategy that incrementally allocates variables instead of incrementally spilling variables. The evaluation performed on standard benchmarks shows that this new approach is near-optimal.

6.10. Interaction Between Spilling and Scheduling

Participants: Quentin Colombet, Alain Darte, Fabrice Rastello.

As explained in Section 6.9, it is possible to decouple the register allocation problem in two successive phases: a first *spilling* phase places `load` and `store` instructions so that the register pressure at all program points is small enough, a second *assignment* and *coalescing* phase maps the remaining variables to physical registers and reduces the number of move instructions among registers. At CASES'11 [18], we presented a new integer linear programming (ILP) formulation, for load-store architectures, to capture “optimal” spilling in a more accurate and more expressive way than previous approaches. Among other features, we can express SSA ϕ -functions, memory-to-memory copies, and the fact that a value can be stored simultaneously in a register and in memory.

We used this ILP formulation to experimentally analyze the impact of the different heuristic strategies and compare them with optimal solutions. While “optimal” solutions show significant improvements for static spill costs, it turned out that runtime performances were disappointing (if not random). We conducted various experiments to understand this behavior and discovered that the interaction with scheduling is actually higher than expected. Micro-architectural features (e.g., memory latencies that can be hidden by prefetching, bundling that can hide cycles) have to be accounted for in the model, which is never done. These experiments and analysis are described in Chapter 4 of Quentin Colombet's PhD thesis [1].

6.11. Elimination of Parallel Copies Using Code Motion on Data Dependence Graphs

Participants: Florian Brandner, Quentin Colombet.

Traditional approaches to copy elimination during register allocation are based on interference graphs and register coalescing. Variables are represented as nodes in a graph, which are coalesced, if they can be assigned the same register. However, decoupled approaches strive to avoid interference graphs and thus often resort to local recoloring.

A common assumption of existing coalescing and recoloring approaches is that the original ordering of the instructions in the program is not changed. We developed an extension of a local recoloring technique called Parallel Copy Motion. We perform code motion on data dependence graphs in order to eliminate useless copies and reorder instructions, while at the same time a valid register assignment is preserved. Our results show that even after traditional register allocation with coalescing our technique is able to eliminate an additional 3% (up to 9%) of the remaining copies and reduce the weighted costs of register copies by up to 25% for the SPECINT 2000 benchmarks. In comparison to Parallel Copy Motion, our technique removes 11% (up to 20%) more copies and up to 39% more of the copy costs.

These results have been accepted for publication at SAC'12 [10] and, in a longer version, in the journal *Computer Languages, Systems, and Structures* [5].

CONVECS Team

6. New Results

6.1. New Formal Languages and their Concurrent 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 the CONVECS team for industrial case studies and applications (see § 6.5) and serves also in university courses on concurrency, in particular at ENSIMAG (Grenoble) and at the 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 2012, in addition to 12 bug fixes, the following enhancements have been brought to these tools:

- We improved the ergonomics of the LNT2LOTOS translator by refining certain command-line options and by making some warning messages more user-friendly.
- We optimized the generated LOTOS code of the “disrupt” and “parallel” composition operators, so as to reduce the number of spurious warnings about impossible synchronizations and, more importantly, to meet the subset of LOTOS supported by the CAESAR compiler (static bound on the number of parallel processes).
- We improved the support for LNT programs that contain several modules by allowing the main process to be defined in any module (not only in the main module).
- We added new predefined functions for the generic data types (lists, sorted lists, and sets), and we updated accordingly the reference manual of LNT. The set types are now implemented correctly by avoiding duplicate elements.

6.1.2. Distributed Code Generation for Process Algebras

Participants: Hugues Evrard, Frédéric Lang.

One goal of CONVECS is to build a tool that generates automatically a distributed implementation of a system specified in LNT. This requires a protocol to realize process synchronization. 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 2012, we explored the bibliography on synchronization protocols. Among almost twenty references studied, we selected three existing distributed synchronization protocols that seemed appropriate to our problem. In order to validate these protocols, we designed a tool chain that, given a system described as a parallel composition of LNT processes, generates an LNT specification of an implementation of the system (called the *implementation model*), by incorporating the protocol in the specification to realize the synchronizations. We then used CADP to check for livelocks and deadlocks possibly introduced in the implementation model by the protocol (using MCL and EVALUATOR 4.0), and to verify that the implementation model mimics the behaviour of the system by equivalence checking (using BISIMULATOR).

Among the three protocols considered, we selected the most promising one [57], which is suitable for generalization to implement synchronization vectors (and hence, the generalized parallel composition operator of LNT). Using the methodology mentioned above, we discovered a previously unknown error in this protocol, which leads to deadlocks in certain situations, and we proposed a correction. An article has been submitted to an international conference.

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. So far, only a few verification tools have been designed for analyzing π -calculus specifications automatically. Our objective is to provide analysis features for π -calculus specifications by reusing the verification technology already available for value-passing process algebras without mobility. Our approach is based on a novel translation from the finite control fragment of π -calculus to LNT. To the best of our knowledge, this is the first π -calculus translation having a standard process algebra as target language.

In this work, we have also extended the original polyadic π -calculus with data-handling features. This results in a general-purpose applied π -calculus, which offers a good level of expressiveness for specifying mobile concurrent systems, and therefore for widening its possible application domains. As language for describing data types and functions, a natural choice was LNT itself: in this way, the data types and functions used in the π -calculus specification can be directly imported into the LNT code produced by translation.

The translation is fully automated by the tool PIC2LNT 2.0. This enables the analysis of applied π -calculus specifications using all verification tools of CADP, in particular the EVALUATOR 4.0 on-the-fly model checker, which evaluates temporal properties involving channel names and data values. PIC2LNT 2.0 was used for teaching mobile concurrency at Saarland University. A paper describing this work was accepted for publication 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 have considered a translation from the EB3 language [39] for information systems to LNT. EB3 is inspired from a process algebra, but has the particularity to contain so-called *attribute functions*, whose semantics depend on the history of events. Therefore, the history of events becomes part of the state of an EB3 specification, which is unusual in process algebras.

Since EB3 is not equipped with native verification tools, we have proposed a translation from EB3 to LNT, which would enable EB3 specifications to be formally verified using CADP. Our formal translation scheme ensures the strong equivalence between the LTS corresponding to an EB3 specification and the LTS corresponding to the LNT code generated. The history of events is encoded as a particular LNT process “*memory*” synchronized on all EB3 events with the rest of the system. The memory process thus acts as a monitor that changes its state according to the occurring events and answers requests emitted by the attribute functions when needed. A prototype translator has been developed at University Paris-Est Créteil and a paper describing this work has been submitted to an international conference.

6.1.5. Coverage Analysis for LNT

Participants: Gwen Salaün, Lina Ye.

In the classic verification setting, we have an LNT specification of a system, a set of temporal properties to be verified on the LTS model corresponding to the LNT specification, and a data set of examples (test cases) we use 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.

Coverage analysis aims at proposing and developing techniques for automatically detecting parts of an LNT specification not (yet) covered during verification. Such LNT coverage analysis techniques would be very helpful for (i) extending the set of test cases with new inputs covering parts of the LNT specification that have not been analyzed yet, (ii) eliminating dead code in the LNT specification, and (iii) extending the set of temporal properties with new ones.

We have already identified four criteria (action, decision, block, property) and developed a prototype tool that automatically returns coverage values for these four criteria. We have applied our tool to LNT specifications of existing protocols, such as a reconfiguration protocol for component-based architectures [34], and found several cases of dead code and missing test cases.

6.1.6. Other Software Developments

Participants: Hubert Garavel, Frédéric Lang, Wendelin Serwe.

In addition of correcting 23 bugs in various CADP tools, we also brought the following enhancements:

- The EUCALYPTUS interface was improved regarding ergonomics and customization.
- The CADP tools for 32-bit and 64-bit Intel/Linux architectures were upgraded to use recent compilers and libraries, and CADP was modified to support Mac OS X 10.8 “Mountain Lion”.
- The usability of the libraries for writing BCG files was improved to detect and signal an improper ordering of the primitives in application programs.
- The SYNTAX parser generator was improved by correcting two subtle errors, one of them causing an infinite looping on certain erroneous input programs. The CADP compilers developed using SYNTAX were enhanced to perform a better diagnosis of the situations when SYNTAX corrected syntactic errors automatically in erroneous programs.
- We improved an optimization of the CAESAR compiler for LOTOS, leading to a significant reduction of the execution time (from one hour and 51 minutes down to 58 seconds) for some examples of LOTOS programs with many variables. We optimized the CAESAR.OPEN script to invoke the CAESAR compiler directly whenever possible (instead of the GENERATOR tool), which improves the performance of graph generation, in particular for LNT.OPEN.
- Four demonstration examples of CADP were extended with LNT descriptions to illustrate the usage of the LNT language and of its compiler. Two examples have been simplified using the latest features of SVL, which can now handle the “ n among m ” parallel composition operator of LNT. Also, three examples have been reorganized for a better clarity and two couples of examples, which were closely related, have been merged into single examples.

6.2. Parallel and Distributed Verification

Participants: Hubert Garavel, Radu Mateescu, Wendelin Serwe.

For distributed verification, CADP provides the PBG format, which implements the theoretical concept of *Partitioned LTS* introduced in [46] and provides a unified access to an LTS distributed over a set of remote machines. The PBG format is equipped with the DISTRIBUTOR and PBG_MERGE (previously called BCG_MERGE [45]) tools, which perform the distributed generation of a partitioned LTS and the conversion of a partitioned LTS represented in the PBG format into a monolithic LTS stored in a BCG file.

To facilitate the manipulation of partitioned LTSs, CADP provides the PBG_CP, PBG_MV, and PBG_RM tools for copying, moving, and removing PBG files, maintaining consistency during these operations. The PBG_INFO tool provides several functionalities to inspect PBG files, such as checking consistency (i.e., existence and readability of all fragment files), calculating the size (number of states and transitions) of the corresponding LTS, displaying the list of labels, and concatenating remote log files (this is useful, e.g., to understand the reason why a PBG generation fails, and to compute global statistics about CPU and memory usage by the worker processes).

In 2012, in addition to correcting two bugs in DISTRIBUTOR and several bugs in the CAESAR_NETWORK_1 communication library used by the distributed verification tools, we also improved these tools as follows:

- We enhanced DISTRIBUTOR to support more than 256 distributed processes.
- We enhanced CAESAR_NETWORK_1 with a debugging facility, which enables traces of all distributed processes to be generated.
- We enhanced the graphical monitor of DISTRIBUTOR with the option of sorting the labels alphabetically, which facilitates their visual inspection.
- We extended PBG_INFO to enable the display of all labels in a partitioned LTS.

We also developed a prototype tool, named 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.

We experimented all these tools on the Grid'5000 computing infrastructure [35] 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 [12].

6.3. Timed, Probabilistic, and Stochastic Extensions

Participant: Hubert Garavel.

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 2012, we undertook a study to merge probabilistic and stochastic aspects into modern high-level languages such as LNT. This work is done at Saarland University under the aegis of the Alexander von Humboldt foundation, in collaboration also with RWTH Aachen and Oxford University. We investigated the theoretical concepts, as well as their integration into modeling languages, together with the corresponding behavioural equivalences and temporal logics.

We also started experimenting with state-of-the-art software implementations, such as MODEST and PASS (Saarland University), COMPASS and MRMC (RWTH Aachen), and PRISM (Oxford University). Two of these tools (namely, MODEST and PRISM) have been used for lab exercises in the *Applied Concurrency Theory* block course created by H. Garavel at Saarland University. Following these experiments, evaluation reports have been produced, which provide feedback about issues and suggestions for enhancements. These reports have been addressed to the respective authors of each tool and already led to improvements in certain tools.

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 [29]. 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 2012, we have continued the development of our prototype tools for partial model checking of the regular alternation-free μ -calculus supporting all features of the input language of EXP.OPEN 2.1. We have also extended our work to handle useful fairness operators of alternation depth 2 in linear time, without developing the complex machinery that would be necessary to evaluate general μ -calculus formulas of alternation depth 2. A paper has been published in an international conference [15] and an extended version has been submitted to an international journal.

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 a (hardware) implementation is conform to a formal model written in LNT. Our approach is inspired by the theory of conformance testing [59], as implemented for instance in TGV [51] and JTorX [33].

We developed two prototype tools supporting conformance testing. The first tool implements a dedicated OPEN/CAESAR-compliant compiler for the particular asymmetric synchronous product of the model and the test purpose. This tool is a generic component for on-the-fly graph manipulation, taking as input two graphs and producing as output the graph of the asymmetric synchronous product. 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. This tool was built from (slightly extended) existing generic components for on-the-fly graph manipulation (τ -compression and τ -confluence reductions, determinization, resolution of Boolean equation systems). The main advantage of our approach compared to existing tools is the use of LNT for test purposes, which facilitates the manipulation of data values.

6.5. Real-Life Applications and Case Studies

6.5.1. ACE Cache Coherency Protocol

Participants: Hubert Garavel, Abderahman Kriouile, Radu Mateescu, Wendelin Serwe.

In the context of a CIFRE convention with STMicroelectronics (see § 7.1), we studied the 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 validation techniques grows exponentially. As an alternative, we study formal verification.

In 2012, 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 2600 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, including an abstract operation that represents any other ACE operation. Using the OCIS simulator, we were able to explore the behavior of the system interactively, which has been found helpful by STMicroelectronics engineers.

Currently, our formal model supports a representative subset of five elementary operations of the ACE protocol (MakeUnique, ReadOnce, ReadShared, ReadUnique, and WriteBack). For each of these operations, we have written a liveness property in MCL expressing that the operation is executed until its termination. Using parametric SVL scripts (about 250 lines) and the EVALUATOR 4.0 model checker, we verified these properties on the fly for up to three memory lines and two processors with two cache lines each. We also generated the corresponding LTS (up to 250 million states and one billion transitions).

We also started considering data integrity properties. This required to translate a state-based property (namely, the consistency between the values stored in memory and in the local caches of the processors) into our action-based setting. This enabled us to automatically exhibit a known error present in a previous version of the ARM specification of the ACE protocol (which was corrected in a subsequent version of the specification). Using the LNT model corresponding to the latest version of the ACE specification, we spotted several potential data integrity issues that we reported to STMicroelectronics, where they are currently under investigation.

6.5.2. *Realizability of Choreographies*

Participants: Alexandre Dumont, Matthias Gdemann, Gwen Salan.

Choreographies allow business and service architects to specify, with a global perspective, the requirements of applications built over distributed and interacting software entities. In collaboration with Pascal Poizat (University Paris-Sud), we proposed new techniques for verifying BPMN 2.0 choreographies, and particularly the *realizability* property. Realizability ensures that peers obtained via projection from a choreography interact as prescribed in the choreography requirements. Our approach is formally grounded on a model transformation into the LNT process algebra and the use of equivalence checking. It is also completely tool-supported through interaction with the Eclipse BPMN2 modeler and CADP. These results have been published in an international conference [17].

In collaboration with Meriem Ouederni (University of Nantes), we extended our techniques for analyzing choreographies to restore realizability for non-realizable, but *repairable* choreographies. For this we exploit the counterexamples generated by the equivalence checker BISIMULATOR to identify problematic messages in the choreography. For those messages we add distributed local monitors to the system which delay message sending if necessary, to restore correct message sequences. This iterative approach introduces the minimal number of necessary additional messages to restore realizability, and the monitors are generated in the most permissive way, i.e., by considering all possible interleavings given the behaviour of the peers participating to the choreography. It is fully automated by a prototype tool we implemented. These results have been published in an international conference [14].

We developed a common formal description language, named CIF (*Choreography Intermediate Format*), for the verification of choreographies. CIF is based on an XML representation for easy exchange between programs, an XSD schema for validation, and a translation to LNT for verification. CIF is used as an intermediate language to specify choreographies, but can also serve as target language for translating various choreography specification languages, such as BPMN 2.0. The back-end connection to CADP via LNT enables the automation of some key choreography analysis tasks (repairability, realizability, conformance, etc.). Our framework is extensible with other front-end and back-end connections to, respectively, other choreography languages and formal verification tools.

6.5.3. *Self-Configuration Protocol for Distributed Cloud Applications*

Participants: Rim Abid, Gwen Salan.

We collaborate with Nol de Palma and Fabienne Boyer (University Joseph Fourier), Xavier Etchevers and Thierry Coupaye (Orange Labs) in the field of cloud computing applications, which are complex, distributed artifacts involving multiple software components running on separate virtual machines. Setting up, (re)configuring, and monitoring these applications are complicated tasks because a software application may depend on several remote software and virtual machine configurations. These management tasks involve many complex protocols, which fully automate these tasks while preserving application consistency as well as some key properties.

In this work, we focus on a self-configuration protocol, which is able to configure a whole distributed application without requiring any centralized server. The high degree of parallelism involved in this protocol makes its design complicated and error-prone. In order to check that this protocol works as expected, we specify it in LNT and verify it using the CADP toolbox. The use of these formal techniques and tools helped to detect a bug in the protocol, and served as a workbench to experiment with several possible communication models. These results led to a publication in an international conference [18].

We are currently studying two variants of the self-configuration protocol, one handling virtual machine failures, and one allowing dynamicity in the system (addition and removal of virtual machines) using a publish-subscribe communication framework.

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 (to be defined within the project) 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, namely model checking and equivalence checking.

In 2012, in interaction with Crouzet engineers, we studied the PLC language of Crouzet to understand precisely its static and dynamic semantics. We specified manually in LNT several examples of control applications provided by Crouzet, with the goal of identifying the principles of translating the PLC language of Crouzet to LNT. We formulated in MCL several safety and liveness properties concerning the temporal ordering of input and output events by the control applications, and we successfully verified them on the LNT specifications. We also started to study the network communication mechanisms between PLCs to identify a suitable LNT abstraction of the communication layer.

6.5.5. Other Case Studies

Participants: Frédéric Lang, Radu Mateescu, Wendelin Serwe.

Continuing a study [53] started in the context of the Multival project (see <http://vasy.inria.fr/multival>), we considered the Platform 2012 architecture proposed by STMicroelectronics, focusing on the Dynamic Task Dispatcher (DTD), a hardware block that assigns a set of application tasks to a set of processors. In 2012, we extended our LNT model and the corresponding MCL properties in order to handle heterogeneous processors equipped with different kinds of processor extensions. We also used constraints on the initialization phase, which reduced the size of the LTS by a factor of up to ten and hence enabled the generation of the LTS for up to eight processors (instead of only six). Both extensions together enabled to discover the possibility of a livelock.

We attempted to investigate this issue further by cosimulation (using the EXEC/CAESAR framework) with the original C++ model of the architect. Unfortunately, the C++ model did not behave correctly for the particular aforementioned application scenario. It was not possible to change this model because the recent evolutions of Platform 2012 excluded the DTD, as its requirements in terms of silicon surface were considered too large. This work, including the LNT model as appendix, has been accepted for publication in an international journal [5].

In collaboration with Nuno Mendes and Claudine Chaouiya (Gulbenkian Institute, Portugal), Yves-Stan Le Cornec (IBISC, University Evry Val d'Essonne) and Grégory Batt (CONTRAINTE project-team, Inria Paris-Rocquencourt), we have studied the use of CADP for checking the reachability of stable states in genetic regulatory networks. A compositional and logical model of genetic regulatory networks called *logical regulatory modules* was defined and translated to LNT processes and EXP.OPEN 2.1 networks of LTSs.

Compositional minimization modulo safety equivalence was applied to the generated network, so as to palliate state explosion while preserving the reachability property. The approach has been illustrated on the segment polarity module involved in the segmentation of the fruit fly embryo and on the delta-notch module involved in cell differentiation in crucial steps of embryonic development of several species. A paper has been submitted to an international journal.

POP ART Project-Team

6. New Results

6.1. Dependable Distributed Real-time Embedded Systems

Participants: Gwenaël Delaval, Pascal Fradet, Alain Girault [contact person], Emil Dumitrescu.

6.1.1. Tradeoff exploration between reliability, power consumption, and execution time

For autonomous critical real-time embedded systems (*e.g.*, satellite), 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 tricriteria 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 [17], [11]. 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. The new contribution for 2012 has been the formulation of a new multi-criteria cost function for our ready list scheduling heuristics, such that we are able to prove rigorously that the static schedules we generate meet both the reliability constraint and the power consumption constraint.

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 the ILP. Secondly, we have compared TSH versus the ECS heuristic (Energy-Conscious Scheduling [68]); 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. Controller Synthesis for the Safe Design of Embedded Systems

Participants: Gwenaël Delaval [contact person], Gregor Goessler, Sebti Mouelhi.

6.2.1. 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.*, [73] 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 [71]. 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 [54].

These notions enabled the computation of approximately equivalent discrete abstractions for several classes of dynamical systems, including nonlinear control systems with or without disturbances, and switched systems. These approaches are based on sampling of time and space where the sampling parameters must satisfy some relation in order to obtain abstractions of a prescribed precision. In particular, the smaller the time sampling parameter, the finer the lattice used for approximating the state-space; this may result in abstractions with a very large number of states when the sampling period is small. However, there are a number of applications where sampling has to be fast; though this is generally necessary only on a small part of the state-space.

In [45] 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 extended the approach of [45] to the synthesis of controllers for time-bounded reachability. Furthermore we have implemented the algorithms for safety and time-bounded reachability in COSYMA, a tool for automatic controller synthesis for incrementally stable switched systems based on multi-scale discrete abstractions. 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.

We are currently exploring, in the SYMBAD project, controller synthesis for switched systems based on a different approach for the construction of multi-scale abstractions. The goal is to further improve the trade-off between cost and precision.

6.2.2. Modular discrete controller synthesis

Discrete controller synthesis (DCS) [71] allows to design programs in a mixed imperative/declarative way. From a program with some freedom degrees left by the programmer (*e.g.*, free controllable variables), and a temporal property to enforce which is not *a priori* verified by the initial program, DCS tools compute off-line automatically a *controller* which will constrain the program (by *e.g.*, giving values to controllable variables) such that, whatever the values of inputs from the environment, the *controlled program* satisfies the temporal property.

Our motivation *w.r.t.* DCS concerns its modular application, improving the scalability of the technique by using contract enforcement and abstraction of components. Moreover, our aim is to integrate DCS into a compilation chain, and thereby improve its usability by programmers, not experts in discrete control. This work has been implemented into the HEPTAGON/BZR language and compiler [50]. This work is done in collaboration with Hervé Marchand (VERTECS team from Rennes) and Eric Rutten (SARDES team from Grenoble).

The implemented tool allows the generation of the synthesized controller under the form of an HEPTAGON node, which can in turn be analyzed and compiled, together with the HEPTAGON source from which it has been generated. This full integration allows this method to aim different target languages (currently C, JAVA or VHDL), and its integrated use in different contexts.

A formal semantics of BZR has been defined, taking into account its underlying nondeterminism related to the presence of controllable variables. A new implementation has been achieved, including an abstraction method based on [47]. We have used BZR for demonstrating the use of Control Theory and Techniques to the administration of computing systems in a closed-loop management [19].

6.3. Automatic Distribution of Synchronous Programs

Participants: Gwenaël Delaval [contact person], Alain Girault, Gregor Goessler, Xavier Nicollin, Gideon Smeding.

6.3.1. 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 is conducted 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 [38]). 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 [48], [49], 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, which 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).

6.3.2. Distribution of synchronous programs under real-time constraints

With the objective to distribute synchronous data-flow programs (*e.g.*, LUSTRE) over GALS architectures, such that the difference between the original and synchronous systems satisfy given bounds, we have developed a quantitative clock calculus to (1) describe timing properties of the architecture's clock domain, and (2) describe the acceptable difference between the original and distributed programs. The clock calculus is inspired by the network calculus [67], with the difference that clocks are described only with respect to one-another, not with respect to real-time.

As a first result, we have applied our clock calculus to analyze the properties of periodic synchronous data-flow programs executed on a network of processors. Because our clock calculus is relational, it can model and preserve correlated variations of streams. In particular, the common case of a data-flow system that splits a stream for separate treatment, and joins them afterwards, this analysis yields more precise result than comparable methods [24].

We have been able to use the clock calculus as an abstract domain to perform abstract interpretation of synchronous boolean data-flow programs and their distribution on synchronous nodes that communicate asynchronously by sampling shared memory. The analysis discovers the relative clock drift of all clocks of the distributed system as well as bounds on the distance from the original program.

In case the guaranteed maximal distance is too large, we provide methods to synthesize bounds on the relative drift of the architecture's clocks that ensure an acceptable distance. Given the synthesized bounds, we use the known clock drifts and program behavior to synthesize light weight protocols.

6.4. New Programming Languages for Embedded Systems

Participants: Alain Girault [contact person], Pascal Fradet, Vagelis Bebelis, Bertrand Jeannet, Peter Schrammel.

6.4.1. 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, 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.

We have introduced the *schedulable parametric data-flow (SPDF)* MoC for dynamic streaming applications [20]. SPDF extends the standard dataflow model by allowing rates to be parametric (*e.g.*, of the form $2xy$, where x and y are parameters, the value of which can change at run-time). SPDF was designed to be statically analyzable while retaining sufficient expressive power. We formulated sufficient and general static criteria for boundedness and liveness. In SPDF, parameters can be changed dynamically even within iterations. The safety of dynamic parameter changes can be checked and their implementation made explicit in the graph. These different analyses are made possible using well-defined static operations on symbolic expressions. The same holds for quasi-static scheduling which is the first step towards code generation for multi-core systems.

We are now focusing on the parallel scheduling of parametric dataflow models. We have proposed a generic and flexible framework to generate parallel ASAP schedules targeted to the new STHORM many-core platform designed by STMicroelectronics. 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. Our case studies are video decoders for high definition video streaming such as VC-1 or HEVC.

This research is the central topic of Vagelis Bebelis' PhD thesis. It is conducted in collaboration with STMicroelectronics.

6.5. Static Analysis and Abstract Interpretation

Participants: Alain Girault, Bertrand Jeannot [contact person], Peter Schrammel.

6.5.1. *Translating data-flow languages for hybrid systems simulation to hybrid automata for hybrid systems verification*

Hybrid systems are used to model embedded computing systems interacting with their physical environment. There is a conceptual mismatch between high-level hybrid system languages like SIMULINK³⁴, which are used for simulation, and hybrid automata, the most suitable representation for safety verification. Indeed, in simulation languages the interaction between discrete and continuous execution steps is specified using the concept of zero-crossings, whereas hybrid automata exploit the notion of staying conditions.

In the context of the INRIA large scale action SYNCHRONICS (see §8.1.1.1), we studied how to translate the ZELUS hybrid data-flow language [34] developed in this project into logico-numerical hybrid automata by carefully pointing out this issue. We investigated various zero-crossing semantics, proposed a sound translation, and discussed to which extent the original semantics is preserved.

This work is part of the PhD thesis of Peter Schrammel and was presented at the conference HSCC'2012 (Hybrid Systems: Computation and Control) [22], [27].

6.5.2. *Abstract Acceleration of general linear loops*

We investigated abstract acceleration techniques for computing loop invariants for linear loops with linear assignments in their body and guards defined by the conjunction of linear inequalities.

While standard abstract interpretation considers over approximations over the set of reachable states at any loop iteration, and relies on extrapolation (*a.k.a.* widening) for making the analysis converge, abstract acceleration captures the effect of the loop with a single, non-iterative transfer function applied to the reachable states at the loop head. The concept of abstract acceleration has already been applied to restricted form of linear loops, by us [16] and others [58], and extended to logico-numerical loops [16]; the novelty here is to investigate general linear loops.

³⁴<http://www.mathworks.com>

The main idea we developed is to over-approximate the set of transformation matrices associated to any number of iterations of the loop body and to apply this “accelerated” transformation to the initial states. This over-approximation is based on the Jordan normal form decomposition that allows deriving closed form symbolic expressions for the entries of the matrix modeling the effect of exactly n iterations of the loop. We then discover linear relationships between the symbolic expressions that hold for any number of iterations, and we obtain a set of matrices described by a polyhedra on its coefficients, which can be applied to a set of vectors also described by a convex polyhedra.

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.

These ideas were implemented and evaluated on a series of simple loops, alone or inside outer loops, exhibiting classical behaviors: polynomial, stable and unstable exponential, inward spirals (damped oscillators), Our approach enables proofs that are out of the reach of most other techniques, 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.

This work was initiated during a visit to the University of Colorado-Boulder and is under review.

6.5.3. Logico-Numerical Max-Strategy Iteration

Strategy iteration methods aim at solving fixed point equations and are an alternative to abstract acceleration for fighting against the loss of precision incurred by extrapolation in classical interpretation. It has been shown that they improve precision in static analysis based on abstract interpretation and template abstract domains, *e.g.*, intervals, octagons or template polyhedra. However, they are limited to numerical programs.

We investigated a method for applying max-strategy iteration to logico-numerical programs, that is, programs with numerical and Boolean variables, without explicitly enumerating the Boolean state space. The method is optimal in the sense that it computes the *least fixed point* w.r.t. the abstract domain.

Our experiments showed that the resulting logico-numerical max-strategy iteration gains one order of magnitude in terms of efficiency in comparison to the purely numerical approach while being almost as precise. Moreover, they are the first experimental results of applying max-strategy iteration to larger programs. This work has been accepted at VMCAI'2013 [23].

6.6. Component-Based Construction

Participants: Alain Girault, Gregor Goessler [contact person], Roopak Sinha, Gideon Smeding.

6.6.1. Incremental converter synthesis

We have proposed and implemented a formal incremental converter-generation algorithm for system-on-chip (SoC) designs. The approach generates a converter, if one exists, to control the interaction between multiple intellectual property (IP) protocols with possible control and data mismatches, and allows pre-converted systems to be re-converted with additional IPs in the future. IP protocols are represented using labeled transition systems (LTS), a simple but elegant abstraction framework which can be extracted from and converted to standard IP description languages such as VHDL. The user can provide control properties, each stated as an LTS with accepting states, to describe desired aspects of the converted system, including fairness and liveness. Furthermore, data specifications can be provided to bound data channels between interacting IPs such that they do not over/under flow. The approach takes into account the uncontrollable environment of a system by allowing users to identify signals exchanged between the SoC and the environment, which the converter can neither suppress nor generate.

Given these inputs, the conversion algorithm first computes the reachable state-space of a maximal non-deterministic converter that ensures (i) the satisfaction of the given data specifications and (ii) the trace equivalence with the given control specifications, using a greatest fix-point computation. It then checks, using the standard algorithm for Büchi games, whether the converter can ensure the satisfaction of the given control specifications (reachability of accepting states) regardless of how the environment behaves. If this is

found to be true, deterministic converters can be automatically generated from the maximal non-deterministic converter generated during the first step. The algorithm is proven to be sound and complete, with a polynomial complexity in the state-space sizes of given IP protocols and specifications. It is also shown that it can be used for incremental design of SoCs, where IPs and specifications are added to an SoC in steps. Incremental design allows to constrain the combinatorial explosion of the explored state-space in each step, and also reduces on-chip wire congestion by decentralizing the conversion process.

A Java implementation has been created, and experimental results show that the algorithm can handle complex IP mismatches and specifications in medium to large AMBA based SoC systems. Future work involves creating a library of commonly-encountered specifications in SoC design such as sharing of control signals between interacting IPs using buffers, signal lifespans, and the generation of optimal converters based on quantitative criteria such as minimal power usage.

This work has been done within the AFMES associated team with the Electric and Computer Engineering Department of the University of Auckland.

6.6.2. 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 [62] and its instantiations. However, this approach relies on a causal model that may not be known, for instance in presence of black-box components.

In [6] we have proposed a formal framework for reasoning about causality, based on black-box components interacting according to well identified *interaction models* [5].

We are currently extending to framework to other models of computation and communication, in particular, to timed automata, and developing a refinement of our original approach that reduces the number of false positives.

6.6.3. A Theory of fault recovery for component-based models

In [35][18] we have introduced a theory of fault recovery for component-based models. A model is specified in terms of a set of atomic components that are incrementally composed and synchronized by a set of glue operators. We have defined what it means for such models to provide a recovery mechanism, so that the model converges to its normal behavior in the presence of faults (*e.g.*, in self-stabilizing systems). We have identified corrector components whose presence in a model is essential to guarantee recovery after the occurrence of faults. We have also formalized component based models that effectively separate recovery from functional concerns. We have shown that any model that provides fault recovery can be transformed into an equivalent model, where functional and recovery tasks are modularized in different components.

6.7. Aspect-Oriented Programming

Participants: Dmitry Burlyayev, Pascal Fradet [contact person], Alain Girault.

The goal of Aspect-Oriented Programming (AOP) is to isolate aspects (such as security, synchronization, or error handling) which cross-cut the program basic functionality and whose implementation usually yields tangled code. In AOP, such aspects are specified separately and integrated into the program by an automatic transformation process called *weaving*.

Although this paradigm has great practical potential, it still lacks formalization and undisciplined uses make reasoning on programs very difficult. Our work on AOP addresses these issues by studying foundational issues (semantics, analysis, verification) and by considering domain-specific aspects (availability, fault tolerance or refinement aspects) as formal properties.

6.7.1. Aspects preserving properties

Aspect Oriented Programming can arbitrarily distort the semantics of programs. In particular, weaving can invalidate crucial safety and liveness properties of the base program.

We have identified categories of aspects that preserve some classes of properties [13]. Our categories of aspects comprise, among others, observers, aborters, and confiners. For example, observers do not modify the base program's state and control-flow (*e.g.*, persistence, profiling, and debugging aspects). These categories are defined formally based on a language independent abstract semantic framework. The classes of properties are defined as subsets of LTL for deterministic programs and CTL* for non-deterministic ones. We have formally proved that, for any program, the weaving of any aspect in a category preserves any property in the related class. In a second step, we have designed for each aspect category a specialized aspect language which ensures that any aspect written in that language belongs to the corresponding category. These languages preserve the corresponding classes of properties by construction.

This work was conducted in collaboration with Rémi Douence from the ASCOLA INRIA team at École des Mines de Nantes.

6.7.2. Fault tolerance aspects

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 fault-tolerance aspects in digital circuits. To this aim, we consider program transformations for hardware description languages (HDL). Our goal is to design an aspect language allowing users to specify and tune a wide range of fault tolerance techniques, while ensuring that the woven HDL program remains synthesizable. The advantage would be to produce fault-tolerant circuits by specifying fault-tolerant strategies separately from the functional specifications.

We have reviewed the different fault tolerant techniques used in integrated circuits: concurrent error detection, error detecting and correcting codes (Hamming, Berger codes, ...), spatial and time redundancy. We have designed a simple hardware description language inspired from LUSTRE and Lucid Sychrone. It is a core functional language manipulating synchronous boolean streams. Faults are represented by bit flips and we take into account all fault models of the form "at most 1 faults within n clock signals". The language semantics as well as the fault model have been formalized in Coq. Many basic (library) properties have been shown on that language.

We are currently expressing different variants of triple modular redundancy (TMR) as program transformations. We are also studying optimizations to prevent the insertion of useless voters in TMR. The next step is to prove that these transformations make the programs fault tolerant *w.r.t.* specific fault models. Further work also includes the study of mixed techniques (*e.g.*, spatial and time redundancy), their high level specification using an AOP-like language and their implementation as transformations.

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. Other works consist of analysing kinematic restitution laws based on the use of the kinetic energy metric. We also performed an analysis of the rocking block motion in terms of the kinetic angles between the two unilateral constraints. Results are in [21], [22] [55]. 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 [20].

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 [18]. In [23] we have provided a tutorial on similar types of systems like relay systems, and their relationships with other formalisms like complementarity systems, or switching dynamical systems. This follows in fact a research program proposed in [4].

6.3. Optimization

6.3.1. Optimization algorithms for large-scale machine learning problems, and applications in computer vision

Participant: Jérôme Malick.

This collaboration with Zaid Harchaoui (Inria, LEAR Team) has been growing since summer 2010. It also involves Miro Dudik (Microsoft Research NYC) and a student who just started his PhD in october 2012 (after his master with us).

The explosion of data that we are experiencing (Big Data) lead us to huge-scale learning problems, new challenges for statistical learning and numerical optimisation algorithms. For example, the new databases for image categorization are large-scale in the three dimensions (large number of examples, high-dimension feature description, and large number of categories). The resulting learning problem is out of reach by standard optimization problems.

We developed a new approach exploiting the hidden underlying lower-dimension structure of this big data. We proposed a new family of algorithms (of the type coordinate results, or conditional gradient), whose iterations have an algorithmic complexity lower than an order compared to standard methods. For example, applied to learning problems with trace-norm penalization, our algorithm [26] exploit the atomic decomposition of the norm and compute only an approximate largest singular vector pair (instead of the whole singular value decomposition). Promising results [27] have been obtained on the image database Imagenet, where we show significant improvements compare to the state-of-the-art approaches (one-vs-rest strategies).

6.3.2. Semidefinite programming and combinatorial optimization

Participants: Nathan Krislock, Jérôme Malick.

We have worked with Frederic Roupin (Prof. at Paris XIII) on the use of semidefinite programming to solve combinatorial optimization problems to optimality. Within exact resolution schemes (branch-and-bound), “good” bounds are those with a “good” balance between tightness and computing times.

We proposed a new family of semidefinite bounds for 0-1 quadratic problems with linear or quadratic constraints [50]. The paradigm is to trade computing time for a (small) deterioration of the quality of the usual semidefinite bounds, in view of enhancing this efficiency in exact resolution schemes. Extensive numerical comparisons et tests showed the superior quality of our bounds, when embedded within branch-and-bound schemes, on standard test-problems (unconstrained 0-1 quadratic problems, heaviest k -subgraph problems, and graph bisection problems).

We have embedded the new bounds within branch-and-bound algorithms to solve 2 standard combinatorial optimization problems to optimality.

- *Max-cut.* We developed [34] 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 ([52]) on the large test-problems.
- *Heaviest k -subgraph problems.* Our previous work [51] takes advantage of the new bounds in their basic form to prune very well in the search tree. Its performances are then comparable with the best method (based on convex quadratic relaxation using CPLEX as an engine). Adapting and incorporating the techniques we developed for the max-cut problem, we propose in [35] 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).

Finally, we have worked on making our data sets available online together with a web interface for our solvers. We have also started working on a generic online semidefinite-based solver for binary quadratic problems using the generality of [50]. All this is publicly available on line at <http://lipn.univ-paris13.fr/BiqCrunch/>.

6.3.3. Unified theory of inaccurate bundle methods

Participants: Claude Lemaréchal, Welington Oliveira.

Convergence of bundle methods is an intricate subject. The situation is even worse in the inexact case, where many variants exist, each with its specific *ad hoc* proof techniques.

With C. Sagastizábal (Rio de Janeiro), we have developed a synthetic theory to single out the successive steps when proving convergence of a generic algorithm, as well as the specific hypotheses that they need. Our pattern covers all variants published so far and suggests a new one. The corresponding paper is being finalized.

6.3.4. Stabilizing marginal prices in electricity production

Participants: Claude Lemaréchal, 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 [48]. However, the computed optimal dual variables show a noisy and unstable behaviour, that could prevent their use as price indicator. We propose a simple and controllable way to stabilize the dual optimal solutions, by penalizing the total variation of the prices [36]. Our illustrations on the daily electricity production optimization of EDF show a striking stabilization at a negligible cost.

6.4. Robotics

6.4.1. Hierarchic QP solver

Participants: Pierre-Brice Wieber, Dimitar Dimitrov.

We are working in collaboration with the LAAS-CNRS and the CEA-LIST on solving multi-objective Quadratic Programs with Lexicographic ordering: Hierarchic QPs [47]. The focus this year has been on enabling fast computations in the case of time-varying Hierarchic QPs through warm-starting the active set method. This has been possible by developing an active set method for lexicographic multi-objective ordering [44], [45]. The main difference with respect to classical active set methods is in the use of a “lexicographic” (sometimes called “multi-dimensional”) Lagrange multiplier.

6.4.2. Modeling of human balance in public transports

Participants: Pierre-Brice Wieber, Zohaib Aftab.

Zohaib Aftab finished his PhD thesis in collaboration with IFSTTAR (previously INRETS) on modeling human balance in public transports. A Model Predictive Control scheme has been developed for the prediction of recovery motions, including ankle and hip strategies as well as stepping with adaptive step locations and timings [37]. This MPC scheme has been validated against a balance recovery scenario found in the biomechanics literature [38].

6.4.3. Model Predictive Control for Biped Walking

Participants: Pierre-Brice Wieber, Andrei Herdt, Jory Lafaye.

In collaboration with the DLR in Munich, we designed an MPC scheme for biped walking based on the “Capture Point”. This is just a simple change of variable $\xi = x + \frac{1}{\omega}\dot{x}$ that transforms the second-order dynamics of the Center of Mass x of the robot into a cascade of two first-order dynamics, one stable and one unstable. This MPC scheme has been evaluated successfully on the DLR biped robot [49].

Since fast computations are always a key objective for feedback controllers, we designed a change of variable in the underlying QP in order to expose the specific structure between time-varying and time-invariant parts of the Hessian matrix and compute its Cholesky decomposition in an efficient way by pre-computing the decomposition of the time-invariant part.

6.5. Computational Toxicology

Participant:

It is now well recognized that toxicology has entered a new era. Previously mainly based on animal testing, toxicology is now turning to *in vitro* and *in silico* experiments. To assess the risk of chemicals but also to gather and to interpret the massive amounts of experimental data generated by modern toxicology, the development of mathematical and computational tools are essential. An important element in risk assessment of chemicals is the human bioaccumulative potential. We developed a predictive tool for human bioaccumulation assesment using a physiologically based toxicokinetic model [53].

6.6. Computational Biology

Participant:

Biological oscillations have attracted widespread interest from experimentalists, with the *in vivo* design of synthetic oscillators, and from mathematicians, with the study of limit cycles. Oscillations in protein concentrations or gene expression are supposed to be involved in the generation of the rhythms observed in the cell. In many situations, oscillations are originated by negative feedback loops. In [54] we have studied the oscillatory regimes of a negative feedback oscillator and derived the probability of having oscillations.

6.7. Mechanical rods

6.7.1. High-order models of mechanical rods

Participants: Florence Bertails-Descoubes, Romain Casati.

Reduced-coordinates models for rods such as the articulated rigid body model or the super-helix model [39] 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-coordinates 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 we investigated the clothoidal case ($d = 1$) in the 2d case [19], relying on Romberg numerical integration. This 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. All these results will we submitted for publication early 2013.

6.7.2. Inverse modeling of mechanical rods

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

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 [43] we noted that reduced-coordinates models such as the super-helix are well-suited for static inversion in presence of gravity. The main difficulty then amounts to fitting a piecewise helix to an arbitrary input curve. This year in A. Derouet-Jourdan's PhD's thesis, we solved this problem by extending to 3d the floating tangents algorithm introduced in 2d in [43]. 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 method proves to be efficient and robust as it can successfully handle large and complex datasets from real curve acquisitions, such as the capture of hair fibers or the magnetic field of a star. This result was submitted for publication to Computer-Aided Geometric Design in Spring, and is currently under minor revision.

6.8. High-accuracy time-stepping schemes

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 first proposed to improve the global order of accuracy over periods when the evolution is sufficiently smooth by mixing standard higher order schemes for Differential Algebraic equations and the Moreau-Jean's scheme [16]. We also proposed self-adapting schemes by applying time-discontinuous Galerkin methods to the measure differential equation in [24]. 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 [17]. Finally, the energetic behavior of the standard Moreau-Jean scheme has been addressed in [25] by developing a Newmark-type scheme for nonsmooth dynamics.

6.9. 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. 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 [46]. It is shown that excepted when the system is state lossless and $\theta = 0.5$, the conditions are very stringent.

6.9.1. Multivalued Lur'e dynamical systems

Participant: Bernard Brogliato.

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 [40]. The results in [42] extend those in [41] with an accurate characterization of the maximal monotonicity of the central operator of these systems. Concrete and verifiable criteria are provided for the above classes (complementarity, relay systems).

MISTIS Project-Team

6. New Results

6.1. Mixture models

6.1.1. Taking into account the curse of dimensionality

Participant: Stéphane Girard.

Joint work with: Bouveyron, C. (Université Paris 1), Fauvel, M. (ENSAT Toulouse)

In the PhD work of Charles Bouveyron (co-advised by Cordelia Schmid from the Inria LEAR team) [53], 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]. Also, the description of the R package is published in [11]. Our recent work consists in adding a kernel in the previous methods to deal with nonlinear data classification [27], [45].

6.1.2. Robust mixture modelling using skewed multivariate distributions with variable amounts of tailweight

Participants: Florence Forbes, Darren Wraith.

Clustering concerns the assignment of each of N , possibly multidimensional, observations y_1, \dots, 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). In such cases, the multivariate student t distribution is motivated as a heavy-tailed alternative to the multivariate Gaussian distribution. The additional flexibility of the multivariate t comes from introducing an additional degree of freedom parameter (*dof*) which can be viewed as a robust tuning parameter.

A useful representation of the t -distribution is as a so-called *infinite mixture of scaled Gaussians* or *Gaussian scale mixture*,

$$p(y; \mu, \Sigma, \theta) = \int_0^\infty \mathcal{N}_M(y; \mu, \Sigma/w) f_W(w; \theta) dw \quad (1)$$

where $\mathcal{N}_M(\cdot; \mu, \Sigma/w)$ denotes the M -dimensional Gaussian distribution with mean μ and covariance Σ/w and f_W is the probability distribution of a univariate positive variable W referred to as the weight variable. When f_W is a Gamma distribution $\mathcal{G}(\nu/2, \nu/2)$ where ν denotes the degrees of freedom, we recover the multivariate t distribution. The weight variable W in this case effectively acts to govern the tail behaviour of the distributional form from light tails ($\nu \rightarrow \infty$) to heavy tails ($\nu \rightarrow 0$) depending on the value of ν .

For many applications, the distribution of the data may also be highly asymmetric in addition to being heavy tailed (or affected by outliers). A natural extension to the Gaussian scale mixture case is to consider *location and scale Gaussian mixtures* of the form,

$$p(y; \mu, \Sigma, \theta) = \int_0^\infty \mathcal{N}_M(y; \mu + w\beta\Sigma, w\Sigma) f_W(w; \theta) dw, \quad (2)$$

where β is an additional M -dimensional vector parameter for skewness and the determinant of Σ equals 1 for parameter identifiability. When f_W is a Generalized Inverse Gaussian distribution ($GIG(y; \lambda, \delta, \gamma)$), we recover the family of Generalized Hyperbolic (GH) distributions. Depending on the parameter choice for the GIG, special cases of the GH family, include: the multivariate GH distribution with hyperbolic margins ($\lambda = 1$); the normal inverse Gaussian distribution ($\lambda = -1/2$); the multivariate hyperbolic ($\lambda = \frac{M+1}{2}$) distribution; the hyperboloid distribution ($\lambda = 0$); the hyperbolic skew- t distribution ($\lambda = -\nu, \gamma = 0$); and the normal gamma distribution ($\lambda > 0, \mu = 0, \delta = 0$) amongst others. For applied problems, the most popular of these forms appears to be the Normal Inverse Gaussian (NIG) distribution, with extensive use in financial applications. Another distributional form allowing for skewness and heavy or light tails includes different forms of the multivariate skew- t . Most of these distributional forms are also able to be represented as *location and scale Gaussian mixtures*.

Although the above approaches provide for great flexibility in modelling data of highly asymmetric and heavy tailed form the above approaches assume f_W to be a univariate distribution and hence each dimension is governed by the same amount of tailweight. There have been various approaches to address this issue in the statistics literature for both symmetric and asymmetric distributional forms. In his work, [66] proposes a dependent bivariate t -distribution with marginals of different degrees of freedom but the tractability of the extension to the multivariate case is unclear. Additional proposals are reviewed in chapters 4 and 5 of [67] but these formulations tend to be appreciably more complicated, often already in the expression of the probability density function. Increasingly, there has been much research on copula approaches to account for flexible distributional forms but the choice as to which one to use in this case and the applicability to (even) moderate dimensions is also not clear. In general the papers take various approaches whose relationships have been characterized in the bivariate case by [73]. However, most of the existing approaches suffer either from the non-existence of a closed-form pdf or from a difficult generalization to more than two dimensions.

In this work, we show that the location and scale mixture representation can be further explored and propose a framework that is considerably simpler than those previously proposed with distributions exhibiting interesting properties. Using the normal inverse Gaussian distribution (NIG) as an example, we extend the standard *location and scale mixture of Gaussian representation* to allow for the tail behaviour to be set or estimated differently in each dimension of the variable space. The key elements of the approach are the introduction of multidimensional weights and a decomposition of the matrix Σ in (6) which facilitates the separate estimation and also allows for arbitrary correlation between dimensions. We outline an approach for maximum likelihood estimation of the parameters via the EM algorithm and explore the performance of the approach on several simulated and real data sets in the context of clustering.

6.1.3. Robust clustering for high dimensional data

Participants: Florence Forbes, Darren Wraith, Minwoo Lee.

For a clustering problem, a parametric mixture model is one of the popular approaches. Most of all, Gaussian mixture models are widely used in various fields of study such as data mining, pattern recognition, machine learning, and statistical analysis. The modeling and computational flexibility of the Gaussian mixture model makes it possible to model a rich class of density, and provides a simple mathematical form of cluster models.

Despite the success of Gaussian mixtures, the parameter estimations can be severely affected by outliers. By adding an additional degrees of freedom (dof) parameter, a robustness tuning parameter, the robust improvement in clustering has been achieved. Although adopting t distribution loses the closed-form solution, it is still tractable by representing t distribution as Gaussian scale mixture (GSM), which consists of a Gaussian random vector that is weighted by a hidden scaling variable. Recent work that uses the multivariate t distribution has showed the improved robustness.

Along with robustness from t distribution, for the practical use, efficient handling of a high dimensional data is critical. High dimensional data often make most of clustering methods perform poorly. To overcome the curse of dimensionality, Bouveyron et al. [54] proposed the model-based high dimensional data clustering (HDDC). HDDC searches the intrinsic dimension of each class with the BIC criterion or the scree-test of Cattell; this allows them to limit the number of parameters by taking into account only the specific subspace that each class is located. The parameterization makes HDDC not only computationally efficient but robust with respect to the ill-conditioning or the singularity of empirical covariance matrix.

This work proposes an approach that combines robust clustering with the HDDC. The use of the mixture of multivariate t distribution on the basis of HDDC develops robust high dimensional clustering methods that can capture various kinds of density models. Further, extending the mixture model with multiple t distributions for each dimension, we propose more flexible model that can be applicable to various data. We suggest a model-based approach for this method.

6.1.4. Partially Supervised Mapping: A Unified Model for Regression and Dimensionality Reduction

Participant: Florence Forbes.

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

We cast dimensionality reduction and regression in a unified latent variable model. We propose a two-step strategy consisting of characterizing a non-linear *reversed* output-to-input regression with a generative piecewise-linear model, followed by Bayes inversion to obtain an output density given an input. We describe and analyze the most general case of this model, namely when only some components of the output variables are observed while the other components are latent. We provide two EM inference procedures and their initialization. Using simulated and real data, we show that the proposed method outperforms several existing ones.

6.1.5. Variational EM for Binaural Sound-Source Separation and Localization

Participant: Florence Forbes.

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

We addressed the problem of sound-source separation and localization in real-world conditions with two microphones. Both tasks are solved within a unified formulation using supervised mapping. While the parameters of the direct mapping are learned during a training stage that uses sources emitting white noise (calibration), the inverse mapping is estimated using a variational EM formulation. The proposed algorithm can deal with natural sound sources such as speech which are known to yield sparse spectrograms, and is able to locate multiple sources both in azimuth and in elevation. Extensive experiments with real data show that the method outperform state-of-the-art both in separation and localization.

6.2. Statistical models for Neuroscience

6.2.1. Variational approach for the joint estimation-detection of Brain activity from functional MRI data

Participants: Florence Forbes, Lotfi Chaari, Thomas Vincent.

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

In standard within-subject analyses of event-related fMRI data, two steps are usually performed separately: detection of brain activity and estimation of the hemodynamic response. Because these two steps are inherently linked, we adopt the so-called region-based Joint Detection-Estimation (JDE) framework that addresses this joint issue using a multivariate inference for detection and estimation. JDE is built by making use of a regional bilinear generative model of the BOLD response and constraining the parameter estimation by physiological priors using temporal and spatial information in a Markovian model. In contrast to previous works that use Markov Chain Monte Carlo (MCMC) techniques to sample the resulting intractable posterior distribution, we recast the JDE into a missing data framework and derive a Variational Expectation-Maximization (VEM) algorithm for its inference. A variational approximation is used to approximate the Markovian model in the unsupervised spatially adaptive JDE inference, which allows automatic fine-tuning of spatial regularization parameters. It provides a new algorithm that exhibits interesting properties terms of estimation error and computational cost compared to the previously used MCMC-based approach. Experiments on artificial and real data show that VEM-JDE is robust to model mis-specification and provides computational gain while maintaining good performance in terms of activation detection and hemodynamic shape recovery. Main corresponding paper [13]

6.2.2. *Hemodynamic-informed parcellation of fMRI data in a Joint Detection Estimation framework*

Participants: Florence Forbes, Lotfi Chaari, Thomas Vincent.

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

Identifying brain hemodynamics in event-related functional MRI (fMRI) data is a crucial issue to disentangle the vascular response from the neuronal activity in the BOLD signal. This question is usually addressed by estimating the so-called Hemodynamic Response Function (HRF). Voxelwise or region-/parcelwise inference schemes have been proposed to achieve this goal but so far all known contributions commit to pre-specified spatial supports for the hemodynamic territories by defining these supports either as individual voxels or a priori fixed brain parcels. In this paper, we introduce a Joint Parcellation-Detection-Estimation (JPDE) procedure that incorporates an adaptive parcel identification step based upon local hemodynamic properties. Efficient inference of both evoked activity, HRF shapes and *supports* is then achieved using variational approximations. Validation on synthetic and real fMRI data demonstrate the JPDE performance over standard detection estimation schemes and suggest it as a new brain exploration tool. Corresponding papers [29], [28].

6.2.3. *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. 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 [25]. Corresponding paper [26].

6.2.4. *Bayesian BOLD and perfusion source separation and deconvolution from functional ASL imaging*

Participants: Florence Forbes, Thomas Vincent.

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

In many neuroscience applications, the Arterial Spin Labeling (ASL) fMRI modality arises as a preferable choice to the standard BOLD modality due to its ability to provide a quantitative measure of the Cerebral Blood Flow (CBF). Such a quantification is central but generally performed without consideration of a specific modeling of the perfusion component in the signal often handled via standard GLM approaches using the BOLD canonical response function as regressor. In this work, we propose a novel Bayesian hierarchical model of the ASL signal which allows activation detection and both the extraction of a perfusion and a hemodynamic component. Validation on synthetic and real data sets from event-related ASL show the ability of our model to address the source separation and double deconvolution problems inherent to ASL data analysis.

6.2.5. *Extraction of physiological components in functional ASL data*

Participants: Florence Forbes, Thomas Vincent, Lotfi Chaari, Marc Guillotin.

In the context of ARC AINSI project, joint work with: Jan Warnking (Grenoble Institute of Neuroscience) and Philippe Ciuciu from Neurospin, CEA in Saclay.

The internship of Marc Guillotin has been supported by Le pole Cognition de Grenoble.

The goal of this work was to investigate Independent component analysis techniques to identify the part of the ASL signal due to physiological sources such as respiratory and cardiac components. Once identified those physiological components should be removed to produce an uncontaminated ASL signal. This preliminary work showed that the physiological effects were affecting all signal components and were therefore not easy to extract without removing some of the useful signal. More experiments should be made on real data from the GIN.

6.2.6. *Comparison of processing workflows for ASL data analysis*

Participant: Thomas Vincent.

In the context of ARC AINSI project, joint work with: Michel Dojat (Grenoble Institute of Neuroscience), Philippe Ciuciu from Neurospin, CEA in Saclay, Remi Dubujet, Elise Bannier, Isabelle Courouge, Christian Barillot, Camille Maudet from EPI Visages in Rennes.

We assessed and compared the performance of different ASL processing pipelines in order to promote one using specific indexes (Contrast to noise ratio, partial volume effect, et). We proposed to assess the impact of the pipelines based on the quality of the final corrected ASL images using a common set of subjects for all workflows. We leaned on the expertise of the Visages and GIN teams on ASL, and first started from existing attempts made in the teams. At the moment, there is a striking lack of such guidelines. The recent toolbox ASLtbx proposes a number of procedures that are based on very standard tools (e.g. SPM) and do not make use of more efficient approaches from more recent literature. Similarly, in the BIRN project, processing pipelines are mentioned but none are currently available.

6.3. Markov models

6.3.1. *Spatial risk mapping for rare disease with hidden Markov fields and variational EM*

Participants: Florence Forbes, Senan James Doyle.

Joint work with: Lamiae Azizi, David Abrial and Myriam Garrido from INRA Clermont-Ferrand-Theix.

Current risk mapping models for pooled data focus on the estimated risk for each geographical unit. A risk classification, *i.e.* grouping of geographical units with similar risk, is then necessary to easily draw interpretable maps, with clearly delimited zones in which protection measures can be applied. As an illustration, we focus on the Bovine Spongiform Encephalopathy (BSE) disease that threatened the bovine production in Europe and generated drastic cow culling. This example features typical animal disease risk analysis issues with very low risk values, small numbers of observed cases and population sizes that increase the difficulty of an automatic classification. We propose to handle this task in a spatial clustering framework using a non standard discrete hidden Markov model prior designed to favor a smooth risk variation. The model parameters are estimated using an EM algorithm and a mean field approximation for which we develop a new initialization strategy appropriate for spatial Poisson mixtures. Using both simulated and our BSE data, we show that our strategy performs well in dealing with low population sizes and accurately determines high risk regions, both in terms of localization and risk level estimation.

Main corresponding paper [14].

6.3.2. *Spatial modelling of biodiversity from high-throughput 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 EPI Bamboo

Biodiversity has been acknowledged as a vital resource for ecosystem health and stability, faced with an unprecedented global decline. In order to be effective, conservation actions need to be based on reliable and fast analysis. Recent advances in DNA sequencing methods now enable DNA-based identification of multiple species from only few, even potentially degraded environmental samples (metabarcoding.org, [74]). This offers a new way of biodiversity assessment and is of particular interest where large-scale individual-based diversity assessment is difficult, for example in tropical environments. Due to their comparatively low demand in cost and effort, these methods are characterized by their high throughput; they are expected to produce vast amounts of data as they gain in popularity over the coming years. The specific properties of these data (e.g. bias from sequencing errors, notion of species) and their high dimensionality provides new statistical and computational challenges for biodiversity assessment. This project aims at extending existing summary statistics to be used with data from metabarcoding surveys and, where this is not adequate, to develop new methodology. A special focus is on the spatial mapping of biodiversity and the co-occurrence of species. In a first instance, we investigate spatial clustering algorithms based on Markov random fields (software SpaCEM3, <http://spacem3.gforge.inria.fr/>) to identify regions of high species occurrence as well as structured additive regression models and their implementation to estimate cross-correlations between species occurrences in space [61], [72], [71]. At present, results have been derived in form of species occurrence maps, which take into account pairwise cross-correlation, and interaction graphs.

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 dependencies 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 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. The graph building stage was based on exploring the space of possible chain graph models, which required defining a notion of neighbourhood of these graphs. A parametric distribution was associated with each graph. It was obtained by combining families of univariate and multivariate distributions or regression models. These were chosen by selection model procedures among different parametric families.

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. 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)

The aim of this work was to characterize genetic determinisms of the alternation of flowering in apple tree progenies. 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. *Post-Reflow Automated Optical Inspection of Lead Defects*

Participants: Florence Forbes, Kai Qin, Huu Giao Nguyen, Darren Wraith, Ludovic Leau-mercier.

This is joint work with VI-Technology in the context of the IVP project.

Quality and throughput in printed circuit board (PCB) assembly lines constitute a continuous challenge, especially when placing smaller components on boards that are becoming increasingly dense. Automated optical inspection (AOI) technology allows PCB assembly lines to keep operating at a high throughput while visually inspecting production quality in term of paste deposits, mounted components and solder joints in an automatic and non-contact manner. In the AOI, high definition cameras precisely move in both X- and Y-direction to scan the device under test lit by special lighting techniques, e.g. light-emitting diode (LED) lighting. The captured images are then analyzed using specific inspection algorithms to identify defects. The AOI systems can be placed at several stages during the manufacturing process, such as bare board inspection, solder paste inspection, pre-reflow inspection and post-reflow inspection, which usually need some time to be programmed via offline learning of verified boards and expert expertise before online inspection starts. Vi TECHNOLOGY (VIT) offers a wide range of AOI solutions to increase productivity throughout electronics manufacturing lines while enhancing the quality of products. Post-reflow AOI is implemented after the reflow procedure in PCB assembly lines to enable inspection of the major post-reflow defects. This work focus on certain types of post-reflow defects occurring on leaded components, i.e. lifted lead, no solder, excess of solder, contamination on lead, insufficient solder, bad wedding and dry joint. We aim at developing efficient post-reflow lead defect detection approaches by synergizing image analysis, pattern recognition, machine learning, and statistics techniques to improve performance of VIT commercial post-reflow AOI solutions from two aspects: 1) Reducing both detection escape rate and false detection rate; 2) Minimizing programming efforts. The exact nature of the work is confidential.

6.4.2. *An Improved CUDA-Based Implementation of Differential Evolution on GPU*

Participants: Kai Qin, Florence Forbes.

Modern GPUs enable widely affordable personal computers to carry out massively parallel computation tasks. NVIDIA's CUDA technology provides a wieldy parallel computing platform. Many state-of-the-art algorithms arising from different fields have been redesigned based on CUDA to achieve computational speedup. Differential evolution (DE), as a very promising evolutionary algorithm, is highly suitable for parallelization owing to its data parallel algorithmic structure. However, most existing CUDA based DE implementations suffer from excessive low-throughput memory access and less efficient device utilization. This work presents an improved CUDA-based DE to optimize memory and device utilization: several logically-related kernels are combined into one composite kernel to reduce global memory access; kernel execution configuration parameters are automatically determined to maximize device occupancy; streams are employed to enable concurrent kernel execution to maximize device utilization. Experimental results on several numerical problems demonstrate superior computational time efficiency of the proposed method over two recent CUDA-based DE and the sequential DE across varying problem dimensions and algorithmic population sizes.

This work was nominated for the best paper award (finalist) in the Digital Entertainment Technologies and Arts / Parallel Evolutionary Systems session of the Genetic and Evolutionary Computation Conference 2012 (GECCO12) conference [33].

6.4.3. *Augmented cumulative distribution networks for multivariate extreme value modelling*

Participants: Stéphane Girard, Gildas Mazo, Florence Forbes.

Max-stable distribution functions are theoretically grounded models for modelling multivariate extreme values. However they suffer from some striking limitations when applied to real data analysis due to the intractability of the likelihood when the number of variables becomes high. Cumulative Distribution Networks (CDN's) have been introduced recently in the machine learning community and allow the construction of max-stable distribution functions for which the density can be computed. Unfortunately, we show in this work that the dependence structure expected in the data may not be accurately reflected by max-stable CDN's. To face this limitation, we therefore propose to augment max-stable CDN's with the more standard Gumbel max-stable distribution function in order to enrich the dependence structure [32].

6.4.4. Modelling extremal events

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

Joint work with: Guillou, A. and Gardes, L. (Univ. Strasbourg).

We introduced a new model of tail distributions depending on two parameters $\tau \in [0, 1]$ and $\theta > 0$. This model includes very different distribution tail behaviors from Fréchet and Gumbel maximum domains of attraction. In the particular cases of Pareto type tails ($\tau = 1$) or Weibull tails ($\tau = 0$), our estimators coincide with classical ones proposed in the literature, thus permitting us to retrieve their asymptotic normality in a unified way. The first year of the PhD work of Jonathan El-methni has been dedicated to the definition of an estimator of the parameter τ . This permits the construction of new estimators of extreme quantiles. The results are published in [17]. Our future work will consist in proposing a test procedure in order to discriminate between Pareto and Weibull tails.

We are also 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 [64], [63]). This work is in collaboration with El-Hadji Deme, a PhD student from the Université de Saint-Louis (Sénégal). El-Hadji Deme obtained a one-year mobility grant to work within the Mistis team on extreme-value statistics. The results are submitted for publication [49]. We also proposed reduced-bias estimators of the Proportional Hazard Premium for heavy-tailed distributions. The results are submitted for publication [50].

6.4.5. Conditional extremal events

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

Joint work with: L. Gardes, Amblard, C. (TimB in TIMC laboratory, Univ. Grenoble I) and Daouia, A. (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 [59] 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) and the tail of the distribution is heavy, we focus on kernel methods [18]. We extension to all kind of tails in investigated in [15].

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 LTHER (Laboratoire d'étude des Transferts en Hydrologie et Environnement) located in Grenoble.

More future work will include the study of multivariate and spatial extreme values. With this aim, a research on some particular copulas [1] has been initiated with Cécile Amblard, since they are the key tool for building multivariate distributions [69]. The PhD theses of Jonathan El-methni and Gildas Mazo should address this issue too.

6.4.6. Level sets estimation

Participant: Stéphane Girard.

Joint work with: Guillou, A. and Gardes, L. (Univ. Strasbourg), Stupfler, G. (Univ. Strasbourg) and Daouia, A. (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].

In collaboration with A. Daouia, we investigate the application of such methods in econometrics [42], [48]: 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 the PhD thesis of Gilles Stupfler (co-directed by Armelle Guillou and Stéphane Girard), 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 [19], [20], [21].

6.4.7. Quantifying uncertainties on extreme rainfall estimations

Participant: Stéphane Girard.

Joint work with: Carreau, J. (Hydrosiences Montpellier), Gardes, L. (univ. Strasbourg) and Molinié, G. from Laboratoire d'Etude des Transferts en Hydrologie et Environnement (LTHE), France.

Extreme rainfalls are generally associated with two different precipitation regimes. Extreme cumulated rainfall over 24 hours results from stratiform clouds on which the relief forcing is of primary importance. Extreme rainfall rates are defined as rainfall rates with low probability of occurrence, typically with higher mean return-levels than the maximum observed level. For example Figure 2 presents the return levels for the Cévennes-Vivarais region that can be obtained. It is then of primary importance to study the sensitivity of the extreme rainfall estimation to the estimation method considered.

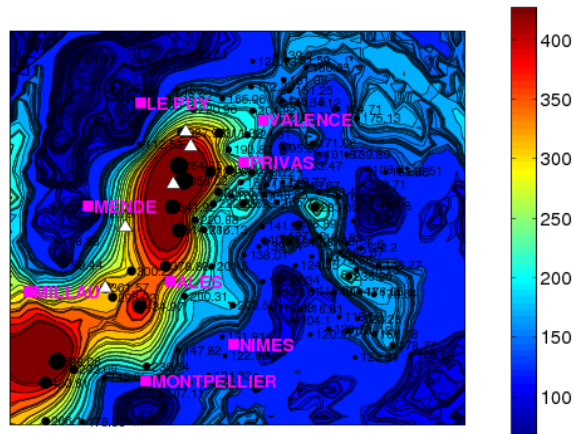


Figure 2. Map of the mean return-levels (in mm) for a period of 10 years.

The obtained results are published in [12].

6.4.8. Retrieval of Mars surface physical properties from OMEGA hyperspectral images.

Participant: Stéphane Girard.

Joint work with: Douté, S. from Laboratoire de Planétologie de Grenoble, France and Saracco, J (University 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) [47]. We have also defined an adaptive version of the method which is able to deal with block-wise evolving data streams [46].

6.4.9. Statistical modelling development for low power processor.

Participant: Stéphane Girard.

Joint work with: A. Lombardot and S. Joshi (ST Crolles).

With scaling down technologies to the nanometer regime, the static power dissipation in semiconductor devices is becoming more and more important. Techniques to accurately estimate System On Chip static power dissipation are becoming essential. Traditionally, designers use a standard corner based approach to optimize and check their devices. However, this approach can drastically underestimate or over-estimate process variations impact and leads to important errors.

The need for an effective modeling of process variation for static power analysis has led to the introduction of Statistical static power analysis. Some publication state that it is possible to save up to 50% static power using statistical approach. However, most of the statistical approaches are based on Monte Carlo analysis, and such methods are not suited to large devices. It is thus necessary to develop solutions for large devices integrated in an industrial design flow. Our objective is to model the total consumption of the circuit from the probability distribution of consumption of each individual gate. Our preliminary results are published in [23].

NANO-D Team

6. New Results

6.1. Adaptively Restrained Particle Simulations

Participants: Svetlana Artemova, Stephane Redon.

Last year, we have introduced a novel, general approach to speed up particle simulations that we call Adaptively Restrained Particle Simulations (ARPS). This year we continued working on this approach. The obtained results have been published in Physical Review Letters [3], and the patent describing the theoretical basis and the algorithms for the numerical realization of ARPS has been deposited.

Particle simulations are widely used in physics, chemistry, biology [13], [14], and even computer graphics [9], and faster simulations (in particular ARPS) may result in progress on many challenging problems, e.g., protein folding, diffusion across bio-membranes, fracture in metals, ion implantation, etc.

ARPS rely on an adaptively restrained (AR) Hamiltonian used to describe a system of N particles:

$$H_{AR}(\mathbf{q}, \mathbf{p}) = \frac{1}{2} \mathbf{p}^T \Phi(\mathbf{q}, \mathbf{p}) \mathbf{p} + V(\mathbf{q}).$$

This Hamiltonian has an unusual inverse inertia matrix $\Phi(\mathbf{q}, \mathbf{p})$, which is made a general function of phase-space coordinates. The precise form of this matrix can be chosen according to the system under study and the problem stated.

We have proposed a particular (diagonal) form of the inverse inertia matrix for the simulations in Cartesian coordinates. In this case, Φ adaptively switches on and off positional degrees of freedom of individual particles while letting particle momenta evolve. The decision whether the particle is restrained or not depends on the particle's momentum, and, precisely, on its kinetic energy. Two user-defined thresholds regulate the amount of simplification of the particle's motion. When a module of a particle's momentum becomes small enough (without necessarily becoming zero), the particle completely stops moving. Even when a particle is fully restrained, though, its momentum may continue to change, and its kinetic energy might become large enough again for the particle to resume moving. In general, ARPS restrain and release particles repeatedly over time.

This approach has numerous advantages: (a) it is mathematically grounded and is able to produce long, stable simulations; (b) it does not require modifications to the simulated interaction potential, so that any suitable existing force-field can be directly used with ARPS; (c) under frequently-used assumptions on the interaction potential, ARPS make it possible to reduce the number of forces that have to be updated at each time step, which may significantly speed up simulations; (d) when performing constant-energy simulations, ARPS allow users to finely and continuously trade between precision and computational cost, and rapidly obtain approximate trajectories; (e) the trade-off between precision and cost may be chosen for each particle independently, so that users may arbitrarily focus ARPS on specific regions of the simulated system (e.g., a polymer in a solvent); (f) most important, when performing Adaptively Restrained Molecular Dynamics (ARMD) in the canonical (NVT) ensemble, correct static equilibrium properties can be computed.

We have demonstrated the advantages of ARPS on several numerical experiments. For example, a planar collision cascade study in Fig. 7 shows how ARPS make it possible to smoothly trade between precision and speed of the simulation. Reference simulations were derived from the usual Hamiltonian $H(\mathbf{q}, \mathbf{p}) = \frac{1}{2} \mathbf{p}^T \mathbf{M}^{-1} \mathbf{p} + V(\mathbf{q})$.

6.2. Hierarchical Adaptively Restrained Particle Simulations

Participants: Svetlana Artemova, Stephane Redon.

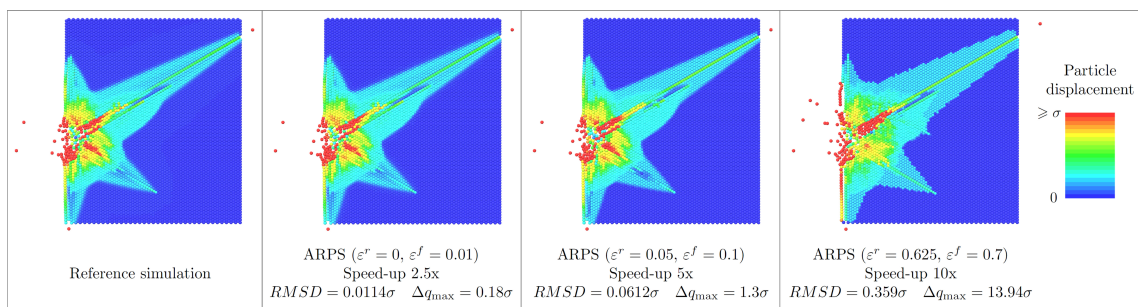


Figure 7. Simulating a collision cascade with controlled precision. Adaptively restrained simulations allow us to smoothly trade between precision and speed. Even for large speed-ups (up to 10x) the features of the shock are extremely well preserved.

It has been shown that algorithms relying on hierarchical representations of molecular systems may accelerate molecular simulations: for example, divide-and-conquer approach for simulations in internal coordinates [10], [11], adaptive algorithms for dynamics of articulated bodies [15], algorithms for neighbor search for system with symmetries [12] or for large rigid molecules [8].

Therefore, we were interested in combining hierarchically-based algorithms with Adaptively Restrained Particle Simulations (ARPS). Precisely, as with classical ARPS, we have considered the adaptively restrained (AR) Hamiltonian:

$$H_{AR}(\mathbf{q}, \mathbf{p}) = \frac{1}{2} \mathbf{p}^T \Phi(\mathbf{q}, \mathbf{p}) \mathbf{p} + V(\mathbf{q}),$$

but we have introduced a different form of the inverse inertia matrix $\Phi(\mathbf{q}, \mathbf{p})$. In this case, again, positional degrees of freedom are adaptively switched on and off during the simulation, but, these are *relative* positional degrees of freedom in the system, and not the positional degrees of freedom of individual particles. Precisely, particles are grouped together into rigid bodies according to the tree representation and released repeatedly during the simulation. We call this approach hierarchical Adaptively Restrained Particle Simulations (hierarchical ARPS).

We have performed several numerical experiments to illustrate this new approach. For example, in Fig. 8 we present the planar collision cascade study.

For hierarchical AR simulations, obtained results depend on the tree representation of the system: for the results demonstrated in Fig. 8 the tree was constructed in a top-down manner by recursive dividing of the system in halves and, therefore, the squares of different levels are being activated by the shock.

To clearly demonstrate the effect of the tree, we provide the results for the same four simulations with another tree built in a bottom-up manner by grouping the particles pairwise according to their sequence number (they were enumerated, first, along the y -axis, vertically, and then, along the x -axis, horizontally). These results are shown in Fig. 9, and are rather different from those in Fig. 8: vertical lines are being activated when the central part of the plane is reached by the shock.

The patent reporting the principles and the algorithms used to implement hierarchical ARPS has been deposited.

6.3. Interactive quantum chemistry

Participants: Mael Bosson, Caroline Richard, Antoine Plet, Sergei Grudin, Stephane Redon.

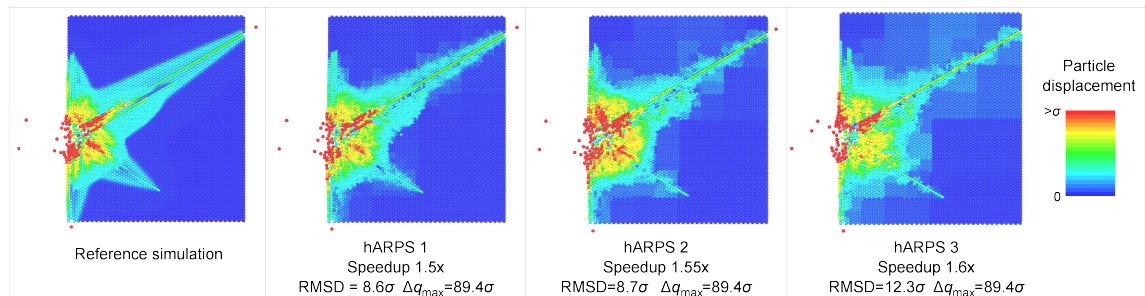


Figure 8. Simulating a collision cascade with controlled precision. Hierarchical adaptively restrained simulations allow us to smoothly trade between precision and speed. The main features of the shock are preserved. The binary tree representation was constructed top-down.

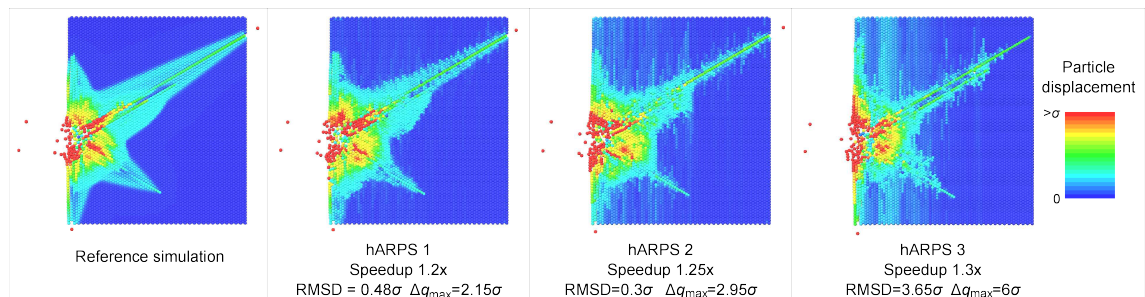


Figure 9. Simulating a collision cascade with controlled precision. Hierarchical adaptively restrained simulations allow us to smoothly trade between precision and speed. The main features of the shock are preserved. The binary tree representation was constructed bottom-up.

Interactive simulation tools allow users to take advantage of their knowledge and intuition to understand physical properties and prototype new devices. To accurately describe bond breaking, bond formation, charge transfer or other electronic phenomena, interactive simulation should ideally be based on quantum mechanics. However, solving quantum chemistry models at interactive rates is a challenging task. Thanks to the algorithms developed in the group, SAMSON is the first software to propose interactive quantum chemistry.

A first contribution allows for interactive quantum chemistry with systems up to a few hundred atoms [6]. The method is based on a divide-and-conquer (D&C) approach. The D&C technique subdivides the system into many subsystems (a–h on the Figure 10). Each of them involves a diagonalisation at each time step. To treat larger systems, we introduce a new algorithm: Block-Adaptive Quantum Mechanics (BAQM) [5] from the combination of two new components.

- **Block-adaptive Cartesian mechanics**

By freezing atomic positions in some subsystems (d–h on the Figure 10) (with atoms in blue), we may avoid updating some eigenproblems. The Block-adaptive Cartesian mechanics component takes advantage of this to control the simulation cost by adaptively adjusting the number of diagonalisations, based on the forces applied to the atoms. Only the subsystems with the largest applied forces are allowed to have mobile atoms.

- **Adaptive reduced-basis quantum mechanics**

Solving even just one of the subsystem's eigenproblem may be too costly to achieve interactive rates. The Adaptive reduced-basis quantum mechanics component projects the equation in an adaptive reduced basis composed of low-energy eigenvectors that have been computed at a previous time step, to benefit from temporal coherence between successive eigenproblems (subsystems (b) and (c) with atoms in black and white on the Figure 10). We use a simple distance to decide on the fly when to automatically update the reduced basis during the simulation (subsystem (a) with atoms in red on the Figure 10).

We demonstrated that BAQM may accelerate geometry optimization for several atomic systems. Indeed, each step is solved significantly faster by constraining some nuclei and electrons, and, by focusing computational resources on the most active parts of the system, we obtain a faster potential energy descent. The proposed BAQM approach also allows for interactive rates with many atomic systems.

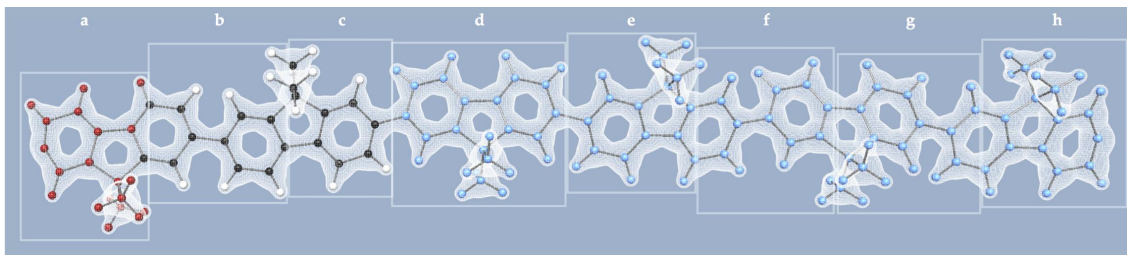


Figure 10. Interactive editing of a polyfluorene molecule with the BAQM algorithm

6.4. Molecular Docking

6.4.1. Development of a new Knowledge-Based Potential for Protein-Ligand Interactions

Participants: Sergei Grudin, Georgy Cheremovskiy.

Macromolecular complexes formed by proteins with small molecules (ligands) play an important role in many biological processes such as signal transduction, cell regulation, etc. Experimental methods for determining the structures of molecular complexes have a very high cost and still involve many difficulties. Therefore, computational methods, such as molecular docking, are typically used for predicting binding modes and affinities, which are essential to understand molecular interaction mechanisms and design new drugs.

Databases containing three-dimensional protein-ligand structures determined by experimental techniques grow very rapidly. In 2011, the PDB (Protein Data Bank) contained about 70,000 of protein structures, with almost 8,000 structures of protein-ligand complexes having refined binding affinity data. The CSD (Cambridge Structural Database), a database for small molecules, contained about 500,000 entries at the beginning of 2012. Thus, we believe that computational tools based on statistical information extracted from three-dimensional structures of protein-ligand complexes will play an ever more increasing role in the functional study of proteins as well as in structure-based drug design and other fields.

We proposed and validated a new statistical method that predicts binding modes and affinities of protein-ligand complexes. To do so, we have developed a novel machine-learning-based approach. Precisely, we have formulated a new optimization problem with 30,000 unknowns, whose solution is a scoring function. We trained the scoring function on 6,000 structures of protein-ligand complexes of high accuracy from the PDB database. Despite the very high dimensionality of the optimization problem, we manage to solve it on a desktop computer in just a few hours.

Our scoring function has three major applications in drug-design:

- Docking: determination of the binding site of a ligand bound to a protein.
- Ranking: identifying a set of ligands with the highest binding affinity for the given protein target by screening a large ligand database.
- Binding constants prediction: prediction of the absolute value of the binding constant of a protein-ligand complex.

The success rates of our method rank it among the top three methods currently available. Thus, we believe that our scoring function is the first one that performs well in all three major applications in drug-design.

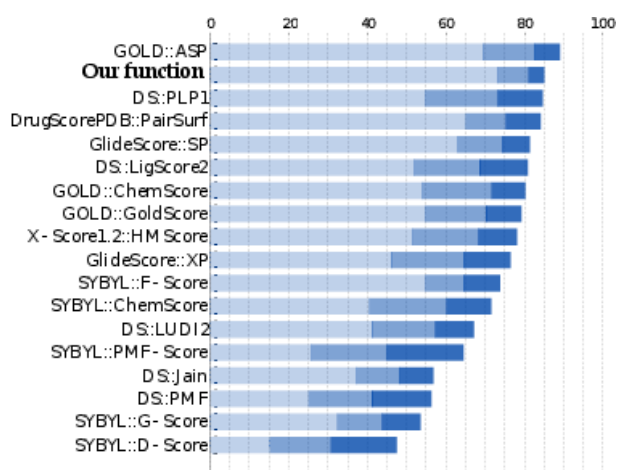


Figure 11. Comparison of the success rates of scoring functions when the best-scored binding pose differs from the true one by RMSD < 1.0 Å (light bars), < 2.0 Å (darker bars) or < 3.0 Å (the darkest bars), respectively. Scoring functions are ranked by success rates when the ligand binding pose is found within RMSD < 3.0 Å.

6.4.2. DockTrina

Participants: Sergei Grudin, Petr Popov.

We derived analytical formulas for fast evaluation of the Root-Mean-Square-Deviation (RMSD) between rigid protein structures. This work resulted in a RMSD library containing algorithms to calculate the RMSD between two proteins in constant time. Based on this library we introduced an efficient algorithm to predict triangular protein structures and implemented it into the DockTrina software. We collected bound benchmarks of 220 protein trimers with and without symmetry properties from the Protein Data Bank and demonstrated the superiority of DockTrina over standard combinatorial algorithms aimed at predicting nonsymmetrical protein trimers.

6.4.3. Machine Learning for Structural Biology

Participants: Sergei Grudin, Petr Popov, Mathias Louboutin.

We developed a new formulation of the machine learning optimization problem to predict protein-protein interactions. We implemented several optimization strategies, both in *dual* and *primal*. We studied the effect of different types of loss-functions on the quality of the prediction. We also tested the efficiency of three descent algorithms, Nesterov descent, gradient descent, and stochastic descent. We demonstrated that generally, primal optimization is faster compared to dual optimization. In the primal, Nesterov descent has a better convergence compared to the gradient descent. Finally, stochastic algorithms often provide a better convergence compared to deterministic algorithms. All the studied algorithms were implemented as a stand-alone library.

6.5. Software Engineering

Participants: Jocelyn Gate, 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). The interface has been improved:

- The visualization of the data graph has been improved. Users may now drag and drop models and parts between layers, as well as directly drag and drop files into SAMSON.
- The undo/redo stack can now be visualized.
- We have begun to work on selection and highlighting.

The software engineering process has been improved as well, in particular to help base and modules developers:

- We have reorganized the file hierarchy so that modules can have associated data.
- We have developed a system to build SAMSON automatically on virtual machines (e.g., ubuntu 12.04 32bit, ubuntu 12.04 64 bit, fedora 17 32 bit, etc.).
- Tools have been created to let modules developers easily write new modules.
- We have begun to develop a mechanism to make it easy to install and update SAMSON automatically.

We have also developed several *SAMSON apps* to test various concepts, including scripting, manipulating molecules with haptic feedback, etc. Figure 12 shows the current user interface of SAMSON.

We have deposited the first version of SAMSON's code base at the APP ("Agence de Protection des Programmes").

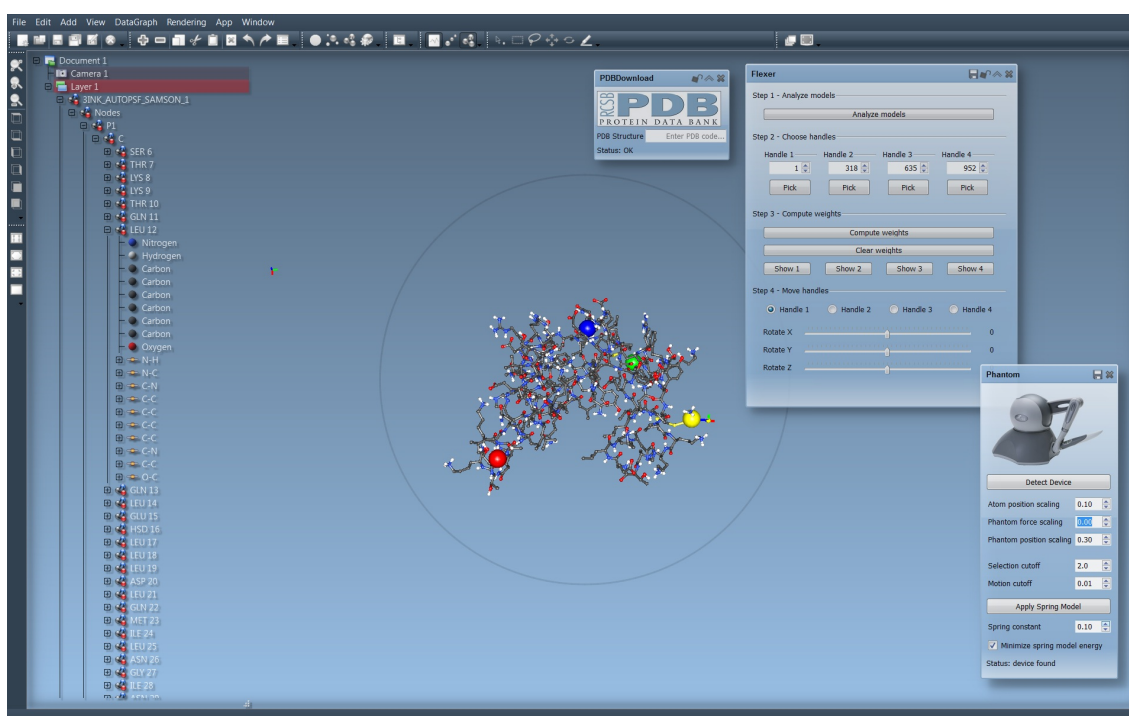


Figure 12. The current user interface of SAMSON, showing an app to download molecules directly from the Protein Data Bank, an app to deform molecules, and an app for haptic interaction. The data graph on the left shows the hierarchical structure of the data graph.

NECS Project-Team

6. New Results

6.1. Communication and control co-design for networked systems

6.1.1. *Energy-aware communication and control co-design in wireless networked control systems*

Participants: C. Canudas de Wit [Contact person], N. Cardoso de Castro, F. Garin, D. Quevedo [Newcastle Univ., Australia].

This work is the topic of the PhD thesis of N. Cardoso de Castro [12]. 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, as we had pointed out in the review paper [63], 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 [31] or over a finite receding horizon [32].

6.1.2. *System-theoretic analysis of modern error correcting codes (serial turbo codes)*

Participants: F. Garin [Contact person], G. Como [Lund Univ., Sweden], F. Fagnani [Polit. Torino, Italy].

Serial turbo codes are a family of codes for error correction in point-to-point digital communication. The encoder can be described as the composition of three linear maps, the intermediate one being a permutation (called interleaver) while the inner and outer one are convolutional codes, i.e., linear dynamical systems where state, input and output belong to a vector space over the finite field $GF(2)$. The decoding is performed with iterative low-complexity algorithms which give a good approximation of the optimal maximum-likelihood (ML) decoder. Using system-theoretic properties of the constituent convolutional codes and probabilistic arguments, we study the average and the typical behavior of ensembles of such codes (with fixed convolutional codes, and random interleaver), asymptotically in the block-length [18]. We disprove the common conjecture that the typical behavior concentrates around the average: indeed, the average error decays polynomially in the block-length N , while the typical code has a faster error decay (exponential in some fractional power of N); however, the typical-code analysis confirms the same design parameters for the convolutional codes that were already suggested by the study of the ensemble average: free distance of the outer encoder, and effective free distance of the inner encoder.

6.2. Networked systems and Graph analysis

6.2.1. *Observability in consensus networks*

Participants: A. Kibangou [Contact person], C. Commault [Gipsa-Lab].

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 in [70] for paths and circular graphs where the study was carried out based on rules on number theory. Herein, we have considered families of graphs admitting an association scheme [62] 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 [79]. Based on the so-called Bose-Mesner algebra [60], 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 so-called 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. In particular, we have shown that observability is ensured in such graphs only if $DK \geq N - 1$ where D is the number of classes of the association scheme, N the number of nodes, and K the valency of the graph, i.e. the cardinality of the neighborhood.

6.2.2. Distributed graph discovery

Participants: A. Kibangou [Contact person], F. Garin [Contact person], C. Commault [Gipsa-Lab], D. Tran, D. Varagnolo [KTH], K.H. Johansson [KTH].

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. Two approaches have been developed. For the first one [38], based on properties of the observability matrix, we have shown that Laplacian eigenvalues can be recovered by solving a local eigenvalue decomposition on an appropriately constructed matrix of observed data. Unlike FFT based methods recently proposed in the literature (see [65], [73]), in our proposed method we are also able to estimate the multiplicities of the eigenvalues. However, this method is only applicable to networks having nodes with sufficient storage and computation capabilities. That's why we have proposed a second method requiring much less computation and storage capabilities in [76]. Based on a recent result showing that the average consensus matrix can be factored in D Laplacian based consensus matrices, where D stands for the number of nonzero distinct Laplacian eigenvalues [40], we have shown how carrying out such a factorization in a fully distributed way. The proposed solution results on a distributed solution of a constrained consensus problem.

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 algorithms. This is particularly difficult in the case when the agents do not have or do not want to disclose their identifiers (IDs), either for technological reasons (in time-varying self-organized networks, assigning unique identifiers to agents is a challenge) or for privacy concerns. In a recent work [78] the authors proposed an algorithm which allows each agent to find an estimate of the number of agents in the network, in an anonymous way. Such an algorithm is based on the generation of pseudo-random numbers, on some consensus algorithms (for distributed computation either of average or of maximum), and on statistical inference. In our work [37], we show how the same algorithm, with some minor modifications, can provide more information: approximations of each node's eccentricity, of the graph diameter and of the graph radius. We study the quality of such approximations, providing tight bounds on the error.

6.3. Distributed methods for control

6.3.1. Distributed control

Participants: A. Seuret [Contact person], G. Rodrigues de Campos, L. Brinon-Arranz, D.V. Dimarogonas [KTH], K.H. Johansson [KTH].

Another particular effort has been provided to the design of distributed control laws for multi-agents systems. Three main contributions have been produced and can be summarized as follows.

In [44], a new consensus algorithms for heterogeneous multi-agent systems is provided. A control strategy based on a consensus algorithm which is decoupled from the original systems is proposed. Consequently, its major advantage remains in the separation of the stability analysis of each subsystem and the distributed control algorithm. It is shown that our method allows using classical distributed consensus algorithms such as simple integrator consensus (with or without delay) and distributed consensus filter algorithms.

For many multi-robot applications it is interesting to impose a particular configuration for the robotic agents. This paper discusses the design and analysis of a distributed algorithm for the compact deployment of agents, where the behavior of each vehicle is only dependent on local information. The objective of the paper [72] is to achieve the most compact formation possible. To solve this problem we propose, in a first step, two uncorrelated controllers: one designed for dispersion with connectivity maintenance and a second designed to minimize inter-agent angles. An improved controller including variable gains, particularly designed to avoid singular configurations, is also provided. Lastly, we propose a sequential strategy composed of the two previously mentioned controllers and a stability analysis based on hybrid systems theory. Finally, some simulation results for different configurations supporting our theoretical results are presented.

6.3.2. Collaborative source seeking control

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

The problem of source localization consists in finding the point or the spatial region from which a quantity of interest is being emitted; this goal can be pursued by one or several agents possibly cooperating each other. 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.

Within the FeedNetBack European project, we have addressed the problem of collaborative source seeking with a fleet of autonomous underwater vehicles (UAVs). This topic was explored in the PhD thesis of Lara Brinon [61], where a solution was proposed, based on circular formations with the center of the formation following a 2-dimensional movement in the direction of the gradient of the source. The gradient computation was achieved through an approximation using the point-wise measurements from the various vehicles.

In a more recent work [29], we leave temporarily aside all issues of coordination and communication failures well-addressed in [61], and we focus on the gradient computation formula. Under some assumptions on the source emission (isotropic diffusive source in steady-state, whose solution satisfies the Laplace equation), we show that there is an exact integral formula (based on the Poisson integral of harmonic functions) for the computation of the gradient at the center of a circle, using pointwise measurements along the circumference. This approach has two main advantages: it can be generalized in three (or more) dimensions, and it allows to compute also higher-order derivatives, which allow to find higher-order control laws, useful e.g. for non-holonomic vehicles. A relevant property is that such an integral formula exploits mathematical properties of the source density distribution (the fact that it is harmonic), but does not require the knowledge of an explicit expression for the density function. This makes our approach different from the main source-seeking techniques present in the literature, which 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.

The latter work is part of the research of Ruggero Fabbiano during his Ph.D. studies.

6.3.3. Distributed real-time Simulation of numerical models

Participants: D. Simon [Contact person], A. Ben Khaled [IFPEN], M. Ben Gaid [IFPEN].

The need of quick innovation in the automotive domain made simulation necessary at early stages of the development cycle. Vehicles and powertrains are complex systems where different domains are involved. Representative phenomenological models of powertrains have been developed and have been used in the design phase under domain dedicated tools. However, their use for controls validation using Model-In-the-Loop (MIL) and Hardware-In-the-Loop (HIL) was prevented due to performance limitation of widely used single-solver/single-core simulation approaches.

Multicore simulation for complex systems has been studied with a focus on simulation duration speedup. The methodology of parallelization across the model has been selected for such problem where strong interactions between the model components are observed. The current study showed that decoupling the model parts by relaxing their data dependencies is promising in term of simulation speed (by increasing the parallelism) and results accuracy. Besides, tests results on engine model showed that, with the model partitioning, it is possible to use efficiently variable-step solvers thanks to the decrease of the number of discontinuities, so the number of integration interrupts, in each subsystem [26].

Further work will investigate in the combination of the use of variable-step solvers in split model with the use of multicore architecture for parallel computing, in order to improve the simulation speedup while keeping results accuracy under control.

6.4. Distributed average consensus algorithms

6.4.1. Finite-time average consensus protocols

Participants: A. Kibangou [Contact person], 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 majority of the 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. 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 [40], [39]. 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 [65], [73], [76]). In [77], 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. Unlike the method in [40], the factor matrices are not necessarily symmetric and the number of these factor matrices is exactly equal to the diameter of the graph.

6.4.2. Quadratic indices for performance evaluation of consensus algorithms

Participants: F. Garin [Contact person], S. Zampieri [Università di Padova], E. Lovisari [Università di Padova and Lund Univ.].

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. Examples of various relevant costs can be found in the book chapter [66] and in the references therein.

We are interested in evaluating the effect of the topology of the communication graph on performance, in particular for large-scale graphs. Motivated by the study of wireless sensor networks, our main objective is to understand the limitations which arise when agents are limited to truly local interactions, i.e., the neighborhoods are determined by being ‘near’ in a geometric (Euclidean) way, differently from graphs with few but possibly ‘distant’ connections, such as in small world models. At first [19] we consider graphs which are regular lattices (infinite lattices, or grids on tori, or grids on hyper-cubes), which are examples of geometrically local interactions, but also have a very rich structure: their symmetries allow to exploit powerful algebraic tools, such as the discrete Fourier transform over rings, to compute their eigenvalues, and then find bounds on the associated costs. Then, we extend the results to a more general class of graphs, thus showing that the behavior of lattices is mainly due to the local nature of interactions and not to the spatial invariance (the richness of the automorphism group). To do so, we exploit the analogy between reversible Markov chains and resistive electrical networks, which allows to study some perturbed grids, with less regularity but still exhibiting the same dimension-dependent asymptotic behavior. This latter work is part of the Ph.D. thesis of E. Lovisari at University of Padova, Italy, and the topic of a journal paper in preparation.

6.5. Distributed Estimation and Data fusion

6.5.1. Distributed joint state and input estimation

Participants: A. Kibangou [Contact person], F. Garin [Contact person], A. Esna Ashari.

Three consensus-based distributed algorithms have been developed for joint state and input estimation in discrete-time systems. The methods are proper substitutes for distributed Kalman filter in the case in which there are additive faults to the system. Previously developed centralized estimation methods have been reformulated so that the estimator can be used for distributed sensor networks. These new forms are similar to the information form of Kalman filter [34], [35]. The new forms can be used to propose distributed algorithms based on the consensus of the nodes on calculation of some matrices and vectors. Also a second algorithm is proposed, based on the consensus of the local estimators on local state estimations. This algorithm has less computation effort than the first, but gives a sub-optimal solution in the sense of covariance error. Finally, a third method based on covariance intersection method for diffusing local estimations was proposed in addition. This method also provides a sub-optimal solution. Compared with the second approach, the diffusion of local data is less complicated, however it requires more message communication between nodes.

6.5.2. Data fusion approaches for motion Capture by Inertial and Magnetic Sensors

Participants: H. Fourati [Contact person], A. Makni.

We are interested to motion capture (or attitude) by fusing Inertial and Magnetic Sensors. In [17], we present a viable quaternion-based Complementary Observer (CO) which is designed for rigid body attitude estimation. We claim that this approach is an alternative one to overcome the limitations of the Extended Kalman Filter (EKF). 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. These results are promising for human motion applications, in particular future ambulatory monitoring. The estimated attitude is used to reconstitute the linear acceleration, linear velocity and finally the 3D position from a usual integration procedure (in the case of foot motion) [36]. The problem of attitude estimation is also recently studied within the PhD thesis of Aida Makni. Our goal is to

develop a new attitude estimation methods in the case of aerial vehicles (hexa-rotors) by the use of intermittent measures of gyroscopes with the goal to reduce the energy consumption and to gain in the autonomy of the battery.

6.6. Stability and control design of asynchronous interconnected systems

6.6.1. New approaches for stability analysis of time-delay systems

Participants: A. Seuret [Contact person], F. Gouaisbaut.

A particular attention has been paid to the stability analysis of time delay systems. Indeed delays represent a classical phenomenon which appears in networked control systems cite. This corresponds to the fact that data are not transmitted instantaneously from one node to its neighbors. In this context some effort has been provided in order to reduce the conservatism of the stability conditions. This works represents some fundamental researches to develop accurate stability conditions to networked control systems. More especially we produced a paper [45] which addresses the stability problem of linear time delay system. In the literature, the most popular approach to tackle this problem relies on the use of Lyapunov-Krasovskii functionals. Many results have proposed new functionals and techniques for deriving less and less conservative stability conditions. Nevertheless, all these approaches use the same trick, the well-known Jensen's inequality which generally induces some conservatism difficult to overcome. In light of those observations, we propose to reduce the conservatism of Lyapunov-Krasovskii functionals by introducing new classes of integral inequalities called Wirtinger's inequalities. This integral type inequality is firstly shown to encompass Jensen's inequality and is then employed to derive new stability conditions. To this end, a slightly modified Lyapunov functional is proposed. Several examples illustrate the effectiveness of our methodology. Further efforts on this topics have been provided and several improved articles are now submitted to servals journals.

6.6.2. Stability and control of asynchronous sampled-data systems

Participants: A. Seuret [Contact person], C. Briat [ETHZ], J. Gomes Da Silva Jr. [UFRGS], M. M. Peet [Illinois Institute of Technology].

Sampled-data systems have been extensively studied in the literature and the references therein. It is now reasonable to design controllers which guarantee the robustness of the solutions of a closed-loop system under periodic samplings. However the case of asynchronous samplings still leads to several open problems. This corresponds to the realistic situation where the difference between two successive sampling instants is time-varying. Several articles drive the problem of time-varying periods based on a discrete-time approach, input delay approach using the framework of Lyapunov-Krasovskii theorem, using the small gain theorem or the analysis of impulsive systems. These last approaches are very relevant to this problem because they cope with time-varying sampling periods as well as with uncertain systems in a simple manner. Nevertheless, these sufficient conditions are still more conservative than discrete-time approaches. In [24], we proposed a novel approach to assess stability of continuous linear systems with sampled-data inputs. The method, which is based on a particular type of functionals, called 'looped-functionals' provided easy tractable stability conditions for the continuous-time model. This method has been extended to various cases dealing with sampled-data systems. Indeed a method to constructs such class of functionals using the Sum of Squares framework was developed in [23]. Another extensions was also proposed in order to include saturations in the actuators [21]. Extensions to the case of communication delays and asynchronous samplings was also provided in [22].

Based on this new type of Lyapunov functional, several works have been provided in the more general context of hybrid system. Indeed sampled-data systems can be seen as a particular type of hybrid systems. This has been provided in several study done by A.R. Teel, Dragan Netic and many other researchers. Thus the idea was to show that the previous approach was also able to provide efficient stability conditions for impulsive systems [16], [27], [28], [54] or switched systems [53].

6.6.3. Event-based control

Participants: A. Seuret [Contact person], N. Marchand [Gipsa-Lab], C. Prieur [GIPSA-Lab], S. Durand [CINVESTAV].

Usually feedback laws are implemented in a periodic fashion on digital hardware. The main reason for using this periodicity in the hardware comes from the difficulties to analyze the stability of aperiodic or asynchronous systems. However it also seems natural to hold the same control input longer if the system behaves in a suitable way or shorter if the system requires an updated input. In [9], an algorithm is suggested to sample the control input based on the behavior of a Lyapunov-like function. This algorithm is called event-triggered since the Lyapunov-like function directly depends on the state of the systems. Using a Lyapunov-like function, two algorithms for the design of event-triggered algorithm are designed. It is assumed that a stabilizing controller for the continuous control system is given. Both event-triggered algorithms need to consider a closed-loop system with a mixed discrete/continuous dynamics (namely this is a hybrid system). Some numerical simulations illustrate the stability properties of both algorithms. In a future work, the performance issue should be analyzed. It is remarked that the event-triggered algorithms have a different performance. The first one seems to ensure a good speed of convergence on numerical simulations, whereas the second event-triggered algorithm allows less jumps and thus needs to compute less often the control variables. The advantages and disadvantages of each algorithm will be studied more precisely in a future work, for a theoretical point of view (e.g. by estimating a priori the number of switches), or on applications (to understand which algorithm is better depending on the application). Regarding this remark a journal paper has been submitted to IMA Journal of Mathematic Control and Information lately in 2012.

6.6.4. Feedback under slacken real-time

Participants: D. Simon [Contact person], A. Seuret, P. Andrianiaina [AIRBUS].

Robustness in control usually deals with the plant's parameter uncertainties, but the insensitivity or adaptability w.r.t. timing deviations from the theoretical pattern, such as jitter or deadlines misses can be exploited. The interesting point is that a feedback control system which is robust w.r.t. the plants parameters uncertainties is also robust, to some extent, w.r.t. timing deviations. Hence a feedback control system is not as hard as it is often considered in the literature, but should be better considered as *weakly hard*, i.e. able to tolerate specified timing deviations without leaving its requested performance [46].

A weakened implementation scheme for real-time feedback controllers is proposed to reduce the conservatism due to traditional worst-cases considerations. To save wasted computing resources, new real-time scheduling scenarios allowed for reducing the time slots allocated to control tasks below the value of the Worst Case Execution Time which is traditionally used to implement embedded control software. The stability of the control system under occasional deadlines miss is assessed using robustness arguments, using Lyapunov-Krasovskii functionals and LMIs solving based on [10]. The methodology has been successfully assessed for a fighter aircraft pitch controller, which show that the stability of the plant can be kept (and even improved) using the new scheduling schemes using less computing resources than traditional implementations [25], [11].

6.6.5. Varying sampling for LPV systems

Participants: D. Simon [Contact person], O. Sename, E. Roche.

In the context of network-controlled systems the idea of using varying control intervals naturally arises when the available computing power devoted to feedback control is limited, e.g. in embedded systems. It can be easily shown that decreasing the control frequency directly decreases the amount of computing needed for control. However, the stability of the feedback controller under varying sampling must be assessed for all the allowed variations of the sampling intervals [8].

The Linear Fractional Transform (LFR) formulation is widely used in robust analysis to study the influence of the plant's uncertain parameters on the stability and performances of a closed-loop system. Usually it is used to build a parameter dependent model of a dynamical system, depending on a known set of parameters. Here the set of varying parameters has been extended with the sampling interval of the control system, thus allowing to handle both varying sampling and plants uncertainties in a single framework (Figure 6).

Here, P_d is a on-line discretized model of the plant, Δ represents the uncertain parameters of the plant and δ is the variation of the sampling interval around its nominal value. From this model a robust controller can be synthesized, enforcing the control system stability for all variations of the sampling interval inside a predefined range.

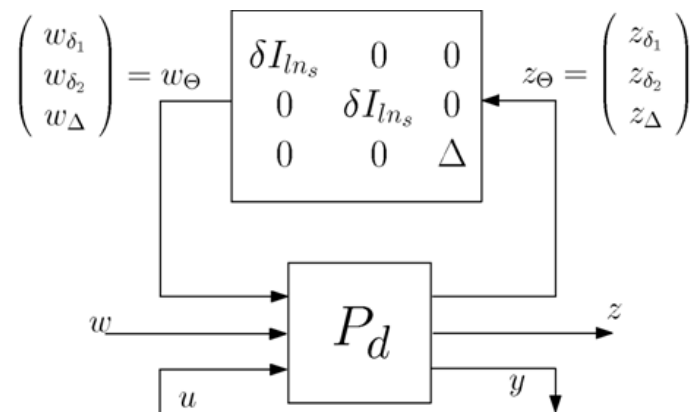


Figure 6. LFR system depending on system parameters and sampling interval variations

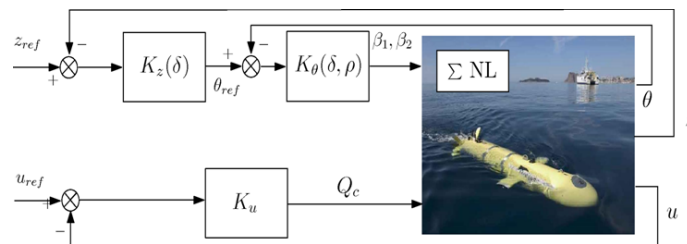


Figure 7. Control of an AUV

The approach have been successfully applied to the pitch and altitude control of a non-linear autonomous underwater vehicle, where the source of sampling variations comes from the altitude ultrasonics sensors [43]. However the approach still suffers from conservatism for which improvements using full block multipliers, or parameter-dependent Lyapunov functions, have been investigated [58].

6.7. Vehicular transportation systems

6.7.1. 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 [30], 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. Simulation results exhibit that the measured densities obtained from the traffic simulator Aimsun and the estimated densities agree closely. In [69], 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 [68].

6.7.2. Traffic control

Participants: C. Canudas de Wit [Contact person], D. Pisarski.

The problem of equilibrium points for the Cell Transmission Model was studied in [42]. The structure of equilibrium sets was analyzed in terms of model parameters and boundary conditions. The goal was to determine constant input flows, so that the resultant steady state of vehicle density was uniformly distributed along a freeway. The necessary and sufficient conditions for the existence of one-to-one relation between input flow and density were derived. The equilibrium sets were described by formulas that allow to design a desired balanced density. A numerical example for the case of a two-cell system was presented. In [41], the problem of optimal balancing of traffic density distributions was explored. The optimization was carried out over the sets of equilibrium points for the Cell Transmission Traffic Model. The goal was to find the optimal balanced density distribution, that maximizes both the Total Travel Distance and the total input flow. The optimization was executed in two steps. At the first step, a nonlinear problem to find a uniform density distribution that maximizes the Total Travel Distance was solved. The second step was to solve a quadratic problem reflecting the trade-off between density balance and input flow maximization. At both steps, decomposition methods were used. The computational algorithms associated to such a problem were given. Finally, in [71], the application of the idea of optimal balancing of traffic density distribution was presented. It was implemented to the Grenoble South Ring in the context of the Grenoble Traffic Lab. The traffic on the ring is represented by the Cell Transmission Model that was tuned by using real data and Aimsun micro-simulator. A special attention was paid to the calibration of a flow merging model. A large-scale optimization problem was solved by using advanced combinatorial procedures. The main difficulties in the implementation as well as the limitations of the designed software were highlighted. Finally, the results of different traffic scenarios on the Grenoble South Ring were presented.

6.7.3. Vehicle control for disabled people

Participants: C. Canudas de Wit [Contact person], V. Ciarla, J. Dumon, F. Quaine [UJF], V. Cahouet [UJF].

The typical architecture of an Electric Power Assistance Steering (EPAS) system includes a static map to provide the correct amplification to the drivers exerted torque. In literature, it is generally known as booster curve. This work concerns the study of the amplification criteria, that are commonly used to these booster curves. The basic concepts of the Electric Power Steering (EPS) systems with a realistic model for the friction contact, that acts on the wheels are discussed. A relation between the assistance and the driver's torque is provided, under the hypothesis of a position-oriented control of the movement and the Stevens' power law [33]. In current works, we want to modify the general architecture of the EPAS system for people driving with two arms. For this purpose, we insert two additional blocks: the first one provides an estimation of the gravitational torque due to the weight of the driver's arm while the second gets as inputs the total driver's torque and the estimated gravitational torque in order to update the driver's torque with the gravitational torque. The updated measure is then given as input to the booster curve for deriving the correct assistance.

6.7.4. 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, corresponding to different levels of energy consumption; therefore, there is 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, but 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 travel time of vehicles in the road network. If all the vehicles can communicate with one another and with the active infrastructure (traffic lights), we can imagine benefits for each of the two problems which can be considered as a whole: on the one hand, for vehicles, 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. More importantly, the traffic manager can send instructions to the vehicle. The aim of the research is to evaluate the potential in terms of energy saving and traffic improvement made possible by communicating vehicle. This work is carried out in collaboration with IFP in the framework of a CIFRE thesis.

OPALE Project-Team

6. New Results

6.1. Mathematical analysis and control of macroscopic traffic flow models

6.1.1. Vehicular traffic

Participants: Maria Laura Delle Monache, Paola Goatin, Mauro Garavello [Piedmont University, Italy], Alexandre Bayen [UC Berkeley, CA, USA].

The activity in traffic flow modeling has been reinforced by the creation of the Associated Team ORESTE between OPALE and the UC Berkeley teams Mobile Millennium and Integrated Corridor Management (ICM) lead by Prof. A. Bayen (see <http://www-sop.inria.fr/members/Paola.Goatin/ORESTE/index.html>). In this framework, three PhD students from US visited Inria during August and September, and M.L. Delle Monache spent two and half months at UC Berkeley.

During this first year of common research we proposed a new junction model for ramp-metering in the continuous and discrete settings. We focused on a junction consisting in a mainline, an on-ramp and an off-ramp. In particular, we introduced a coupled PDE-ODE model, in which the PDE describes the evolution of the cars flow on the mainline 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. At the junction we imposed the maximization of the outgoing flux together with a fixed priority parameter for incoming roads. We were able to prove existence and uniqueness of the solution of the corresponding Riemann problem. This approach has then been extended to networks and discretized using the Godunov scheme. The corresponding discrete optimization problem has been solved using the Adjoint Method and it is now being implemented into a MATLAB code. This model will serve as starting point for a subsequent model for optimal rerouting, which includes multi-commodity flow and partial control.

Besides that, we studied a coupled PDE-ODE system modeling the interaction of a large slow moving vehicle with the surrounding traffic flow. The model consists in a scalar conservation law with moving density constraint describing traffic evolution coupled with an ODE for the slow vehicle trajectory. The constraint location moves due to the surrounding traffic conditions, which in turn are affected by the presence of the slower vehicle, thus resulting in a strong non-trivial coupling. The existence result is given in [60].

The paper [41] is devoted to the study of a traffic flow model on a network composed by an arbitrary number of incoming and outgoing arcs connected together by a node with a buffer. We define the solution to the Riemann problem at the node and we prove existence and well posedness of solutions to the Cauchy problem.

6.1.2. Crowd motion

Participants: Nora Aïssiouene, Christophe Chalons [LJLL, UP7], Régis Duvigneau, Paola Goatin, Matthias Mimault, Massimiliano D. Rosini [ICM, Warsaw University, Poland], Nicolas Seguin [LJLL, UPMC], Monika Twarogowska.

The activity on in pedestrian flow modeling is reinforced by the doctoral thesis of M. Mimault, started in October, and the enrollment of M. Twarogowska on a post-doctoral position.

Concerning crowd motion modeling, we are interested in the optimization of facilities design, in order to maximize pedestrian flow and avoid or limit accidents due to panic situations. To this aim, we are now studying first and second order macroscopic models for crowd movements consisting in one or two scalar conservation law accounting for mass conservation and momentum balance, coupled with an Eikonal equation giving the flux direction depending on the density distribution. From the theoretical point of view, and as a first step, we are studying the problem in one space dimension (for applications, this case corresponds to a crowd moving in a corridor). In collaboration with M. Rosini (supported by the project CROM3, funded by the PHC Polonium 2011), we have established entropy conditions to select physically relevant solutions,

and we have constructed explicit solutions for some simple initial data (these results are presented in [40]). We are now studying existence of solutions of the corresponding initial boundary value problem, using the wave-front tracking approach. In this framework, M. Mimault's internship was devoted to develop a MATLAB code based on wave-front tracking to compute the solutions of Hughes' model of pedestrian motion with generalized running cost. This model displays a non-classical dynamic at the splitting point between the two directions of motion. The wave-front tracking scheme provides us with reference solutions to test numerically the convergence of classical finite volume schemes, which do not treat explicitly the dynamics at the turning point (see [66]). The code can be downloaded at the following URL: <http://www-sop.inria.fr/members/Paola.Goatin/wft.html>

From the numerical point of view, we are implementing some macroscopic models in two space dimensions on triangular meshes on the Num3sis platform. This was partly done by N. El-Khatib (postdoc at Inria from January to August 2011), and is now being completed by M. Twarogowska, with the support of N. Aïssiouene. This will provide a performing numerical tool to solve the related optimization problems arising in the optimization of facilities design, such as the position and size of an obstacle in front of (before) a building exit in order to maximize the outflow through the door and avoid or limit over-compression.

Finally, in collaboration with C. Chalons and N. Seguin, we have generalized the results on conservation laws with local flux constraint obtained in [3], [5] to general flux functions and nonclassical solutions arising for example in pedestrian flow modeling. We first define the constrained Riemann solver and the entropy condition, which singles out the unique admissible solution. We provide a well posedness result based on wave-front tracking approximations and Kruzhkov doubling of variable technique. We then provide the framework to deal with nonclassical solutions and we propose a "front-tracking" finite volume scheme allowing to sharply capture classical and nonclassical discontinuities. Numerical simulations illustrating the Braess paradox are presented as validation of the method. The results are collected in [65].

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 discipline, such as structural mechanics.

Our approach to *competitive optimization* 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. This approach has been demonstrated over a test-case of aero-structural aircraft wing shape optimization, in which the eigensplit-based optimization reveals clearly superior [38].

While the two-discipline method is currently being applied to various complex physical multi-objective situations (see in particular 6.2.4 , 6.2.5 , 6.2.6), the method has been extended to situations involving more than two objectives when the initial point is Pareto-optimal. Then, a particular convex combination of the criteria is locally stationary, and the two-discipline strategy can be applied using this combination as preponderant criterion, and a particular other criterion as secondary one. Whence, the proposed split of territory produces a continuum of Nash equilibrium points *tangent* to the Pareto set. This theoretical result has been illustrated in the context of a simpler numerical experiment by E. Baratchart during his internship [4], see Fig. 2 .

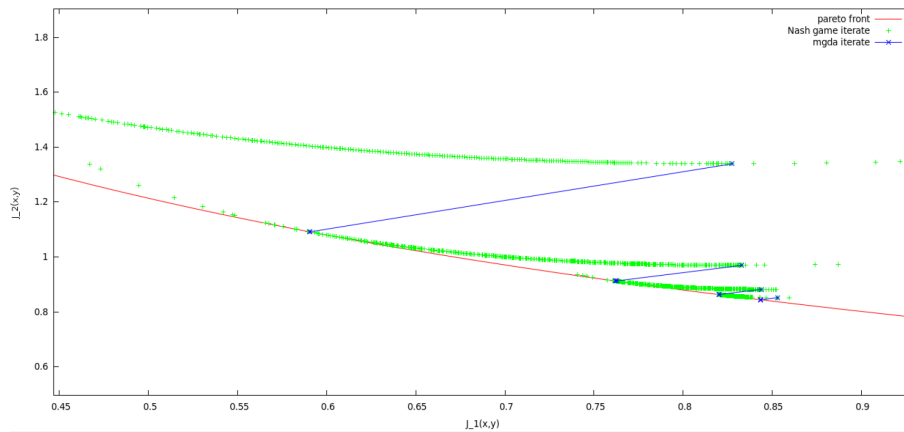


Figure 2. Combination of cooperative and competitive optimization algorithms: in red the Pareto set, in blue MGDA steps directed to the Pareto set, in green steps by Nash games with split of territory tangent to the Pareto set.

Our approach to *cooperative optimization* is based on a result of convex analysis established for a general unconstrained multi-objective problem in which all the gradients are assumed to be known. The theorem [39] 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-optimal. 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 point on the Pareto set, at which a competitive optimization phase can possibly be launched on the basis of the local eigenstructure of the different Hessian matrices. This general formulation fosters several connected studies detailed in 6.2.1 .

6.2.1. Multiple-Gradient Descent Algorithm (MGDA)

Participants: Jean-Antoine Désidéri, Régis Duvigneau, Matteo Giacomini, 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.

In the original report [14], and in [39], we have introduced the notion of *Pareto-stationarity*, and given a first proof that it was the natural necessary condition for Pareto-optimality when the objective-functions are locally smooth in some open domain about the design-point. This report has been revised to provide a more rigorous,

and extended proof. In particular, in the revised version [14] (version 3, 2012), the number of objective-functions n and the dimension of the design space compare arbitrarily. The objective-functions are assumed to be locally convex.

Additionally, we had established that MGDA converges to Pareto-stationary design-points. This had been confirmed by numerical experiments in which MGDA had been tested over a number of classical multi-objective optimization test-cases, and found successful to converge to Pareto-optimal solutions in situations of either convex or concave Pareto sets. Additionally, MGDA [57] and PAES [69] were found to have complementary merits, making a hybrid method promising.

The method was tested successfully in a domain partition model problem in which the sub-solutions to the Poisson equation are matched at the interfaces by minimization of the integral along the interface of the squared normal-derivative jump. This academic exercise has permitted to illustrate the importance of applying an appropriate scaling to the gradients prior to calculating the descent direction [61] [47]. This has led us to define, a novel form of MGDA, consisting of a direct algorithm [62] based on a Gram-Schmidt orthogonalization conducted with a special normalization. The direct method was found more accurate and more efficient. Subsequently, we proposed two enhancements [63], the first to define the order in which the gradients are introduced in the Gram-Schmidt process uniquely and to interrupt the process as soon as the current estimate of the search direction is proved to satisfy the descent property, and the second to optimally scale the gradients when the Hessians are known, or approximated (e.g. by the BFGS estimate).

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. 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 [55]. This more difficult problem has been proposed by Renault within the OMD2 ANR project.

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 6.6) 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: Jean-Antoine Désidéri, Régis Duvigneau, Jérémie Labroquère.

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, a numerical study is conducted to i) improve the understanding of controlled flows ii) develop a methodology to determine optimal control parameters by coupling the controlled flow simulation 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[37].

In this perspective, unsteady Reynolds Averaged Navier-Stokes equations are considered, with the Spalart-Allmaras turbulence closure. A numerical model for synthetic jets has been implemented to simulate the actuation[48], based on imposed velocity boundary conditions. Particular developments have then be carried out to include a noise term into Gaussian Process metamodels, which is used to filter errors arising from unsteady simulations/citelabroquere:hal-00742940. First results have demonstrated the feasibility of the proposed method. A systematic assessment of modeling and numerical errors is in progress, 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, Daigo Maruyama.

This work aims to develop robust design tools for aircraft design 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 consisting of minimizing statistical quantities such as mean and variance of a cost function, typically the drag coefficient under lift constraint. MGDA is used for this purpose.

At present, analytical test cases have been tested, confirming the validity of our approach to identify the Pareto set.

One aspect of the problem is that the evaluation of these statistics and performing their optimization is very cost demanding. One solution could be, for aerodynamic design, to identify the most important variables to be treated as uncertain, possibly by the ANOVA approach, and construct adequate meta-models.

6.2.4. Aero-structural optimization

Participants: Gérald Carrier [Research Engineer, ONERA/DAAP], Jean-Antoine Désidéri, Imane Ghazlane.

In industry, aircraft wings are designed by accounting for several multidisciplinary couplings. Certainly of greatest importance is the coupling, or concurrency, between aerodynamic optimization and structural design. At ONERA, in the former thesis of M. Marcelet, the aerodynamic gradient has been extended to account for (the main terms of) static fluid-structure interaction, commonly referred to as the “aeroelastic gradient”.

In her thesis, I. Ghazlane has extended M. Marcelet’s work to take into account, in the aeroelastic gradient, the terms originating from the differentiation of the wing-structural model. In this development, the wing structure is treated as an equivalent Euler-Bernoulli beam. These formal extensions have been validated by an extensive experimentation. Additionally, special post-processing procedures have been set up to evaluate accurately the various physical contributions to drag. As a result, a realistic aircraft wing optimization has been conducted using a configuration provided by Airbus France as initial design. I. Ghazlane defended successfully her doctoral thesis in December 2012 [34].

Besides, I. Ghazlane has realized a two-objective optimization (drag and mass reduction) via a Nash game using our optimization platform FAMOSA. These results will be included in a common publication on Nash games in preparation.

6.2.5. Sonic boom reduction

Participants: Gérald Carrier [Research Engineer, ONERA/DAAP], Jean-Antoine Désideri, 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 [68] for the calculation of the Whitham function corresponding to the lowest-boom (axisymmetric) shape. Method and results for more general geometries have been presented internationally in [50].

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. These results will be included in a common publication on Nash games in preparation.

6.2.6. Helicopter rotor blade optimization in both situations of hovering and forward flight

Participants: Michel Costes [Research Engineer, ONERA/DAAP], Jean-Antoine Désideri, 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. In the current work, 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 the twist angle along the wing is 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 a forthcoming AIAA Conference, and will also contribute to a common publication on Nash games in preparation.

6.2.7. Optimum design in naval hydrodynamics

Participants: Régis Duvigneau, Louis Blanchard.

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.

In collaboration with GALAAD Project-Team, 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[58]. 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.

6.3. Optimum design in structural mechanics

6.3.1. Shape Optimization in Multidisciplinary Non-Linear Mechanics

Participants: Aalae Benki, Jean-Antoine Désidéri, Abderrahmane Habbal.

In collaboration with the ArcelorMittal's Center for Research in Automotive and Applications, 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. In both cases, only arbitrary territory splitting was used. Application to large scale 3D industrial problems, and the study of intelligent territory splitting algorithms is ongoing.

6.3.2. Optimization of Addendum Surfaces in Stamping

Participants: Fatima Zahra Oujebbour, Jean-Antoine Désidéri, Abderrahmane Habbal.

Within the OASIS Consortium (ArcelorMittal, ErDF, Inria, UTC, EURODECISION, ESILV, NECS, Delta-CAD, SCILAB-DIGITEO), Opale Project 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 thickness distribution.

We have successfully applied the NBI approach coupled to the hybrid SA+SPSA minimizer (Simulated Annealing with local search using the Simultaneous Perturbation Stochastic Approximation) to capture the Pareto front of a simple cross stamping of a high performance steel sheet. The use of cubic spline approximation of the costs (spring-back and failure criteria) turned out to be more reliable than e.g. a krigage method.

6.4. Application of shape and topology design to biology and medicine

6.4.1. Mathematical modeling of dorsal closure DC

Participants: Abderrahmane Habbal, Luis Almeida [University of Nice-Sophia Antipolis], Patrizia Bagnerini [Genova University], Fanny Serman [University of Nice-Sophia Antipolis], Stéphane Noselli [University of Nice-Sophia Antipolis], Glenn Edwards [Duke University].

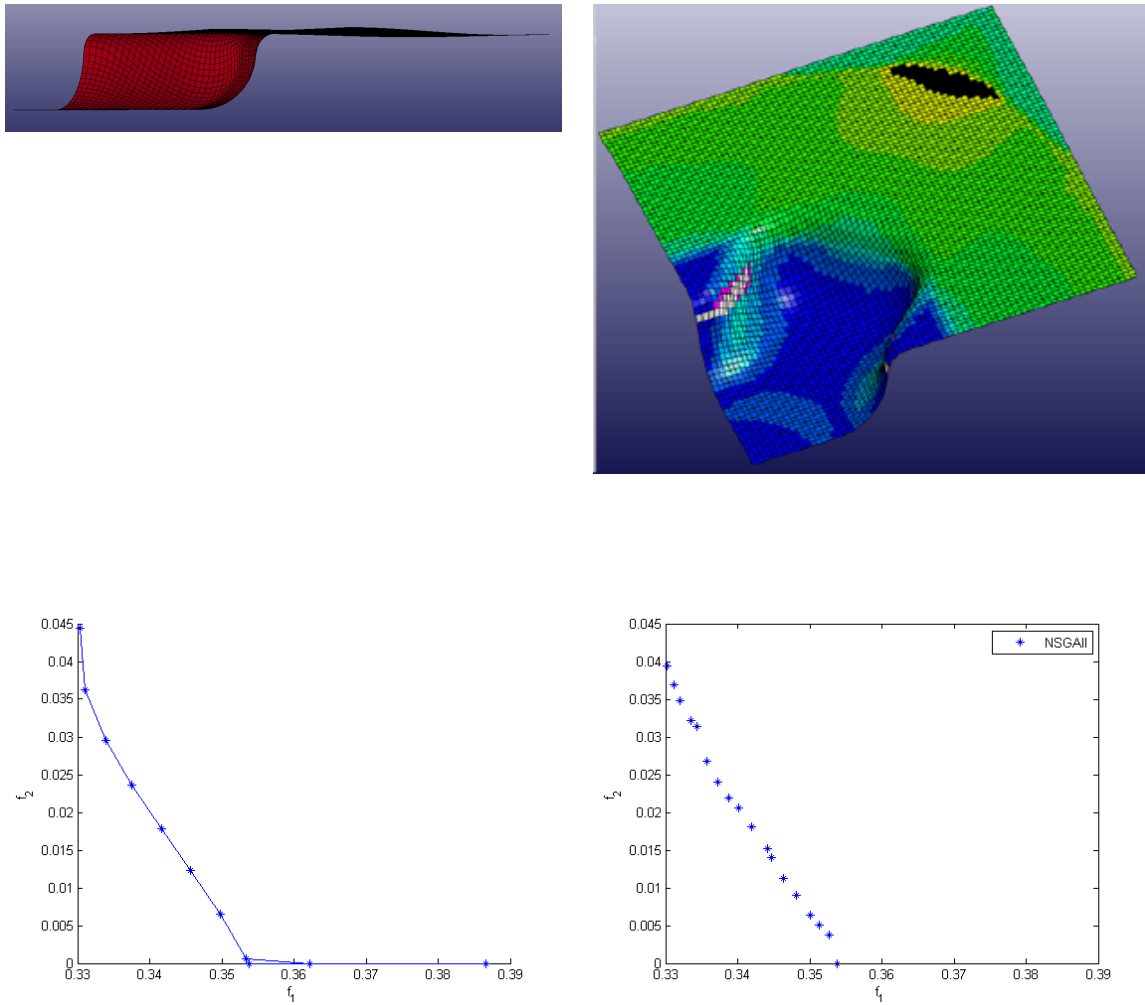


Figure 3. Multiobjective design of the stamping process of a high performance steel sheet. The costs are elastic spring-back (upper-left) and failure (upper-right). The Pareto front obtained by NNCM (lower-left) is compared to a NSGA-II one (lower-right).

A mathematical model for simulation of actin cable contraction, during wound closure for *Drosophila* embryo, which contains an extra term in addition to the curvature flow is developed. The basic mathematical model introduced and validated in [2] is extended in order to include the non-homogeneous wound healing or non-homogeneous dorsal closure. The new model is obtained by adding extra terms that describe the particular process we want to model (lamellipodial crawling, granulation tissue contraction, extension of actin protrusions, epithelial resistance, etc.). We concentrate on the treatment of non-homogeneous forces, i.e. non-constant boundary terms which can be associated with a non-uniform cable, internal pull or zipping force due to the non-uniformity of the biological or physical properties of the boundary cells or of the connective tissue [35].

We also consider a particular yet major aspect of wound healing, namely the one related to the movement of wounded epithelial cell monolayers. The epithelial monolayer cell population, also referred to as cell-sheet, can be seen as a 2 dimensional structure, although it is well known that apical and basal sites play distinctive important roles during the migration, as well as the substrate itself. Immediately after a wound is created, the cells start to move in order to fill in the empty space. This movement, the wound closure, is a highly-coordinated collective behavior yielding a structured cohesive front, the wound leading edge. Even though wound closure involves biochemical and biomechanical processes, still far from being well understood, which are distributed over the whole monolayer, much specific attention was paid to the leading edge evolution, seen as the front of a traveling wave of the cell density function. We show that, for non inhibited wound assays, closure occurs at constant speed of the leading edge, a fact that is commonly shared by biologists and biomathematicians. But we also show that the leading edge may exhibit accelerated profiles, and that when inhibited, then the F-KPP has poor performances in modeling the leading edge dynamics.

6.5. Particular applications of simulation methods

6.5.1. Hermitian interpolation under uncertainties

Participants: Jean-Antoine Désideri, Manuel Bompard [Doctoral Student, ONERA/DSNA until December 2011; currently post-doctoral fellow in Toulouse], Jacques Peter [Research Engineer, ONERA/DSNA].

In PDE-constrained global optimization, iterative algorithms are commonly efficiently accelerated by techniques relying on approximate evaluations of the functional to be minimized by an economical, but lower-fidelity model (meta-model), in a so-called Design of Experiment (DoE). Various types of meta-models exist (interpolation polynomials, neural networks, Kriging models, etc). Such meta-models are constructed by pre-calculation of a database of functional values by the costly high-fidelity model. In adjoint-based numerical methods, derivatives of the functional are also available at the same cost, although usually with poorer accuracy. Thus, a question arises : should the derivative information, available but known to be less accurate, be used to construct the meta-model or ignored ? As a first step to investigate this issue, we have considered the case of the Hermitian interpolation of a function of a single variable, when the function values are known exactly, and the derivatives only approximately, assuming a uniform upper bound ε on this approximation is known. The classical notion of best approximation has been revisited in this context, and a criterion introduced to define the best set of interpolation points. This set was identified by either analytical or numerical means. If $n + 1$ is the number of interpolation points, it is advantageous to account for the derivative information when $\varepsilon \leq \varepsilon_0$, where ε_0 decreases with n , and this is in favor of piecewise, low-degree Hermitian interpolants. In all our numerical tests, we have found that the distribution of Chebyshev points is always close to optimal, and provides bounded approximants with close-to-least sensitivity to the uncertainties [56].

6.5.2. Mesh qualification

Participants: Jean-Antoine Désideri, Maxime Nguyen, Jacques Peter [Research Engineer, ONERA/DSNA].

M. Nguyen Dinh is conducting a CIFRE thesis at ONERA supported by AIRBUS France. The thesis topic is the qualification of CFD simulations by anisotropic mesh adaptation. Methods for refining the 2D or 3D structured mesh by node movement have been examined closely. Secondly, it is investigated how could the local information on the functional gradient $\|dJ/dX\|$ be exploited in a multi-block mesh context. This raises particular questions related to conservation at the interfaces.

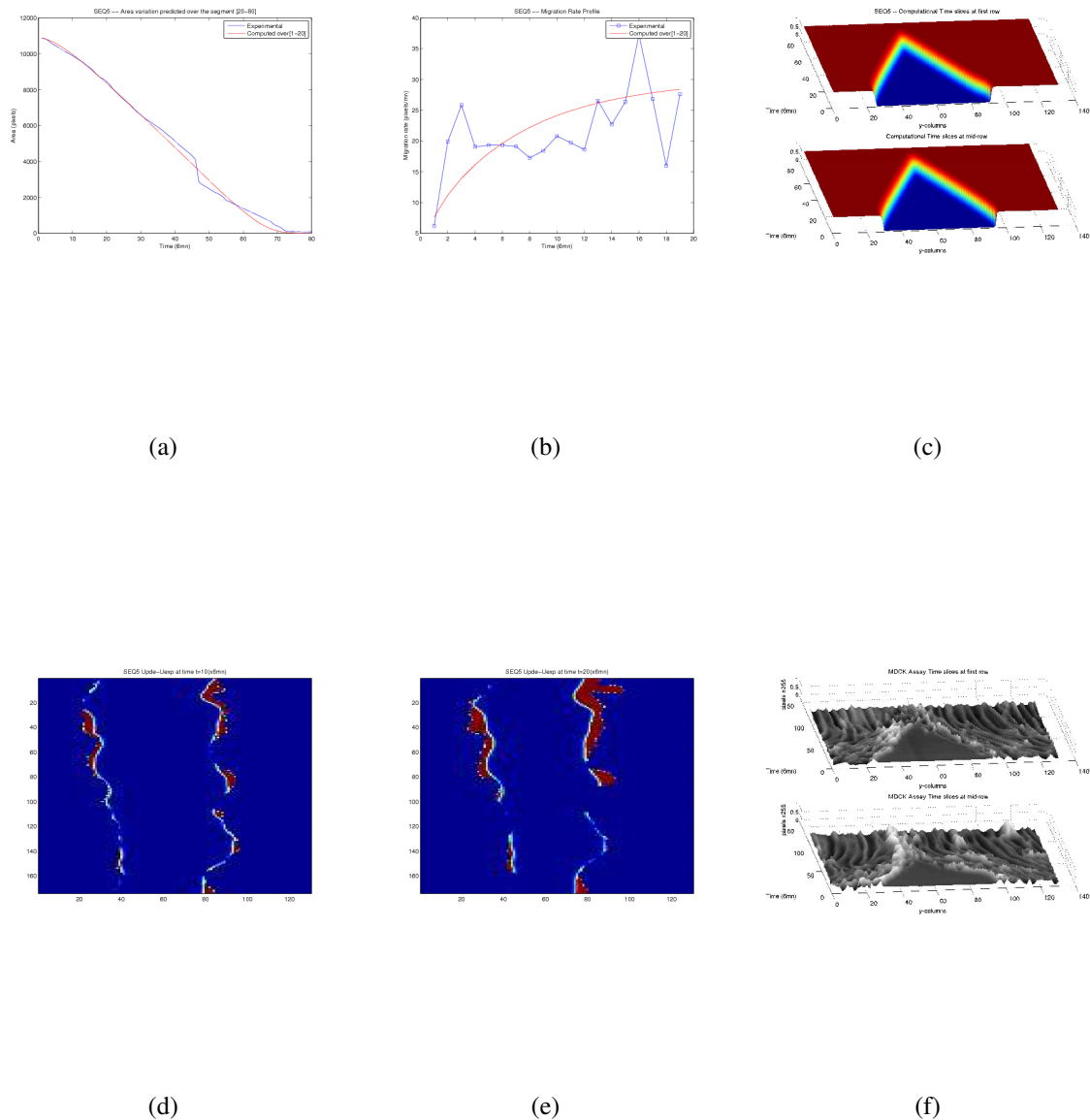


Figure 4. Sequence-5. Computational vs experimental wound evolution. (a) Time variation of experimental (blue) versus computed (red) wound area (in pixels). (b) Time variation of the experimental (blue-dot) versus computed (red) migration rate (in pixels/mm). (c) Computed 3D XT view at first and mid-rows. (d) (e) (f) Traces of the difference between the experimental segmented and binarized cell-sheet images and the computed ones at different times, respectively 1hour (d), and 2hours (e) after the wounding. (f) Experimental 3D XT view at first and mid-rows.

Several criteria have been assessed for mesh qualification in the context of inviscid-flow simulation and are currently being extended to the RANS context. These results have been presented internationally in the communication [54] and the publication [44].

6.5.3. Hybrid meshes

Participants: Sébastien Bourasseau, Jean-Antoine Désideri, Jacques Peter [Research Engineer, ONERA/DSNA], Pierre Trontin [Research Engineer, ONERA/DSNA].

S. Bourasseau has started a CIFRE thesis at ONERA supported by SNECMA. The thesis is on mesh adaption in the context of hybrid meshes, that is, made of both structured and unstructured regions. Again, the aim is to exploit at best the function gradient provided by the adjoint-equation approach. Preliminary experiments have been conducted on geometries of stator blade yielding the sensitivities to global shape parameters.

The on-going developments are related to the extension to the hybrid-mesh context of the full shape gradient in a 3D Eulerian flow computation.

6.5.4. Data Completion Problems Solved as Nash Games

Participants: Abderrahmane Habbal, Moez Kallel [University of Tunis].

The Cauchy problem for an elliptic operator is formulated as a two-player Nash game.

- Player (1) is given the known Dirichlet data, and *uses as strategy variable the Neumann condition* prescribed over the inaccessible part of the boundary.
- Player (2) is given the known Neumann data, and *plays with the Dirichlet condition* prescribed over the inaccessible boundary.
- The two players solve in parallel the associated Boundary Value Problems. Their respective objectives involve the *gap between the non used Neumann/Dirichlet known data and the traces of the BVP's solutions* over the accessible boundary, and are *coupled through a difference term*.

We prove the existence of a unique Nash equilibrium, which turns out to be the reconstructed data when the Cauchy problem has a solution. We also prove that the completion algorithm is stable with respect to noise. Many 3D experiments were performed which illustrate the efficiency and stability of our algorithm [42].

6.6. Isogeometric analysis and design

Participants: Louis Blanchard, Régis Duvigneau, Bernard Mourrain [Galaad Project-Team], Gang Xu [Galaad Project-Team].

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 all design tasks:

- the geometry is defined by NURBS surfaces;
- the computation domain is defined by NURBS volumes instead of meshes;
- the solution fields are obtained by using a finite-element approach that uses NURBS basis functions
- the optimizer controls directly NURBS control points.

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 GALAAD Project-Team, such as the development of methods for adaptive parameterization including an a posteriori error estimate[46], [45]. A collaborative work has also been carried out with the Technical University of Kaiserslautern, concerning the computation of shape gradients for linear elasticity problems[59].

BAMBOO Project-Team

6. New Results

6.1. Partial Enumeration of Traces

Traditional algorithms to solve the problem of sorting by signed reversals output just one optimal solution while the space of all optimal solutions can be huge. A so-called trace represents a group of solutions which share the same set of reversals that must be applied to sort the original permutation following a partial ordering. By using traces, we therefore can represent the set of optimal solutions in a more compact way. Algorithms for enumerating the complete set of traces of solutions were developed. However, due to their exponential complexity, their practical use is limited to small permutations. A partial enumeration of traces is a sampling of the complete set of traces and can be an alternative for the study of distinct evolutionary scenarios of big permutations. Ideally, the sampling should be done uniformly from the space of all optimal solutions. This is however conjectured to be #P-complete.

We proposed and evaluated three algorithms for producing a sampling of the complete set of traces that instead can be shown in practice to preserve some of the characteristics of the space of all solutions [7]. We analysed the distribution of the enumerated traces with respect to their height and average reversal length.

6.2. De-novo calling alternative splicing events from RNA-seq data

We addressed the problem of identifying and quantifying polymorphisms in RNA-seq data when no reference genome is available, without assembling the full transcripts. Based on the fundamental idea that each polymorphism corresponds to a recognisable pattern in a De Bruijn graph constructed from the RNA-seq reads, we proposed a general model for all polymorphisms in such graphs. We then introduced an exact algorithm, called KISSPLICE, to extract alternative splicing events. The first version of KISSPLICE appeared in 2011, but several important improvements were implemented in 2012 [24]. The first improvement was the memory consumption, the new version is much more memory efficient and can handle datasets of approximately 10^8 reads. The second was in the running time, the enumeration step can now be done in parallel, which results in a significant speedup in the overall running time. Finally, an improved event quantification step was added to the method.

Application-wise, we showed that KISSPLICE enables to identify more correct events than general purpose transcriptome assemblers. Additionally, on a 71 M reads dataset from human brain and liver tissues, KISSPLICE identified 3497 alternative splicing events, out of which 56% are not present in the annotations, which confirms recent estimates showing that the complexity of alternative splicing has been largely underestimated so far.

6.3. Efficient bubble and/or cycle enumeration in directed/undirected graphs

Polymorphisms in DNA- or RNA-seq data lead to recognisable patterns in a de Bruijn graph representation of the reads obtained by sequencing. Such patterns have been called mouths, or bubbles in the literature. They correspond to two vertex-disjoint directed paths between a source s and a target t . Due to the high number of such bubbles that may be present in real data, their enumeration is a major issue concerning the efficiency of dedicated algorithms. We developed the first linear delay algorithm to enumerate all bubbles with a given source [31].

By combining the insights from the most efficient but not optimal solution presented by Johnson [SIAM J. Computing, 1975] for simple cycle enumeration in undirected graphs and an amortisation technique previously established by our collaborators Roberto Grossi and Rui Ferreira [ESA, 2011] from the University of Pisa, Italy, we obtained the first optimal solution to list all the simple cycles in an undirected graph G (paper accepted at SODA 2013, to appear). Moreover, we also obtained the first optimal solution to list all the simple paths from s to t in an undirected graph G . This work benefited also from discussions and work from Pierluigi Crescenzi and Marie-France Sagot. The method is not naturally extendable to directed graphs, and the challenge is now to obtain optimal solutions in this case also.

6.4. Simulating RNA-seq experiments

RNAseq experiments now enable to characterise the RNA complement of a cell. However, the series of steps (fragmentation, reverse transcription, sequencing) that separate the initial RNA molecules from the short DNA reads obtained in fine are not well understood although it is widely accepted that they contribute to generating noise in the signal. We introduced the FLUXSIMULATOR [14], a computer program able to reproduce the biases seen in RNAseq data. This pipeline should prove useful both to produce realistic data on which to test programs which aim at reconstructing RNA from short reads, and suggest ways of improving the experimental steps so that they produce less noise.

6.5. Chimeric Transcripts may be Translated

There is now increasing evidence for the existence of so-called Chimeric Transcripts. In contrast to regular transcripts, which are composed of exons located close to each other on the genome, these chimeric transcripts can be composed of exons which are located megabases away, or even on different chromosomes. We showed that these chimeras are lowly expressed, are tissue specific, and that some of them may be translated, yielding proteins with altered function or localisation [13].

6.6. Transcriptomics of symbiosis in the *Asobara tabida*-*Wolbachia* association

Wolbachia has evolved a very peculiar phenotype in the host *Asobara tabida* where it is obligatory for oogenesis. Transcriptomics approaches were developed first using Sanger sequencing of mRNA [19]. It has now been complemented by RNAseq analyses on two lines, which exhibit different ovarian phenotypes in absence of *Wolbachia*. We have currently analysed these data both to isolate genes that are differentially expressed, but also that exhibit polymorphism between the lines. Interesting candidates were detected that are under further investigation and that are involved in the regulation of early oogenesis, apoptosis, autophagy and oxidative stress. This part is in direct connection with the algorithms developed by BAMBOO for the analysis of NGS data without a reference genome (KISSPLICE).

6.7. MicroRNA predictor

We developed a microRNA predictor using structural and target information. The method shows 97% sensitivity and 90% specificity for the *Acyrtosiphon pisum* genome. Comparing to the results of the previous method we developed in 2010 (available in the software CRAVELA) we obtained a better performance (sensitivity 90% and specificity 88%). However, as we are working on a genome wide scale, it is important to obtain even better specificity (obviously, maintaining a reasonable sensitivity). This work is currently in development.

On the other hand, the computational search for novel miRNA precursors often involves also some sort of structural analysis with the aim of identifying which type of structures are recognised and processed by the cellular miRNA-maturation machinery. A natural way to tackle this problem is to perform clustering over the candidate structures along with known miRNA precursor structures. Mixed clusters allows then the identification of candidates that are similar to known precursors. Given the large number of pre-miRNA candidates that can be identified in single-genome approaches, even after applying several filters for precursor robustness and stability, a conventional structural clustering approach is unfeasible. We proposed a method to

represent candidate structures in a feature space which summarises key sequence/structure characteristics of each candidate [21]. We showed that proximity in this feature space is related to sequence/structure similarity, and we selected candidates which have a high similarity to known precursors. Additional filtering steps were then applied to further reduce the number of candidates to those with greater transcriptional potential.

6.8. Genomics of symbiosis

Insect symbioses are model systems for studying the effect of symbionts and the evolution of bacterial genomes. Members from the LBBE described the symbiotic complement of different biotypes of the insect *Bemisia tabaci* in Western Africa. We further obtained the complete genome of different symbionts that co-exist in *Bemisia tabaci*, among which the primary symbiont *Portiera* [25], *Hamiltonella*, *Rickettsia* and *Wolbachia*. Analyses are underway, that concern the possible complementation between *Hamiltonella* and *Portiera* and the comparative analyses of different *Hamiltonella* genomes.

6.9. Representation and curation of metabolic data: UniPathway, Rhea and MNX

These activities are carried out in collaboration with the SwissProt group at the Swiss Institute for Bioinformatics (SIB). UNIPATHWAY (<http://www.unipathway.org>) is a manually curated database of metabolic pathways. It provides the official controlled vocabulary for pathway annotation within UNIPROTKB records since 2009. A complete description of the UNIPATHWAY database and of its relationship with UNIPROTKB has been published in *Nucleic Acids Research* (Jan. 2012 Database Issue) [22]. RHEA (<http://www.ebi.ac.uk/rhea>) is developed jointly with the European Institute for Bioinformatics (EBI) and the SIB. It provides a comprehensive resource of expert-curated biochemical reactions, for use in a large spectrum of applications, including metabolic network reconstruction and pathway inference. The complete description of the RHEA database appeared in the Jan. 2012 *NAR* Database issue [5]. The MNX project is developed in the context of the METANETX project (<http://www.metanetx.org>). It attempts to automate the reconciliation of discrepancies between metabolite or reaction information from distinct resources (BIGG, BRENDA, CHEBI/RHEA, KEGG, METACYC, UNIPATHWAY, THE SEED, REACTOME), thereby alleviating a major bottleneck in the construction of genome-scale metabolic network models. The MNXREF namespace is available at <http://www.metanetx.org/mnxdoc/mnxref.html> and the method to compute the MNXREF namespace is described in [8].

6.10. Annotation of the proteins of *Angomonas deanei* and *Strigomonas culicis*

Angomonas deanei and *Strigomonas culicis* are trypanosomatids that harbour only one beta-proteobacterial endosymbiont and this mutualistic association is an interesting model to study eukaryotic cell evolution. The genomes of these organisms were sequenced by our collaborators at LNCC / MCT (Brazil) and we participated in the functional annotation of these genomes as concerns their metabolism which enabled to reveal new aspects of the *Trypanosomatidae* family. This work has been submitted for publication. It was done with Ana Tereza Vasconcelos in a collaboration with Maria Cristina Machado Motta (UFRJ - Brazil).

6.11. Finding candidate genes for orphan enzymes

Of all biochemically characterized metabolic reactions formalized by the IUBMB, over one out of four have yet to be associated with a nucleic or protein sequence, *i.e.* are sequence-orphan enzymatic activities. Few bioinformatics annotation tools are able to propose candidate genes for such activities by exploiting context-dependent rather than sequence-dependent data, and none are readily accessible and propose result integration across multiple genomes. We introduced CANOE (Candidate genes for Orphan Enzymes), a four-step bioinformatics strategy that proposes ranked candidate genes for sequence-orphan enzymatic activities (or orphan enzymes for short) [26]. Our strategy found over 60,000 genomic metabolons in more than 1,000 prokaryote organisms from the MICROSCOPE platform developed by the group of Claudine Médigue from the Génoscope with whom this work was done, generating candidate genes for many metabolic reactions, of which more than 70 distinct orphan reactions. A computational validation of the approach was discussed and we presented a case study on the anaerobic allantoin degradation pathway in *Escherichia coli* K-12.

6.12. Metabolic cooperation of symbionts and their host trypanosomatids

Trypanosomatids that harbour a symbiotic bacterium (SHTs) are known to have less nutritional requirements when compared to their counterparts without symbionts (RTs). Nutritional and biochemical data indicated that the symbionts largely contributed to the routes for amino acid and vitamin biosynthesis. We analysed the genomic data of 5 SHTs and their respective symbionts and 2 RTs and we found most of the genes related to those pathways in the symbionts. This work will soon be submitted for publication. It is being done with Ana Tereza Vasconcelos in a collaboration with Erney P. Camargo, Marta M.G. Teixeira (USP - Brazil), João M.P. Alves, Gregory A. Buck (VCU - USA), and Maria Cristina Machado Motta (UFRJ - Brazil).

6.13. Structural and dynamical analysis of biological networks

We published a review on the structural and dynamical analysis of biological networks with as main focus explaining the cares that should be taken when this kind of analysis is performed [18]. Correctly distinguishing between potential metabolic networks and their realisations is necessary in choosing the right methods to be used and in the interpretation of their outcomes. In our review, we covered several different techniques, both static and dynamic, for the analysis of metabolic networks such as centrality techniques, flux-balance analysis and kinetic modelling of full-scale networks.

6.14. Network distance analysis

We addressed the diameter computation problem in the case of undirected unweighted graphs, where the diameter D is defined as the maximum distance among all the pairs of nodes and the distance $d(u, v)$ between two nodes u and v is defined as the number of edges contained in the shortest path from u to v . In the context of real-world networks, the textbook method based on performing a breadth-first search (in short, BFS) from every node of the graph, requires a prohibitive cost of $O(nm)$ time, where n is the number of nodes and m is the number of edges of the graph. Our main contribution consists of showing that BFS can indeed be an extremely *powerful* tool to compute the *exact* value of the diameter, whenever it is used in a more clever way. In particular, we have developed the *iterative* Fringe Upper Bound (in short, *iFUB*) algorithm to calculate the exact value of the diameter. This work has been accepted for publication in *Theoretical Computer Science* (to appear).

We then successively generalised the idea of the *iFUB* algorithm, by presenting the directed *iFUB* (in short, *DiFUB*) algorithm, in order to calculate the diameter of the strongly connected components of directed graphs [33]. As far as we know, *DiFUB* is the first algorithm which is able to compute exactly the diameter of the strongly connected components of huge real-world directed graphs. The *DiFUB* algorithm can also return a pair of nodes whose distance is exactly equal to the diameter, and a natural adaptation of it works also for weighted graphs.

6.15. Information spreading in dynamic graphs

We showed how a technique used to analyse the flooding completion time in the case of a special class of random evolving graph model, that is, the *edge-Markovian model*, can be used in order to prove that the flooding completion time of a random evolving graph $(G_t)_{t \geq 0}$ is bounded by $kD + 2C$, where intuitively (1) k is the smallest time necessary for the rising of a giant component, (2) D is the diameter of the giant component, and (3) C is the time required for the nodes outside the giant component to eventually get an edge connecting them to the giant component [30]. Then, based on this result, we developed a general methodology for analysing flooding in sequences of random graphs and we applied this general methodology to the case of power-law evolving graphs (that is, sequences of mutually independent random graphs such that the number y of nodes of degree x distributes like $1/x^\beta$ for some $\beta > 0$), and to the case of an arbitrary given degree distribution.

6.16. Metabolic network comparison

Previous works on minimal gene sets, when analysing host-dependent bacteria, found small common sets of metabolic genes. When such analyses are restricted to bacteria with similar lifestyles, larger portions of metabolism are expected to be shared and their composition is worth investigating. Comparing the small molecule metabolism of 58 bacteria carefully selected and representing a range of lifestyles, we found not a single enzymatic reaction common to all of them. While obligate intracellular symbionts have no core of reactions within their group, extracellular and cell-associated symbionts do have a small core enriched in biosynthetic processes composed of disconnected fragments. As more genomes are added, we expect, based on our simulations, that the core of cell-associated and extracellular bacteria continues to diminish, converging to approximately 60 reactions. These results were in preparation in 2011 and are now published [17]. The work was done with Ana Tereza Vasconcelos and in a collaboration with Ludovic Cottret (INSA Toulouse).

6.17. Core and periphery of metabolic networks

The core metabolism can be defined as the reactions present in every organism, however it is not robust considering that adding or removing one organism in the study will modify the resulting set. An alternative way is to include in the core the reaction that is present in a large enough proportion of species. For that, we proposed a method where the threshold to decide what is large enough is not set by the user (thus relying on a subjective choice), but rather automatically selected by the method, relying on the information contained in the data. Two approaches are being proposed, one is EM (Expectation Maximization) which relies only on the information of presence / absence of a reaction in a species while the second (NEM - Neighboring Expectation Maximization) relies on a neighbouring relation between reactions. The latter tends to classify in a same group (core or periphery) a reaction for which a majority of neighbours belong to a same group. The work is being done with Ana Tereza Vasconcelos in a collaboration with Catherine Matias, Christophe Ambroise, Yolande Diaz (Genopole, CNRS).

6.18. Metabolic stories

Enumerating stories, *i.e.*, enumerating maximal directed acyclic graphs with sets of sources and targets contained in a given subset of the nodes, is an algorithmic approach we proposed for interpreting metabolomics experiments. The modelling, algorithms and complexity results were recently accepted for publication [2]. The complexity of the enumeration problem remains unknown. There are also further modelling issues that could be dealt with in a near future. Both considerations were also detailed in a talk given in August at St. Petersburg, in the First RECOMB Satellite Conference on Open Problems in Algorithmic Biology.

We then applied our enumerating method on real data. We analysed data on the detoxification process of yeast cells exposed to cadmium. Our method allowed to recover known pathways involved in the process but also to propose alternative scenarios. The method was also investigated in order to automatically propose metabolic pathways through an experiment in which we try to recover known metabolic pathways using only minimal information (*e.g.*, their entries and endpoints). A paper is in preparation and should soon be submitted for publication. This work is being done in collaboration with Fabien Jourdan and Ludovic Cottret from the INRA at Toulouse, and with Christophe Junot from the CEA in Paris.

6.19. Minimal precursor sets

We proposed two new, more efficient algorithms for the enumeration of minimal precursor sets: PITUFINA and PAPA PITUFO [3]. The model of minimal precursor sets we had previously published was the first to formally take into account cycles, which are a common event in metabolic networks. The new methods avoid the memory issues of our previous approach by traversing directly the metabolic network structure instead of building a secondary tree representation. PAPA PITUFO additionally saves pre-computed solutions by a local modification of the network.

6.20. Minimum ratio cover of matrix columns by extreme rays of its induced cone

Given a matrix S and a subset of columns R , we studied the problem of finding a cover of R with extreme rays of the cone $\mathcal{F} = \{v \in \mathbb{R}^n \mid Sv = \mathbf{0}, v \geq \mathbf{0}\}$, where an extreme ray v covers a column k if $v_k > 0$ [34]. In order to measure how proportional a cover is, we introduced two different minimisation problems, namely the MINIMUM GLOBAL RATIO COVER (MGRC) and the MINIMUM LOCAL RATIO COVER (MLRC) problems. In both cases, we applied the notion of the *ratio* of a vector v , which is given by $\frac{\max_i v_i}{\min_{j \mid v_j > 0} v_j}$. These problems were originally motivated by a biological question on metabolic networks. This notion of ratio is also of interest in the field of *exact linear programming*, where current algorithms for scaling a matrix have a complexity that depends on the ratio of its elements. We showed that these two problems are NP-hard, even in the case in which $|R| = 1$. We introduced a mixed integer programming formulation for the MGRC problem, which is solvable in polynomial time if all columns should be covered, and introduce a branch-and-cut algorithm for the MLRC problem. Finally, we presented computational experiments on data obtained from real metabolic networks.

6.21. Optimal flux spaces of genome-scale stoichiometric models

The metabolism of organisms can be studied with comprehensive stoichiometric models of their metabolic networks. Flux balance analysis (FBA) calculates optimal metabolic performance of stoichiometric models. However, detailed biological interpretation of FBA is limited because, in general, a huge number of flux patterns give rise to the same optimal performance. The complete description of the resulting optimal solution spaces was thus far a computationally intractable problem. We introduced COPE-FBA: Comprehensive Polyhedra Enumeration Flux Balance Analysis, a computational method that solves this problem [15]. COPE-FBA indicates that the thousands to millions of optimal flux patterns result from a combinatorial explosion of flux patterns in just a few metabolic sub-networks. The entire optimal solution space can now be compactly described in terms of the topology of these sub-networks. COPE-FBA simplifies the biological interpretation of stoichiometric models of metabolism, and provides a profound understanding of metabolic flexibility in optimal states.

6.22. Lateral gene transfer as a support for the tree of life

We published with Sophie Abby the last results of her PhD work that apply an explicit phylogenetic model of horizontal gene transfer to bacterial and archaeal phyla [1]. We showed that lateral gene transfer allows to discriminate between phylogenetic hypotheses, and that in a typical bacterial gene family, 96-98% of tree branches result from vertical descent and 2-4% from lateral gene transfer.

6.23. Comparative approximability of hybridization number and directed feedback vertex set

We showed that the problem of computing the hybridization number of two rooted binary phylogenetic trees on the same set of taxa X has a constant factor polynomial-time approximation if and only if the problem of computing a minimum-size feedback vertex set in a directed graph (DFVS) has a constant factor polynomial-time approximation. The latter problem, which asks for a minimum number of vertices to be removed from a directed graph to transform it into a directed acyclic graph, is one of the problems in Karp's seminal 1972 list of 21 NP-complete problems. However, despite considerable attention from the combinatorial optimisation community, it remains to this day unknown whether a constant factor polynomial-time approximation exists for DFVS. Our result thus placed the (in)approximability of hybridization number in a much broader complexity context, and as a consequence we obtained that hybridization number inherits inapproximability results from the problem Vertex Cover [16]. On the positive side, we used results from the DFVS literature to give an $O(\log r \log \log r)$ approximation for the hybridization number, where r is the value of an optimal solution to the hybridization number problem. This work is submitted for publication.

6.24. Influence of symbionts on antagonistic interactions

Symbionts are often key players in antagonistic interactions between their hosts and other organisms. In host-parasitoid interactions, both players can be infected by different symbionts. We investigated how a virus and *Wolbachia*, respectively infecting a parasitoid and a drosophila, can shape the host-parasitoid interaction. While only a limited effect *Wolbachia* has been detected, the virus protects the parasitoid from the immune response of *Drosophila* [20]. Protection conferred by symbionts to their insect hosts is a promising avenue for antivectorial programs, but requires a thorough analysis of the evolutionary consequences of protection. We reviewed the literature on this topic [28].

6.25. Mod/Resc Parsimony Inference

We addressed a computational biology problem that aims at understanding a mechanism that could potentially be used to genetically manipulate natural insect populations infected by inherited, intra-cellular parasitic bacteria. In this problem, that we denoted by Mod/Resc Parsimony Inference, we are given a boolean matrix and the goal is to find two other boolean matrices with a minimum number of columns such that an appropriately defined operation on these matrices gives back the input. We showed that this is formally equivalent to the Biclique Edge Cover for Bipartite Graphs problem and derive some complexity results for our problem using this equivalence. We provided a new, fixed parameter tractability approach for solving both problems that slightly improves upon a previously published algorithm for the Biclique Edge Cover for Bipartite Graphs. Finally, we presented experimental results applying some of our techniques to a real-life dataset. This is the augmented journal version [23] of the conference paper that appeared in 2011.

6.26. On the genetic architecture of cytoplasmic incompatibility

Numerous insects carry intracellular bacteria manipulating their reproduction and thus facilitating their own spread. Cytoplasmic incompatibility (CI) is a common form of such manipulation, where a (currently uncharacterized) bacterial modification of male sperm induces the early death of embryos unless the fertilized eggs carry the same bacteria, inherited from the mother. The death of uninfected embryos provides an indirect selective advantage to infected ones, thus enabling the spread of the bacteria. We used and expanded recently developed algorithms (the first being the one described in the previous item) to infer the genetic architecture underlying the complex incompatibility data from the mosquito *Culex pipiens*. We showed that CI requires more genetic determinants 40 than previously believed, and that quantitative variation in gene products potentially contributes to the observed CI patterns. In line with population genetic theory of CI, our analysis suggests that toxin factors (those inducing embryo death) are present in fewer copies in the bacterial genomes than antitoxin factors (those ensuring that infected embryos survive). In combination with comparative genomics, our approach will provide helpful guidance to 45 identify the genetic basis of CI, and more generally of other toxin / anti-toxin systems that can be conceptualised under the same framework. This work is currently submitted for publication. It was done in collaboration with Sylvain Charlat from the LBBE.

6.27. Viral population structure and dynamics

The work which started a few years ago with the Pasteur Institute in Cambodia (Dr. P. Buchy) and the CIRAD at Montpellier (Dr. R. Frutos) on viral population structure and dynamics has been continued in 2012, focusing on the H5N1 and Dengue viruses. The exploratory statistical approach based on MCOA (see the Bamboo annual report for 2011) was used to identify a novel H5N1 endemic sub-clade specific to Cambodia [27] and the work performed last year on Dengue serotype 1 has been extended in 2012 to serotypes 2 and 3 [11] thus providing a more precise view of the virus population dynamics over the last 12 years and demonstrating "synchronized" replacements most probably linked to climatic disasters like flood or drought.

6.28. Charge group partitioning in biomolecular simulation

Molecular simulation techniques are increasingly being used to study biomolecular systems at an atomic level. Such simulations rely on empirical force fields to represent the intermolecular interactions. There are many different force fields available each based on a different set of assumptions and thus requiring different parametrization procedures. Recently, efforts have been made to fully automate the assignment of force-field parameters, including atomic partial charges, for novel molecules. In this work, we focused on a problem arising in the automated parameterisation of molecules for use in combination with the gromos family of force fields: namely, the assignment of atoms to charge groups such that for every charge group the sum of the partial charges is ideally equal to its formal charge. In addition, charge groups are required to have size at most k . We showed NP-hardness and gave an exact algorithm capable of solving practical problem instances to provable optimality in a fraction of a second [32].

BEAGLE Team

6. New Results

6.1. Model of genome reduction

To test whether the effect of the rearrangement rate on genome size holds independently of the artificial chemistry of the aevol (individual-based) model, we have written a simpler, mathematical model of genome size evolution including both genes and intergenic sequences, evolving through small insertions and deletions, large deletions and duplications and through selection based on gene number. The approach was presented this summer as a poster at the SMBE conference (Society for Molecular Biology and Evolution). We have shown analytically that without selective pressure, genomes spontaneously shrink and that large genomes are particularly unstable. When selection is included that favors the highest gene number, simulations show that genome sizes do not grow indefinitely as large genomes cannot be sustained. There is a trade-off between fitness and structural stability. A manuscript is being written and will be submitted in January.

6.2. The Paradoxical Effects of Allelic Recombination on Fitness

D.P. Parsons, C. Knibbe, G. Beslon. [42]

We introduced in the aevol model the possibility of DNA exchange by allelic recombination, in order to study the influence of recombination on the evolution of both fitness and genomic architecture. Surprisingly, despite the theoretical benefits it could confer, there seems to be very little (if any) differences in the fitness of the evolved organisms between the different groups of simulations.

6.3. Genome histories reconstructions

E. Tannier (Beagle), with B. Boussau, G. Szollosi, V. Daubin, L. Duret, M. Gouy, S. Abby (LBBE, Lyon), N. Lartillot (Univ Montreal), C. Chauve (SFU Vancouver)

Lateral gene transfer has been discovered in the 1940's and since has been seen by phylogeneticists as a noise one had to remove before analyses in molecular evolution. This noise was recently considered so important that it would blur the historical signal and leave no hope for reconstructing a phylogeny. In a series of papers [16], [31], [32], we model the lateral gene transfer and prove that it can be used as a signal to

- reinforce the support for the phylogeny of vertical descent [16]
- order in time some bacterial diversification events, and thus provide a unique source for dating the history of life (more than 3/4 of it is prokaryotic and the fossil record is not abundant) [31]
- have a trace of extinct species which did not leave any descendants, if they gave some genes to more successful lineages, which opens the way to include them in molecular phylogenies [32]

We devised methods to trace whole genome evolution, with multi-scale mutations: from nucleotide substitutions to large-scale rearrangements. We provided a mammalian phylogeny accounting for the evolution of several thousand genes [17], and a method to sample among evolutionary scenarios [27].

Eventually we built a model of evolution of relations between pairs of genes, enable us to reconstruct ancestral genome structure or ancestral systems of interactions [18]. In the case of genome structure we also published a method to linearize a set of ancestral relations [26].

6.4. A Theory of Rate Coding Control by Intrinsic Plasticity Effects

H. Berry (Beagle), J. Naudé and B. Delord (ISIR, CNRS UMR 7222, Univ P&M Curie, Paris) and J.T. Paz (Stanford Univ Medical Center, CA, USA).

Over the past decades, experimental and theoretical studies of the cellular basis of learning and memory have mainly focused on synaptic plasticity, the experience-dependent modification of synapses. However, behavioral learning has also been correlated with experience-dependent changes of non-synaptic voltage-dependent ion channels. This intrinsic plasticity changes the neuron's propensity to fire action potentials in response to synaptic inputs. Thus a fundamental problem is to relate changes of the neuron input-output function with voltage-gated conductance modifications. Using a sensitivity analysis in biophysically realistic models, we depicted a generic dichotomy between two classes of voltage-dependent ion channels [28]. These two classes modify the threshold and the slope of the neuron input-output relation, allowing neurons to regulate the range of inputs they respond to and the gain of that response, respectively. We further provide analytical descriptions that enlighten the dynamical mechanisms underlying these effects and propose a concise and realistic framework for assessing the computational impact of intrinsic plasticity in neuron network models. Our results account for a large repertoire of empirical observations and may enlighten functional changes that characterize development, aging and several neural diseases, which also involve changes in voltage-dependent ion channels.

6.5. The influence of topology on calcium wave propagation in 3D astrocyte networks

H. Berry, Jules Lallouette (Beagle)

Glial cells are non-neuronal cells that constitute the majority of cells in the human brain and significantly modulate information processing via permanent cross-talk with the neurons. Astrocytes are also themselves inter-connected as networks and communicate via chemical wave propagation. How astrocyte wave propagation depends on the local properties of the astrocyte networks is however unknown. We have investigated the influence of the characteristics of the network topology on wave propagation [38]. Using a model of realistic astrocyte networks (> 1000 cells embedded in a 3D space), we show that the major classes of propagations reported experimentally can be emulated by a mere variation of the topology. Our study indicates that calcium wave propagation is favored when astrocyte connections are limited by the distance between the cells, which means that propagation is better when the mean-shortest path of the network is larger. This unusual property sheds new light on consistent reports that astrocytes *in vivo* tend to restrict their connections to their nearest neighbors.

6.6. Dynamics of protein aggregation in Escherichia coli

H. Berry, Anne-Sophie Coquel (Beagle) and Ariel Lindner (INSERM U1001, Cochin Medical School, Paris).

Protein aggregation plays a key role in cell decline and leads to several human disease linked to ageing like Alzheimer or Parkinson disease and prion disease. In *Escherichia coli* bacteria, accumulation of damaged proteins and their asymmetric segregation allowed to show ageing signs. This work [14] is focused on the *in vivo* spatial dynamics of protein aggregates in *E. coli*. Protein aggregates can be classified as inclusion bodies and they are amorphous or amyloid with a high order level due to β sheets. Combining a double theoretical and experimental approach, based on modeling and time-lapse and microfluidic microscopy, we studied the mechanism governing the motion of protein aggregates and the long-term vertical transmission of prionoid aggregates for about 10 generations. Our results show clearly that Brownian diffusion governs the motion of protein aggregates and the diffusion coefficient depends on the molecule size. The amyloid proteinopathy study shows the existence of lineages propagating two kind of aggregates : globular or comet-like. Lineages maintaining globular aggregates present an increase of the aggregate size until inhibition of the growth rate while comet-like aggregates are mildly detrimental to growth. We observed also at low frequency in some lineages the presence of both aggregates and a switch between them. Globular foci give born to comet-like aggregates.

6.7. Model of membrane domains emergence

HA Soula, A Coulon, G Beslon (Beagle)

In the classical view, cell membrane proteins undergo isotropic random motion, that is a 2D Brownian diffusion that should result in an homogeneous distribution of concentration. It is, however, far from the reality: Membrane proteins can assemble into so-called microdomains (sometimes called lipid rafts) which also display a specific lipid composition. The amount of this so-called overconcentration at equilibrium is simply related to the ratio of diffusion coefficients between zones of high and low diffusion. Expanding the model to include particle interaction, we show that inhomogeneous diffusion can impact particles clusterization as well. The clusters of particles were more numerous and appear for a lower value of interaction strength in the zones of low diffusion compared to zones of high diffusion. Provided we assume stable viscosity heterogeneity in the membrane, our model proposes a simple mechanism to explain particle concentration heterogeneity and hence domains.

6.8. Deleterious effect of receptor clustering on canonical signaling pathways

BR Caré, HA Soula (Beagle)

Classical framework for analyzing system biology pathways assumed that the cells are a well mixed and stirred medium. This hypothesis can dramatically fail in the case of membrane based stage of signaling. Due to microdomains membrane receptors are colocalized. Using individual based-model we show that this clustering seriously impairs the overall ligands binding as well as several pathways downstream. We contend that this unexpected effect is a very simple tool available for a cell to adjust its response.

6.9. Novel mathematical model of Adipose tissue cells size distribution

HA Soula (Beagle) C. Soulage, A Géloën (CARMEN)

We present a novel model to explain bimodality of size distribution of adipocytes: adipose tissue cells. These cells are dedicated to storing energy excess in form of fat and therefore can experience wide variations of sizes. Ubiquitous to all the species, we tested so far the size distributions are bimodal with no characteristic size. Using data from experiments, we provide a simple surface based model of circulating fats that cells can exchange. We show that in the physiological range for the parameters of the model, we obtain bimodal distribution. We also provide prediction of the size evolution during severe caloric restriction that we were able to verify experimentally as well.

DRACULA Project-Team

6. New Results

6.1. Modelling of Erythroblastic Islands (red blood cell production)

Participants: Fabien Crauste [Contact person], Olivier Gandrillon, Vitaly Volpert.

In collaboration with N. Bessonov, S. Fischer and P. Kurbatova.

The production and regulation of red blood cells (erythropoiesis) occurs in the bone marrow where erythroid cells proliferate and differentiate within particular structures, called erythroblastic islands. A typical structure of these islands consists of a macrophage (white cell) surrounded by immature erythroid cells (progenitors), with more mature cells on the periphery of the island, ready to leave the bone marrow and enter the bloodstream.

We proposed a hybrid model [11], coupling a continuous model (ordinary differential equations) describing intracellular regulation through competition of two key proteins, to a discrete spatial model describing cell-cell interactions, with growth factor diffusion in the medium described by a continuous model (partial differential equations), to investigate the role of the central macrophage in normal erythropoiesis. Intracellular competition of the two proteins leads the erythroid cell either to proliferation, differentiation, or death by apoptosis. This approach allows considering spatial aspects of erythropoiesis, involved for instance in the occurrence of cellular interactions or the access to external factors, as well as dynamics of intracellular and extracellular scales of this complex cellular process, accounting for stochasticity in cell cycle durations and orientation of the mitotic spindle. The analysis of the model showed a strong effect of the central macrophage on the stability of an erythroblastic island, when assuming the macrophage releases prosurvival cytokines. Even though it is not clear whether or not erythroblastic island stability must be required, investigation of the model concludes that stability improves responsiveness of the model, hence stressing out the potential relevance of the central macrophage in normal erythropoiesis.

6.2. Modelling of the CD8 T cell Immune Response

Participants: Fabien Crauste [Contact person], Olivier Gandrillon, Emmanuelle Terry.

In collaboration with J. Marvel and C. Arpin.

The CD8 immune response is a specific immune response triggered by the organism when the innate response is unable to fight a pathogen. We proposed a new model of the CD8 T cell immune response based on the description of feedback controls exerted by the cytotoxic CD8 T cell population on the pathogen and the population itself [14]. This model, a system of ordinary and age-structured partial differential equations, allows describing a classical response, characterized by a cellular expansion following the pathogen-mediated activation, then a contraction phase and the generation of memory CD8 T cells. Moreover, we showed the global asymptotic stability of this system corresponding to the elimination of the virus. This situation is expected and describes for instance what is observed with the flu virus.

A simpler version of this model (based on nonlinear ordinary differential equations) has then been confronted to experimental data generated by Jacqueline Marvel's team in Lyon (immunology team), with 3 different pathogens. A parameter sweep has been performed and some parameters of the model, specific of cellular processes, have been shown to characterize CD8 immune responses against either a virus or a bacterium. This work is in progress and should be submitted soon.

6.3. Modelling of Platelet Thrombus Formation

Participants: Alen Tosenberger, Vitaly Volpert [Contact person].

In collaboration with F. Ataullakhanov, N. Bessonov, A. Butylin, G. Panasenko, M. Panteleev, E. Shnol, I. Sirakov and A. Tokarev.

An injury of a blood vessel requires quick repairing of the wound in order to prevent a loss of blood. This is done by the hemostatic system. The key point of its work is the formation of an aggregate from special blood elements, namely, platelets. The construction of a mathematical model of the formation of a thrombocyte aggregate with an adequate representation of its physical, chemical, and biological processes is an extremely complicated problem. A large size of platelets compared to that of molecules, strong inhomogeneity of their distribution across the blood flow, high shear velocities, the moving boundary of the aggregate, the interdependence of its growth and the blood flux hamper the construction of closed mathematical models convenient for biologists. We proposed a new PDE-based model of a thrombocyte aggregate formation [21], [22]. In this model, the movement of its boundary due to the adhesion and detachment of platelets is determined by the level set method. The model takes into account the distribution inhomogeneity of erythrocytes and platelets across the blood flow, the invertible adhesion of platelets, their activation, secretion, and aggregation. The calculation results are in accordance with the experimental data concerning the kinetics of the ADP-evoked growth of a thrombus in vivo for different flow velocities. The model constructed here can be easily extended to the case of other hemostatic mechanisms and can be integrated into different continuous blood flow models.

6.4. Reaction-Diffusion Model of Atherosclerosis Development

Participant: Vitaly Volpert [Contact person].

In collaboration with N. El Khatib, S. Genieys and B. Kazmierczak.

Atherosclerosis begins as an inflammation in blood vessel walls (intima). The inflammatory response of the organism leads to the recruitment of monocytes. Trapped in the intima, they differentiate into macrophages and foam cells leading to the production of inflammatory cytokines and further recruitment of white blood cells. This self-accelerating process, strongly influenced by low-density lipoproteins (cholesterol), results in a dramatic increase of the width of blood vessel walls, formation of an atherosclerotic plaque and, possibly, of its rupture. We suggested a 2D mathematical model of the initiation and development of atherosclerosis which takes into account the concentration of blood cells inside the intima and of pro- and anti-inflammatory cytokines [18]. The model represents a reaction-diffusion system in a strip with nonlinear boundary conditions which describe the recruitment of monocytes as a function of the concentration of inflammatory cytokines. We proved the existence of travelling waves described by this system and confirmed our previous results which suggest that atherosclerosis develops as a reaction-diffusion wave.

6.5. Hematopoietic model with feedback control

Participants: Mostafa Adimy [Contact person], Lila Sebaa.

In collaboration with O. Angulo and C. Marquet.

We investigate a mathematical model of blood cell production in the bone marrow (hematopoiesis). The model describes both the evolution of primitive hematopoietic stem cells and the maturation of these cells as they differentiate to form the three types of blood cells (red blood cells, white cells and platelets). The primitive hematopoietic stem cells and the first generations of each line (progenitors) are able to self-renew, and can be either in a proliferating or in a resting phase (G_0 -phase). These properties are gradually lost while cells become more and more mature. The three types of progenitors and mature cells are coupled to each other via their common origin in primitive hematopoietic stem cells compartment. Peripheral control loops of primitive hematopoietic stem cells and progenitors as well as a local autoregulatory loop are considered in the model. 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 investigate some fundamental properties of the solutions of this system, such as boundedness and positivity. 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 [7]. 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. This can be related to observations of some periodical hematological disease such as chronic myelogenous leukemia, cyclical neutropenia, cyclical thrombocytopenia, etc.

IBIS Project-Team

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 PL models, as described in Section 4.1. GNA has been integrated with the other bioinformatics tools distributed by Genostar (<http://www.genostar.com/>). Version 8.4 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. A book chapter describing the current version of GNA has been published in a volume on the modeling of bacterial molecular networks [13]. The chapter is a tutorial illustrating the practical use of recent functionalities of GNA like the network editor and the formal verification module by means of an example network in *E. coli*.

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 PL 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 PL models (Section 3.2). This has motivated extensions of the PL 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 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. A paper describing these results is being prepared for submission.

5.2. Experimental mapping of gene regulatory networks in bacteria

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 3.3). 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 of the information contained in reporter gene expression data. Several improvements of the platform for measuring gene expression are the subject of ongoing work, including a novel method for efficiently cloning reporter gene constructions on the chromosome of *E. coli*.

These tools have been used in a series of studies directed at the experimental mapping of gene regulatory networks in *E. coli*. One example, carried out in the framework of the PhD thesis of Guillaume Baptist, is 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 is currently under revision.

Other examples of on-going work are the analysis of the network involved in motility and sessility and the modulation of the RpoS regulon in *E. coli* by Stephan Lacour, the analysis of the regulation of cAMP levels in the bacterial cell by Claire Villiers, and the analysis of various aspects of the regulation of carbon metabolism by Valentin Zulkower and Stéphane Pinhal.

5.3. Analysis of metabolic coupling in gene regulatory networks

The regulation of gene expression is tightly interwoven with metabolism and signal transduction. A realistic view of genetic regulatory networks should therefore not only include direct interactions resulting from transcription regulation, but also indirect regulatory interactions mediated by metabolic effectors and signaling molecules. We coined the term metabolic coupling to denote these indirect interactions mediated by metabolism. Ignoring metabolic coupling during the analysis of the network dynamics may lead crucial feedback loops to be missed.

In previous work, published in *PLoS Computational Biology* in 2010, we showed how indirect interactions arising from metabolic coupling can be derived from a model of the underlying biochemical reaction network. We applied this approach to the carbon assimilation network in *Escherichia coli* investigating how the structural properties of the network are modified by the inclusion of metabolic interactions. Our results showed that the derived gene regulatory network is densely connected, contrary to what is usually assumed. Moreover, we found that the signs of the indirect interactions are largely fixed by the direction of metabolic fluxes, independently of specific parameter values and rate laws, and that a change in flux direction may invert the sign of indirect interactions. This leads to a feedback structure that is at the same time robust to changes in the kinetic properties of enzymes and that has the flexibility to accommodate radical changes in the environment.

It remains an open question, however, to which extent the indirect interactions induced by metabolic coupling affect the dynamics of the system. This is a key issue for understanding the relative contributions of the regulation of gene expression and metabolism during the adaptation of the cell to changes in its environment. In collaboration with Valentina Baldazzi, formerly post-doctoral fellow in IBIS and now research scientist at INRA (Avignon), we have carried out a dynamic analysis by developing a qualitative PL model of the gene regulatory network, including both the direct and indirect interactions.

In order to obtain a clearer view of the dynamic role of metabolic coupling in the adaptation of gene expression, we developed several qualitative models corresponding to a network topology including all, some, or none of the indirect interactions. The dynamical properties of the models were analyzed and compared with available experimental data using the computer tool GNA (Section 4.1). In particular, we compared the steady-state concentrations of enzymes and transcription regulators during growth on glucose and acetate, as well as the dynamic response of gene expression to the exhaustion of glucose and the subsequent assimilation of acetate. We find significant differences between the dynamics of the system in the absence and presence of metabolic coupling. This confirms that indirect interactions are essential for correctly reproducing the observed adaptation of gene expression to a change in carbon source. Our work thus underlines the importance of metabolic coupling in gene regulatory networks, and shows that such indirect interactions cannot be neglected when studying the adaptation of an organism to changes in its environment. A paper describing these results has been published in the *Journal of Theoretical Biology* [5]. Another publication, reviewing the applicability of these and other ideas for multi-scale modeling in plants, has appeared in *Trends in Plant Science* [4].

5.4. Parameter estimation for kinetic 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 comparison with the qualitative PL models described in Sections 5.1 and 5.3, these more general classes of ODE models are intended to provide a quantitative description of the network dynamics, both on the genetic and metabolic

level. New experimental techniques have led to the accumulation of large amounts of data, such as time-course measurements of metabolite, mRNA and protein concentrations and measurements of metabolic fluxes under different growth conditions. However, the estimation of parameter values in the kinetic models from these data remains particularly challenging in biology, mostly because of incomplete knowledge of the molecular mechanisms, noisy, indirect, heterogeneous, and partial observations, and the large size of the systems, with dynamics on different time-scales. We have addressed parameter estimation in the context of the analysis of the interactions between metabolism and gene expression in carbon metabolism in *E. coli*.

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*, as described in an article published in *Bioinformatics* in 2011. The model was based on the use of so-called linlog functions to approximately describe the rates of enzymatic reactions. More precisely, linlog models define reactions rates as proportional to both the enzyme concentrations and a linear combination of the logarithms of metabolite concentrations. The estimation of parameters in the linlog model from metabolomics, transcriptome, proteomics data sets required the development of a new approach, adapted to the occurrence of numerous missing values in the data sets. When applied to the above-mentioned linlog model, exploiting a high-throughput dataset published in the literature, we were able to obtain reasonable estimates of the 100 parameters.

The results of the above application also 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 among the two. In addition, we looked into the implications of this analysis for the reduction of nonidentifiable 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 has been accepted for publication in the *Journal of Mathematical Biology* [6] (see also [11] for a short version with preliminary results).

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, have been calibrated using experimental data from the experimental part of the IBIS group for the gene expression measurements and the group of Jean-Charles Portais at INSA in Toulouse for the measurements of metabolism. The model with the estimated parameter values is currently being tested and used to understand some key mechanisms in the adaptation of *E. coli* to the exhaustion of glucose. The PhD thesis of Manon Morin, which started at the end of this year in the framework of a collaboration supported by a Contrat Jeune Scientifique INRA-Inria, will further develop these research directions.

5.5. Structural identification of gene regulatory networks

In general, structural identification of genetic regulatory networks involves fitting appropriate network structures and parameters to the data. While modern measurement techniques such as reporter gene systems provide data of ever-increasing quality, the problem remains challenging because exploring all possible network structures in the search of the best fitting model is prohibitive.

In order to address the structural identification problem, Eugenio Cinquemani developed in collaboration with the Automatic Control Lab at ETH Zürich (Switzerland) and the Computer Engineering & Systems Science Department of the University of Pavia (Italy), an ODE modelling framework based on so-called unate-like functions, and a method that exploits monotonicity properties of these functions to effectively prune models that are incompatible with the data from the family of all unate-like modelling alternatives. This model invalidation step is based on simple preprocessing of time-course protein concentration and synthesis rate profiles, assumed available, and allows one to reduce the search of the best fitting model to a small subset of viable model structures.

The method, first published in *Bioinformatics* in 2010 and demonstrated on real data from the synthetic network IRMA, allows one to integrate *a-priori* knowledge on the expected network dynamics in a natural way. Leveraging on this, in the context of the same international collaboration, the method has been further developed in particular by considering relevant subclasses of the family of unate-like models that also enjoy certain quasi-convexity properties. For this restricted class, combined use of monotonicity and quasi-convexity properties allows one to ameliorate the model invalidation step, *i.e.* retain even fewer viable model structures based on affordable data preprocessing. These developments have been presented and demonstrated *in silico* in a paper published in the 2012 special issue on System Identification for Biological Systems of the *International Journal of Robust and Nonlinear Control* [9].

We are currently applying the above methods to actual, known or partially unknown, networks. In the framework of the PhD thesis of Diana Stefan, the network inference method has been applied to gene expression data from the network regulating motility of *E. coli*. First encouraging results have suggested further experimental and computational investigations that are currently in progress.

5.6. 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 is tremendous, ranging from a better comprehension of the biochemical regulatory mechanisms underlying life, 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.

In view of the potential and the relevance of the subject, one research line of IBIS is dedicated to the probabilistic modeling of the dynamics of gene regulatory networks at the level of individual cells. Our activity is centered around two main challenges. On the one hand, we address the problem of developing methods for fitting unknown network parameters of stochastic models to experimental data. As a reference case study we consider the network regulating the inset of the arabinose uptake process in *E. coli* upon depletion of glucose in the growth medium. For this system, Eugenio Cinquemani and Michel Page are developing and implementing methods for the inference of unknown parameters from fluorescence microscopy data. On the other hand, we investigate several alternative modelling approaches in an attempt to determine their relevance to different systems and application scenarios. This activity is being developed in collaboration with Gregory Batt (CONTRAINTEs, Inria Paris-Rocquencourt), Giancarlo Ferrari-Trecate (University of Pavia, Italy), and Alfonso Carta (COMORE, Inria Sophia-Antipolis - Méditerranée). First results connected to control applications on real and simulated data have been submitted for presentation at the European Control Conference to be held in 2013. Finally, further ongoing work concerns the study of noise propagation in gene regulatory networks, in collaboration with Irina Mihalcescu (Université Joseph Fourier), and the analysis of data from Fluorescence Recovery After Photobleaching (FRAP) experiments, in collaboration with Marianna Rapsomaniki and Zoi Lygerou (University of Patras, Greece) and John Lygeros (ETH Zürich, Switzerland).

5.7. Control of regulatory networks in bacteria

While systems biology is primarily concerned with natural systems shaped by evolution, synthetic biology opens up a new generation of fundamental research by trying to redesign natural systems or create novel systems from scratch. Mathematical modeling and analysis are essential components of synthetic biology, as they help understanding the consequences of (changes in) the network of interactions on the dynamical

behavior of the system. More specifically, a model can be a powerful tool for the control and regulation of the system towards a desired goal.

Within the projects ColAge and GeMCo (Section 7.2), we attempt to control one of the fundamental physiological properties of bacterial cells, their growth rate. In particular, in order to control the growth rate, we propose to focus on the gene expression machinery of *E. coli*, whose activity is controlled by a complex regulatory network with many components and intertwined feedback loops. Delphine Ropers is developing models of the gene expression machinery and Jérôme Izard, in the context of his PhD thesis, is rewiring part of the network to enable control of the network dynamics. The results on these projects are currently being prepared for publication.

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

Gene expression is controlled by the joint effect of (i) the global physiological state of the cell, in particular the activity of the gene expression machinery, and (ii) 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 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 has been accepted for *Molecular Systems Biology* [7].

MOISE Project-Team

6. New Results

6.1. Mathematical Modelling of the Ocean Dynamics

6.1.1. *Beyond the traditional approximation on the Coriolis force*

Participant: Antoine Rousseau.

Formerly, A. Rousseau has performed some theoretical and numerical studies around the derivation of quasi-hydrostatic models. With C. Lucas, he proved that it is sometimes necessary to take into account the cosine part of the Coriolis force (which is usually neglected, leading to the so-called Traditional Approximation). They have also shown that the non-traditional terms do not raise any additional mathematical difficulty in the primitive equations: well-posedness for both weak and strong solutions.

A. Rousseau and J. McWilliams (UCLA) proposed a mathematical justification of the tilt of convective plumes in the quasi-geostrophic regime, thanks to the account of the complete Coriolis force in the so-called quasi-hydrostatic quasi-geostrophic (QHQG) model. The new model has been presented in international conferences [59] and [60].

6.1.2. *Coupling Methods for Oceanic and Atmospheric Models*

Participants: Eric Blayo, David Cherel, Laurent Debreu, Antoine Rousseau, Manel Tayachi.

6.1.2.1. *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.

A first step towards the coupling between Navier-Stokes and primitive equations has been made in 2012. Starting from the optimized boundary conditions obtained for the Navier-Stokes equations, we performed an asymptotic analysis in order to obtain boundary conditions that should supplement the hydrostatic Navier-Stokes equations. These results have been presented in national and international conferences [47], [46], a paper is in preparation. David Cherel defended his PhD on Dec. 12th, 2012.

6.1.2.2. *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 is the framework of our partnership with EDF in the project MECSICO. In her PhD, M. Tayachi is aimed to build a theoretical and numerical framework to couple 1D, 2D and 3D models for river flows.

In [65], 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 and primitive equations has been started recently.

6.1.3. Numerical schemes for ocean modelling

Participants: Laurent Debreu, Jérémie Demange.

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 a recent paper [87], 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 satisfying the Peclet constraint.

In 2012, in collaboration with F. Lemarié at UCLA, this work has been extended in order to render the biharmonic diffusion operator scheme unconditionally stable [17]. This is particularly needed when the slopes between coordinate 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 begins a work on advection-diffusion schemes for ocean models (Supervisors : L. Debreu, P. Marchesiello (IRD)). His work will focus 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 [49], [48].

Salinity at 1000 m in the Southwest Pacific ocean is shown in figure 1 . 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).

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.

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 has been submitted late 2012.

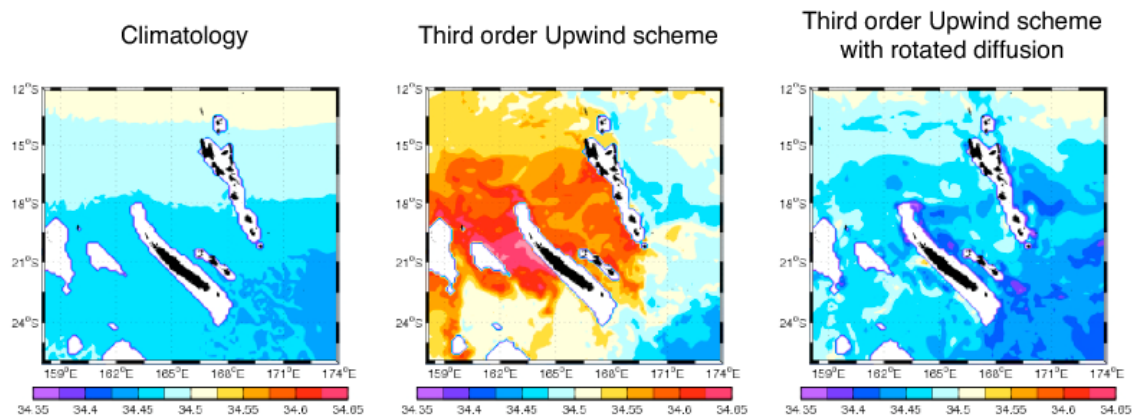


Figure 1. Salinity at 1000m in the Southwest Pacific ocean.

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. This part of the project is now well advanced and encouraging preliminary results are available on an idealized numerical configuration of an oceanic basin (see Figure 2). Several novative diagnostics have been also developed in this framework

6.2.2. Identification of pollution.

Participant: François-Xavier Le Dimet.

The problem is the next : potential sources of pollution are known but the contribution of each source to a local site is unknown. The problem is to identify the contribution of each source. This is a very common situation both at the local scale and at the synoptic scale. Thanks to second order methods we have been able to reach this goal, the theoretical part is done at FSU and application at the Institute of Mechanics of the Vietnamese Academy of Sciences. One paper has been submitted for publication [75]. At FSU M.Y. Hussaini and I. Souopgui are involved in this project

The quality of water ressources is an important problem for Vietnam. With scientists of the Institute of Mechanics (Ha Tran Thua, Hoang Van Lai, Nguyen Ba Hung) in [31] and [53] we have used the methods described in [75] for water pollution studies, in parallel Tran Thu Ha and Pham Dinh Tuan (LJK) have been working on the application of Kalman filter for this problem. Several talks have been given and papers published.

6.2.3. Variational data assimilation 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 adjoint 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.

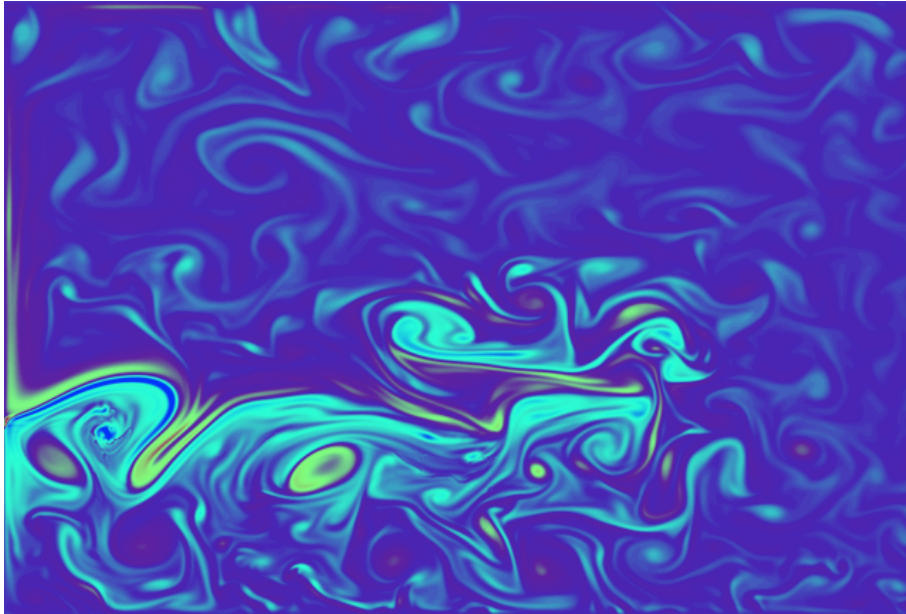


Figure 2. Surface relative vorticity of a 1/24th of a degree NEMO configuration

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. We wish to make the most of all available data and evaluate what new observations to add, where and when. Sensitivity analysis, and in particular the adjoint method, allows to identify the most influential parameters and variables and can help to design the observation network.

B. Bonan started his PhD in September 2010 on this subject. We implemented the 4D-Var algorithm for a flowline Shallow-Ice model, called Winnie, developed by C. Ritz at LGGE. In a simple configuration, we were able to generate the adjoint code by automatic differentiation. First results were encouraging and were presented at EGU [58] and Les Houches Summer School [30].

6.2.4. Ensemble Kalman filtering for large scale ice-sheet models

Participants: Bertrand Bonan, Maëlle Nodet, Catherine Ritz.

We are also interested in comparing variational methods to stochastic filtering. In the framework of B. Bonan PhD, we then implemented Ensemble Transform Kalman Filter (ETKF) on Winnie, which we would like to compare to variational assimilation methods. First results are promising and were presented at three conferences [39], [37], [38].

6.2.5. 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).

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 published in The Cryosphere Discussion [11] and presented at three conferences [52], [28], [57]. Figure 3 presents the reconstructed surface velocities compared to the observations, where we can see a good agreement of the main features, thus success of the assimilation process.

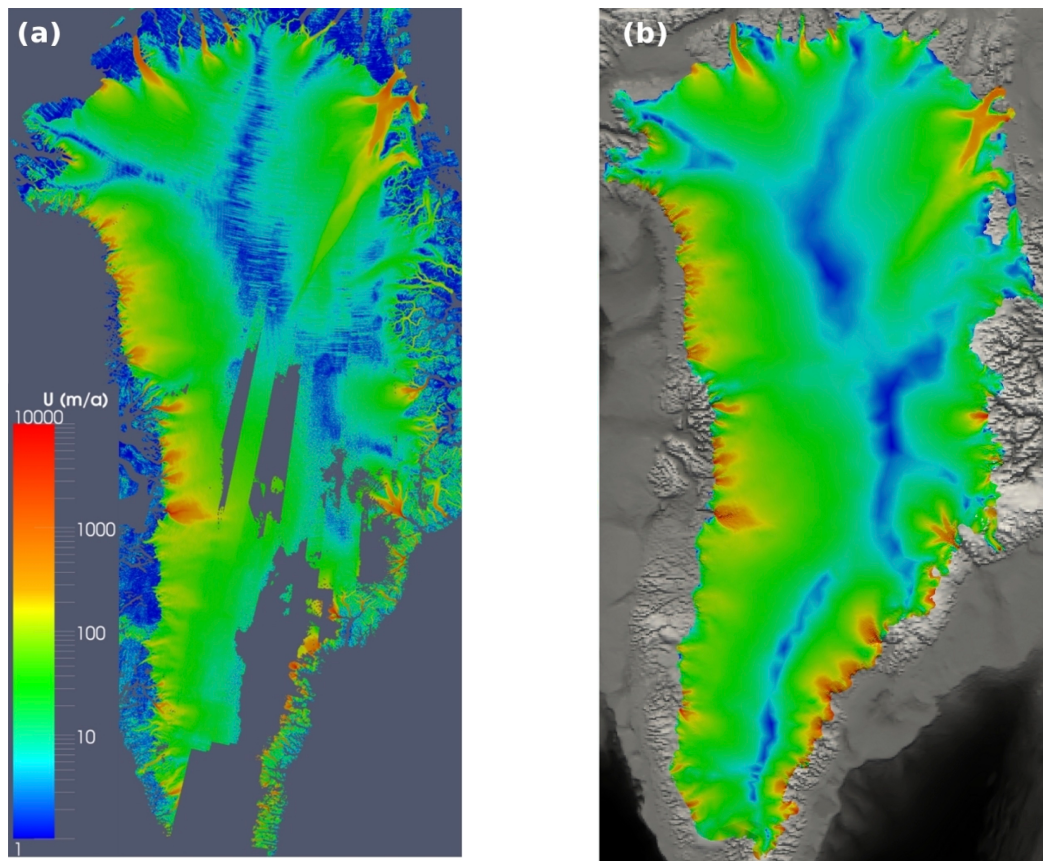


Figure 3. Surface velocities of Greenland Ice Sheet, in meters per year. On the left (a), the velocities which are observed by satellites. On the right (b), velocities obtained after assimilation.

6.2.6. Dating ice matrix and gas bubbles with *DatIce*

Participants: Eric Blayo, Bénédicte Lemieux-Dudon, Habib Toye Mahamadou Kele.

H. Toye Mahamadou Kele joined the MOISE team for 2 years as an Inria young engineer. A shared memory parallelization of the code and a more friendly user interface have been developed. Efforts have been made to calibrate the error covariance matrices by the mean of a posteriori diagnostics.

The MOISE team was involved in the Antarctic Ice Core Chronology 2012 (AICC2012) through a tight collaboration with the Laboratoire de Glaciologie et de Géophysique de l'Environnement (LGGE), the

Laboratoire des Sciences du Climat et de l'Environnement (LSCE), and other European laboratories. The AICC2012 project aimed at constructing an unified chronology for several Antarctic ice cores. A special issue is dedicated to AICC2012 in Climate of the Past http://www.clim-past-discuss.net/special_issue53.html. MOISE efforts on the DaIce code lead to two important articles currently reviewed in the Open Discussion process http://www.clim-past-discuss.net/papers_in_open_discussion.html: [2] and [22].

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 limitation 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. Variational Data Assimilation and Control of Boundary Conditions

Participant: Eugène Kazantsev.

A variational data assimilation technique applied to the identification of the optimal discretization of interpolation operators and derivatives in nodes that are adjacent to the boundary of the domain is discussed in two contexts: a simplified case of a shallow water model and the ORCA-2 configuration of the NEMO model.

Experiments with a non-linear shallow water model in [14] show that controlling the discretization of operators near a rigid boundary can bring the model solution closer to observations both within and beyond the assimilation window. This type of control allows also to improve climatic variability of the model. These properties have been studied in two different configurations: an academic case of assimilation of artificially generated observational data in a square box configuration and assimilation of real observations in a model of the Black sea.

The sensitivity of the shallow water model in the previously described configurations has been studied in detail in [15]. It is shown in both experiments that boundary conditions near a rigid boundary influence the solution higher than the initial conditions. This fact points out the necessity to identify optimal boundary approximation during a model development.

Considering a full-physics global ocean model, we apply the 4D-Var data assimilation technique to ORCA-2 configuration of the NEMO in order to identify the optimal parametrization of boundary conditions on the lateral boundaries as well as on the bottom and on the surface of the ocean [71]. The influence of boundary conditions on the solution is analyzed as in the assimilation window and beyond the window. It is shown that optimal surface and bottom boundary conditions allow us to better represent the jet streams, such as Gulf Stream and Kuroshio. Sea Surface Height in the North Atlantic before and after control is shown in fig.4). 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 [55].

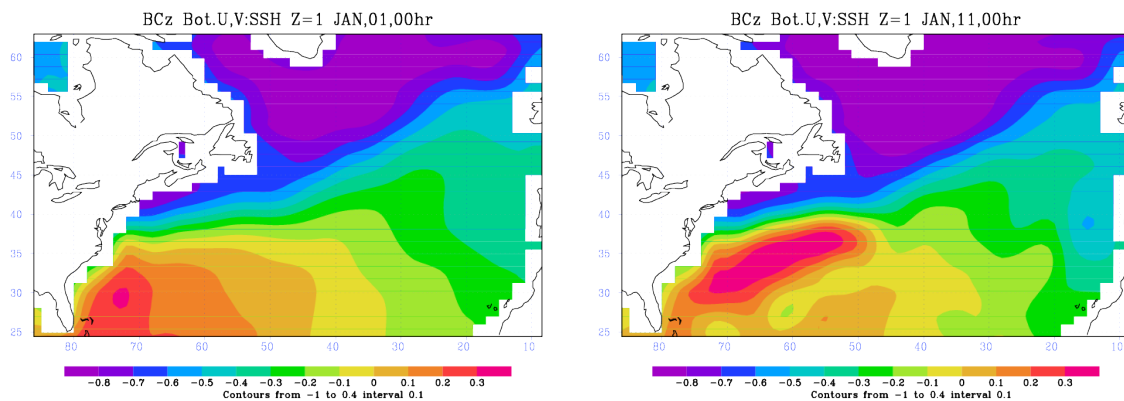


Figure 4. Sea surface elevation in the North Atlantic on the 1st (classical boundary) and on the 11th (optimal boundary) of January, 2006.

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.

6.3.3. 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 current follow-up GeoFluids (along with EPI FLUMINANCE and CLIME, and LMD, IFREMER and Météo-France)

During the ADDISA project we developed Direct Image Sequences Assimilation (DISA) and proposed a new scheme for the regularization of optical flow problems [95]. 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 [78]. 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 mains structures within each image. This can be done using, for example, a wavelet representation of images. 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.

6.3.4. Assimilation of ocean images

Participants: Vincent Chabot, Maëlle Nodet, Nicolas Papadakis, Arthur Vidard, Alexandros Makris.

In addition with the direct assimilation approach previously described, a particular attention has been given to the study of data noise for ocean image assimilation. A journal paper is about to be submitted on this subject in *Tellus A*. In the context of assimilation of structures contained in satellite images, a two step registration approach using computer vision tools has been proposed within the post-doctorate of Alexandros Makris [33], [56].

6.4. Quantifying Uncertainty

6.4.1. Sensitivity analysis for West African monsoon

Participants: Anestis Antoniadis, Céline Helbert, Clémentine Prieur, Laurence Viry.

6.4.1.1. Geophysical context

The West African monsoon is the major atmospheric phenomenon which drives the rainfall regime in Western Africa. Therefore, this is the main phenomenon in water resources over the African continent from the equatorial zone to the sub-Saharan one. Obviously, it has a major impact on agricultural activities and thus on the population itself. The causes of inter-annual spatio-temporal variability of monsoon rainfall have not yet been univocally determined. Spatio-temporal changes on the sea surface temperature (SST) within the Guinea Gulf and Saharian and Sub-Saharan Albedo are identified by a considerable body of evidences as major factors to explain it.

The aim of this study is to simulate the rainfall by a regional atmospheric model (RAM) and to analyze its sensitivity to the variability of these inputs parameters. Once precipitations from RAM are compared to several precipitation data sets we can observe that the RAM simulates the West African monsoon reasonably.

6.4.1.2. Statistical methodology

As mentioned in the previous paragraph, our main goal is to perform a sensitivity analysis for the West African monsoon. Each simulation of the regional atmospheric model (RAM) is time consuming, and we first have to think about a simplified model. We deal here with spatio-temporal dynamics, for which we have to develop functional efficient statistical tools. In our context indeed, both inputs (albedo, SST) and outputs (precipitations) are considered as time and space indexed stochastic processes. A first step consists in proposing a functional modeling for both precipitation and sea surface temperatures, based on a new filtering method. For each spatial grid point in the Gulf of Guinea and each year of observation, the sea surface temperature is measured during the active period on a temporal grid. A Karhunen-Loève decomposition is then performed at each location on the spatial grid [97]. The estimation of the time dependent eigenvalues at different spatial locations generates great amounts of high-dimensional data. Clustering algorithms become then crucial in reducing the dimensionality of such data.

Thanks to the functional clustering performed on the first principal component at each point, we have defined specific subregions in the Gulf of Guinea. On each subregion, we then choose a referent point for which we keep a prescribed number of principal components which define the basis functions. The sea surface temperature at any point in this subregion is modeled by the projection on this truncated basis. The spatial dependence is described by the coefficients of the projection. The same approach is used for precipitation. Hence for both precipitation and sea surface temperatures, we obtain a decomposition where the basis functions are functions depending on time and whose coefficients are spatially indexed and time independent. Then, the most straightforward way to model the dependence of precipitation on sea surface temperatures is through a multivariate response linear regression model with the output (precipitation) spatially indexed coefficients in the above decomposition and the input (SST) spatially indexed coefficients being predictors. A naive approach consists in regressing each response onto the predictors separately; however it is unlikely to produce satisfactory results, as such methods often lead to high variability and over-fitting. Indeed the dimensions of both predictors and responses are large (compared to the sample size).

We apply a novel method recently developed by [91] in integrated genomic studies which takes into account both aspects. The method uses an ℓ_1 -norm penalty to control the overall sparsity of the coefficient matrix of the multivariate linear regression model. In addition, it also imposes a *group* sparse penalty. This penalty puts a constraint on the ℓ_2 norm of regression coefficients for each predictor, which thus controls the total number of predictors entering the model, and consequently facilitates the detection of important predictors. The dimensions of both predictors and responses are large (compared to the sample size). Thus in addition to assuming that only a subset of predictors enter the model, it is also reasonable to assume that a predictor may affect only some but not all responses. By the way we take into account the complex and spatio-temporal dynamics. This work has been published in [1].

6.4.1.3. Distributed Interactive Engineering Toolbox

An important point in the study described above is that the numerical storage and processing of model inputs/outputs requires considerable computation resources. They were performed in a grid computing environment with a middleware (DIET) which takes into account the scheduling of a huge number of computation requests, the data-management and gives a transparent access to a distributed and heterogeneous platform on the regional Grid CIMENT (<http://ciment.ujf-grenoble.fr/>).

Thus, a different DIET module was improved through this application. An automatic support of a data grid software (<http://www.irods.org>) through DIET and a new web interface designed for MAR was provided to physicians.

These works involve also partners from the Inria project/team GRAAL for the computational approach, and from the Laboratory of Glaciology and Geophysical Environment (LGGE) for the use and interpretation of the regional atmospheric model (RAM).

6.4.2. Tracking for mesoscale convective systems

Participants: Anestis Antoniadis, Céline Helbert, Clémentine Prieur, Laurence Viry, Roukaya Keinj.

6.4.2.1. Scientific context

In this section, we are still concerned with the monsoon phenomenon in western Africa and more generally with the impact of climate change. What we propose in this study is to focus on the analysis of rainfall system monitoring provided by satellite remote sensing. The available data are micro-wave and IR satellite data. Such data allow characterizing the behavior of the mesoscale convective systems. We wish to develop stochastic tracking models, allowing for simulating rainfall scenari with uncertainties assessment.

6.4.2.2. Stochastic approach

The chosen approach for tracking these convective systems and estimating the rainfall intensities is a stochastic one. The stochastic modeling approach is promising as it allows developing models for which confidence in the estimates and predictions can be evaluated. The stochastic model will be used for hydro-climatic applications in West Africa. The first part of the work will consist in implementing a model developed in [96] on a test set to evaluate its performances, our ability to infer the parameters, and the meaning of these parameters. Once the model well fitted on toy cases, this algorithm should be run on our data set, and compared with previous results by [89] or by [88]. The model developed by [96] is a continuous time stochastic model to multiple target tracking, which allows in addition to birth and death, splitting and merging of the targets. The location of a target is assumed to behave like a Gaussian Process when it is observable. Targets are allowed to go undetected. Then, a Markov Chain State Model decides when the births, death, splitting or merging of targets arise. The tracking estimate maximizes the conditional density of the unknown variables given the data. The problem of quantifying the confidence in the estimate is also addressed. Roukaya Keinj started working on this topic with a two years postdoctoral position in November 2011. She left the team in October 2012, and is now replaced by Alexandros Makris.

6.4.3. 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, Jean-Yves Tissot, Federico Zertuche.

6.4.3.1. 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.

6.4.3.2. Estimating sensitivity indices

A first task is to develop tools for estimated sensitivity indices. Among various tools a particular attention was first paid to FAST and its derivatives. In [21], the authors present a general way to correct a positive bias which occurs in all the estimators in random balance design method (RBD) and in its hybrid version, RBD-FAST. Both these techniques derive from Fourier amplitude sensitivity test (FAST) and, as a consequence, are faced with most of its inherent issues. And up to now, one of these, the well-known problem of interferences, has always been ignored in RBD. After presenting in which way interferences lead to a positive bias in the estimator of first-order sensitivity indices in RBD, the authors explain how to overcome this issue. They then extend the bias correction method to the estimation of sensitivity indices of any order in RBD-FAST. They also give an economical strategy to estimate all the first-order and second-order sensitivity indices using RBD-FAST. A more theoretical work [77] revisit FAST and RBD in light of the discrete Fourier transform (DFT) on finite subgroups of the torus and randomized orthogonal array sampling. In [77] the authors study the estimation error of both these methods. This allows to improve FAST and to derive explicit rates of convergence of its estimators by using the framework of lattice rules. A natural generalization of the classic RBD is also provided, by using randomized orthogonal arrays having any parameters, and a bias correction method for its estimators is proposed. In variance-based sensitivity analysis, another classical tool is the method of Sobol' [94] 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 [93] and more recently by Owen [90] but the quantities they estimate still require $O(d)$ samples. In a recent work [76] the authors 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. They establish theoretical properties of such a method for the first-order Sobol' indices and discuss the generalization to higher-order indices. As an illustration, they 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.

6.4.3.3. 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 [12], 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. The present preprint is under review. 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 [13]. In a more recent work [69], the authors deal with asymptotic confidence intervals in the double limit where the sample size goes to infinity and the metamodel converges to the true model. Implementations have to be conducted on more general models such as Shallow-Water models. 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 senses has been submitted, dealing with goal oriented uncertainties assessment [70].

6.4.3.4. *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 [82] in the case where input parameters are correlated. A PhD started in October 2010 on this topic (Gaëlle Chastaing). We obtained first results [4], deriving a general functional ANOVA for dependent inputs, allowing defining new variance based sensitivity indices for correlated inputs.

6.4.3.5. *Multy-fidelity modeling for risk analysis*

Federico Zertuche 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 ([83], [92]) : 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.

6.4.4. *Multivariate risk indicators*

In collaboration with Véronique Maume-Deschamps (ISFA Lyon 1), Elena Di Bernardino (CNAM), Anne-Catherine Favre (LTHE Grenoble) and Peggy Cenac (Université de Bourgogne), we are interested in defining and estimating new multivariate risk indicators. This is a major issue with many applications (environmental, insurance, ...). Two papers were accepted for publication and two other ones are submitted. The first submitted one deals with the estimation of bivariate tails [79]. In [81] and [68] we propose estimation procedures for multivariate risk indicators. In [5] we propose to minimize multivariate risk indicators by using a Kiefer-Wolfowitz approach to the mirror stochastic algorithm.

6.4.5. *Quasi-second order analysis for the propagation and characterization of uncertainties in geophysical prediction*

We have developed a new approach for the propagation and characterization of uncertainties in geophysical prediction. Most of the method presently used are based on Monte-Carlo type (ensemble) methods, they are expensive from the computational point of view and have received a poor theoretical justification especially in the case of strongly non linear models. We have proposed a new method based on quasi-second order analysis, with a theoretical background and robust for strongly non linear models. Several papers have been published [20], [10], [19], [51] and the application to complex models are presently under development. Igor Gejadze and Victor Shutuyaev have been staying both for a total of four weeks in MOISE.

6.5. Image processing

6.5.1. *Image processing*

Participant: Nicolas Papadakis.

In collaboration with the Inria team MC2 of the Bordeaux-Sud-Ouest center, we investigate the application of image assimilation to medical issues. The objective is here to use MRI images in order to monitor EDP models dealing with tumor growth in lungs or brains. Using such images, we would like to define a patient specific process allowing to calibrate the numerical model with respect to the observed tumor. First works based on convex relaxation of the binary segmentation problem [34] have been realized in this direction by proposing a 3D segmentation method dedicated to glioblastomas from a set of MRI brain images. The obtained automatic segmentation results are very close to specialist manual segmentations (errors of 5%) and will be used as pseudo-observations for an assimilation system based on the numerical model describing the tumor growth. The final issue will be to define an observation operator linking images with the model in order to realize a direct assimilation.

Next, in collaboration with Vicent Caselles (Pompeu Fabra University, Barcelona, Spain) we tackled the problem of histogram equalization of different images. Our aim has been to include spatial information on color repartition during the histogram transfer for inpainting and shadow removal purposes [18]. We also focused with Jean-François Aujol (Institut de Mathématiques de Bordeaux), on the convexification of non linear problems such as optical flow estimation and submitted a journal paper on this subject in SIAM Journal on Imaging Sciences.

6.5.2. *Optimal transport*

Participants: Maëlle Nodet, Nicolas Papadakis, Arthur Vidard.

Within the optimal transport project TOMMI funded by the ANR white program, some new algorithms had been proposed to take into account the physics (rigidity, elasticity) of the density to transport [40]. A journal paper has been submitted on this topic in M2AN.

6.6. **Mathematical modelling for CFD-environment coupled systems**

Participant: Antoine Rousseau.

6.6.1. *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 inhomogeneity 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. During the internship of S. Barbier (co-advised by A. Rousseau and A. Rapaport (INRA-MODEMIC)) 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. This will lead to a publication. In addition, bioremediation algorithms proposed by the authors have been sent to Inria Technology Transfert Services for a patent registration.

6.6.2. *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. In [9], A. Rousseau and E. Frénod improve 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 5). This was presented in an international conference [60].

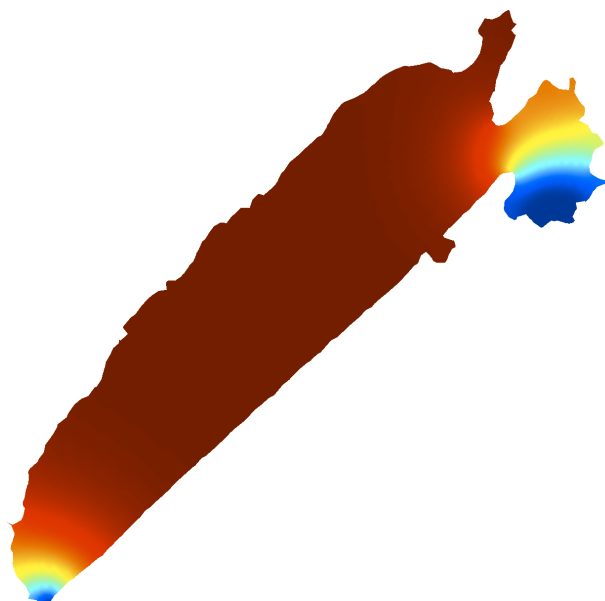


Figure 5. Confinement map in the Thau Lagoon (France). See [60].

6.7. CO₂ 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 CO₂ storage in geological environment. This work is following the study of Franck Diedro in the same subject [8]. 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.

6.8. Land Use and Transport models calibration

Participants: 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 been accepted and will start early 2013.

NUMED Project-Team

6. New Results

6.1. New result 1

Numed has developped a general strategy and generic softwares (to be released soon) to allow populational parametrization on complex models like PDEs.

STEPP Exploratory Action

6. New Results

6.1. Calibration of TRANUS Adjustment Parameters

One of the most difficult steps in calibrating the parameters of the TRANUS land use model, concerns the estimation of its adjustment parameters (so-called shadow prices), that allow to “absorb” imperfections of the model or the data. The main difficulties are the non-linearity of the underlying equations and the fact that some of these equations give rise to loops between intermediate system variables: modifications of some of these variables entail modifications of others and vice-versa. In other words, the concerned part of TRANUS is a dynamic system. Currently, users of TRANUS perform the calibration by semi-automatic (at best) trial-and-error.

We have started investigating more systematic solutions to this. A first step has been to explicitly pose the estimation problem in the form of an optimization problem, with clearly stated cost function and constraints. Next, we have found ways of splitting the problem into separable subproblems, concerning the estimation of adjustment parameters for different economic sectors. In particular, the housing/land sectors can be calibrated independently of the others. A simple gradient descent was shown to be sufficient, both theoretically and experimentally, to achieve this calibration. We are currently investigating strategies to estimate the adjustment parameters of the remaining sectors.

6.2. Calibration of TRANUS Using Maximum Likelihood Estimation

Calibration of the TRANUS land use module typically involves determination of key parameters which dictate land use assignments and prices. As mentioned earlier, It is a difficult task to calibrate a LUTI model as the number of parameters involved are large and are uncertain. Traditionally, these models are calibrated manually by experts, who try to estimate the parameters using their prior experience. However, such a method is difficult as well as time consuming, especially when the parameter space is large and uncertain. Hence, an algorithmic procedure to estimate parameters from mathematical model is desired.

We have proposed an algorithm to calibrate the land use module of TRANUS using maximum likelihood estimation (MLE). The observed outputs of the land use module is modeled to follow a Gaussian process. The covariance matrix is represented as a function of inputs of the land use module and hyperparameters. A MLE optimization problem is then formulated to estimate the parameters of the land use module and the hyperparameters of the Gaussian covariance kernel. The resulting nonlinear programming (NLP) problem is then solved using NLP solvers based on sequential quadratic programming.

The proposed calibration algorithm has been successfully applied to the model of Grenoble, France ; and the performance of the proposed calibration methodology, has been compared to traditional calibration techniques. The metric to judge performance is assumed to be the \mathcal{L}_2 norm of the difference between observed and calculated land use assignments obtained using the calibrated model.

Before this calibration task is performed, a sensitivity analysis has been carried out. Hence, sensitivity analysis of the parameters on the output is important as it helps us identify major sources of uncertainty in terms of their contribution towards output space variability. Here, the *total effect* of the land use parameters on a *quantity of interest* or *QoI* is assessed. The *QoI* is assumed to be the \mathcal{L}_2 norm of the difference between observed and calculated land use assignments. For this Grenoble model, the number of uncertain parameters involved are 100, and finally it is observed that only 3 amongst them contribute towards 99.2% of *QoI* variability. [14], [13]

6.3. Material flows, production and consumption at sub-national geographic levels

As explained earlier, estimating the actual environmental impact of an urban area on the one hand, and the efficiency of (local or national) policy options in reducing these impacts on the other, requires an understanding of the material flows and material uses generated by the considered urban area. It is important to realize that impacts (both local and distant) can vary greatly from one region or département to the next, depending on its agricultural and industrial characteristics. The whole point of this work is to evaluate as best as possible these variations, in order to best adapt public policies in terms of environmental impacts, for given socio-economic conditions and objectives.

The first step in this analysis is to establish a database of production, consumption and exchanges (import and export) at the various geographic levels of interest, and for the various types of material of interest. In practice, the finest scale of available data is a French *département*, and the publicly available data refer to the national, regional or “départemental” level. Only major primary materials are accounted for, through the content of end products and waste in these primary materials (toxic waste are accounted for separately). For example, for cereals such as wheat, production at the département level is available through the national *Agriste* database, variations of stock are small once averaged over a few years period, import and export are obtained from the *Sitram* database (a database initially elaborated by the ministry of transportation and now maintained by the Ministry of Environment), which follows all national and international transport by transportation mode and by type, through annual stratified polls of transportation companies. Productions of non-agricultural products in France is very low except for construction materials (most notably cement), for which the industry maintains its own publicly available database. Following transformations requires information from various industrial sectors, e.g., the flour trade and food industry for wheat use, taking into account animal farming which consumes a non-negligible fraction of primary agricultural products.

Once this database is constructed, one also needs to estimate production, consumption and imports and exports at finer scale than the département. In practice, this is performed by correlating the desired information at the national, regional and départemental scale with another auxiliary quantity serving as proxy, that is also known at the desired smaller scale. For example, wheat production can easily be correlated with available surfaces in wheat growing areas, that are known from the Corine Land Cover database at scales of the order of a few hundred meters. More generally, auxiliary quantities are constructed from relevant demographic and economic and geographic data, that are mostly available through the various INSEE databases. This requires some educated guess-work to find the most likely auxiliary quantities, and evaluate their correlation with the quantities of interest at scales where data on both are available. This aspect of the problem has been completed only for food staples at this stage.

An important aspect of the problem is to estimate the errors in the data. Errors can be detected when quantities of a given material are not conserved through transportation and transformation processes. It appears that the largest source of error comes from the transportation database, because the stratified polling methodology is optimized with respect to total transport from a pair of origin and destination, independently of the nature of the transported goods. It is in principle possible to compute confidence intervals per type of material and not only on total volumes of exchanges, but this requires access to some non public information. Discussions have been initiated with the Ministry to have access to this information, in order to estimate the reliability of this method of transport quantification. If the *Sitram* database turns out to be too imprecise, the method described above to estimate lacking data can be applied to transport as well with appropriate auxiliary quantities, but the results also suffer from various sources of error.

This first stage of the Material Flow analysis is nevertheless largely underway. The two next steps consist in environmental impact evaluation on the one hand, at the present date, and in developing a method of analysis of changes of such impacts under various policy scenarios and options. Both will rely on the use of Life Cycle Analysis databases, as mentioned above.

6.4. Computer vision

Three of our permanent staff have previously been active in computer vision. This activity is gradually coming to an end: the last PhD student has defended his thesis in 2012 and no new projects are started. Since this topic is not central to STEEP, results are only summarized very briefly. The main scientific result has been the development of a novel approach for 3D modeling of semi-transparent objects, which couples both, geometric and photometric information [1]. This allows 3D modeling with fewer input images than previously and potentially, with a higher accuracy. Besides this, our main activity in computer vision has been related to industrial projects, the main goal being to finalize our work of the last years with an industrial transfer.

AVALON Team

6. New Results

6.1. HPC Component Model

Participants: Zhengxiong Hou, Vincent Pichon, Christian Pérez.

6.1.1. L2C: A Low Level Component Model

We have proposed a low level component model (L²C) that supports directly native connectors for typical scenarios of high performance computing, such as MPI, shared memory and method invocation [10]. We have applied it to a typical example of stencil computation, i.e., a 2-D Jacobi application with domain decomposition. The experimental results have shown that L²C can achieve the equivalent performance as native implementations, while gaining benefits such as performance portability on the basis of the software component model.

6.1.2. Auto-tuning of Stencil Based Applications

We started modeling the performance of stencil applications on multi-core clusters. We focused in particular on a 2D Jacobi benchmark application and the NEMO application as well as memory bandwidth performance. We derived a tuning approach including 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 Curie and Juqueen supercomputers. A paper presenting these results is in preparation.

6.2. Cooperative Resource Managers

Participants: Eddy Caron, Cristian Klein, Christian Pérez, Noua Toukourou.

6.2.1. Integration of SALOME with CooRM

We have continued the validation works of the CooRM RMS architecture [52]. To this end, we focused on the SALOME numerical simulation platform developed and used jointly by EDF and CEA. In 2012, we have mostly started the integration of CooRMv1 concepts in SALOME. CooRMv1 targets moldable applications and allows them to efficiently employ their custom resource selection algorithms. We have done the necessary changes in SALOME, thus obtaining a working prototype implementation. Thanks to this, SALOME applications could be published with a custom launcher (implementing a resource selection algorithm) so as to transparently launch applications efficiently, instead of having to leave this burden to the user.

6.2.2. A Distributed Resource Management Architecture for Moldable Applications

In 2011, we have proposed CooRMv1 [52], a centralized RMS architecture to efficiently support moldable applications. Having a centralized architecture is however undesirable for geographically-distributed resources such as Grids or multiple Clouds. For example, if there is a network failure, some users will not be able to access any resources, not even those that are located on their side of the bisection.

To this end, we extended CooRMv1 and proposed a distributed version of it, distCooRM, in collaboration with the Myriads team. It allows moldable applications to efficiently co-allocated resource managed by independent agents. Simulation results show that the approach is feasible and scales well for a reasonable number of applications. In other words, it presents good strong scalability, but not weak scalability, which we intend to address in future work.

6.2.3. A Resource Management Architecture for Fair Scheduling of Optional Computations

In collaboration with two teams from IRIT, we have identified a use-case that is currently badly supported. Some applications, such as Monte-Carlo simulations, contain optional computations: These are not critical, but completing them would improve the results. When executing these application on HPC resources, most resource managers, such as batch schedulers, require the user to submit a predefined number of computing requests. If the user submits too many requests, the platform might become overloaded, whereas if the user submits too few requests, then resources might be left idle.

To solve this issue, we proposed a resource management architecture, called DIET-ethic [42], which auto-tunes the number of optional requests. It improves user happiness, fairness and the number of completed requests, when compared to a system which does not support optional computations.

6.3. Large-Scale Data Management and Processing

Participants: José Saray, Bing Tang, Gilles Fedak, Anthony Simonet.

6.3.1. Data Management on Hybrid Distributed Infrastructure

The BITDEW framework addresses the issue of how to design a programmable environment for automatic and transparent data management on Grids, Clouds and Desktop Grids. BITDEW relies on a specific set of meta-data to drive key data management operations, namely life cycle, distribution, placement, replication and fault-tolerance with a high level of abstraction.

In collaboration with Mohamed Labidi, University of Sfax (Tunisia), we have developed a data-aware and parallel version of Magik, an application for Arabic writing recognition using the BITDEW middleware. We are targeting digital libraries, which require distributed computing infrastructure to store the large number of digitalized books as raw images and at the same time to perform automatic processing of these documents such as OCR, translation, indexing, searching, etc. [20].

In 2012, we have also surveyed P2P strategies (replication, erasure code, replica repair, hybrid storage), which provide reliable and durable storage on top of hybrid distributed infrastructures composed of volatile and stable storage. Following these simulation studies, we are implementing a prototype of the Amazon S3 storage on top of BitDew, which will provide reliable storage by using both Desktop free disk space and volunteered remote Cloud storage [25].

6.3.2. MapReduce Programming Model for Desktop Grid

MapReduce is an emerging programming model for data-intense applications proposed by Google, which has recently attracted a lot of attention. MapReduce borrows from functional programming, where programmer defines Map and Reduce tasks executed on large sets of distributed data. In 2010, we developed an implementation of the MapReduce programming model based on the BitDew middleware. Our prototype features several optimizations which make our approach suitable for large scale and loosely connected Internet Desktop Grid: massive fault tolerance, replica management, barriers-free execution, latency-hiding optimization as well as distributed result checking. We have presented performance evaluations of the prototype both against micro-benchmarks and real MapReduce applications. The scalability test achieved linear speedup on the classical WordCount benchmark. Several scenarios involving larger hosts and host crashes demonstrated that the prototype is able to cope with an experimental context similar to real-world Internet [9].

In collaboration with the Huazhong University of Science & Technology (China), we have developed an emulation framework to assess MapReduce on Internet Desktop Grid. We have made extensive comparison on BitDew-MapReduce and Hadoop using GRID'5000 which show that our approach has all the properties desirable to cope with an Internet deployment, whereas Hadoop fails on several tests [22].

We have published a joint work in collaboration with Virginia Tech (USA), which is a presentation of two alternative implementations of MapReduce for Desktop Grids : Moon and Bitdew [37].

6.4. Computing on Hybrid Distributed Infrastructure

Participants: Simon Delamare, Gilles Fedak, José Saray, Anthony Simonet.

6.4.1. *SpeQuloS: Providing Quality-of-Service to Desktop Grids using Cloud resources*

EDGI is an FP7 European project, following the successful FP7 EDGeS project, whose goal is to build a Grid infrastructure composed of "Desktop Grids", such as BOINC or XtremWeb, where computing resources are provided by Internet volunteers, and "Service Grids", where computing resources are provided by institutional Grid such as EGI, gLite, Unicore and "Clouds systems" such as OpenNebula and Eucalyptus, where resources are provided on-demand. The goal of the EDGI project is to provide an infrastructure where Service Grids are extended with public and institutional Desktop Grids and Clouds.

The main limitation with the current infrastructure is that it cannot give any QoS support for applications running in the Desktop Grid (DG) part of the infrastructure. For example, a public DG system enables clients to return work-unit results in the range of weeks. Although there are EGI applications (e.g., the fusion community's applications) that can tolerate such a long latency most of the user communities want much shorter deadlines.

In 2011, we have developed the SpeQuloS middleware to solve this critical problem. Providing QoS features even in Service Grids is hard and not solved yet satisfactorily. It is even more difficult in an environment where there are no guaranteed resources. In DG systems, resources can leave the system at any time for a long time or forever even after taking several work-units with the promise of computing them. Our approach is based on the extension of DG systems with Cloud resources. For such critical work-units the SpeQuloS system is able to dynamically deploy fast and trustable clients from some Clouds that are available to support the EDGI DG systems. It takes the right decision about assigning the necessary number of trusted clients and Cloud clients for the QoS applications. In 2012, we have conducted extensive simulations to evaluate various strategies of Cloud resources provisioning. Results show that SpeQuloS improve the QoS of BoTs on three aspects: it reduces the makespan by removing the tail effect, it improves the execution stability and it allows to accurately predicts the BoT completion time [14], [21], [35]. The software have now been delivered to the partners and run in production in the European Desktop Grid Infrastructure.

6.4.2. *Scheduling on Hybrid Distributed Computing Infrastructures*

In collaboration with the Mircea Moca, from the Babes-Bolyai University of Cluj-Napoca (Romania), we have investigated new scheduling algorithms for pull-based scheduler, which relies on Promethee method. We have shown that these heuristics perform efficiently on three different kinds of infrastructures, namely Grids, Clouds and Desktop Grids [23].

6.5. Energy Efficiency in Large Scale Systems

Participants: Ghislain Landry Tsafack, Mohammed El Mehdi Diouri, Olivier Glück, Laurent Lefevre.

6.5.1. *Energy Efficiency in HPC Systems*

Modern high performance computing subsystems (HPC) – including processor, network, memory, and I/O — are provided with power management mechanisms. These include dynamic speed scaling and dynamic resource sleeping. Understanding the behavioral patterns of high performance computing systems at runtime can lead to a multitude of optimization opportunities including controlling and limiting their energy usage. We have proposed a general purpose methodology for optimizing energy performance of HPC systems considering processor, disk and network. We have relied on the concept of execution vector along with a partial phase recognition technique for on-the-fly dynamic management without any a priori knowledge of the workload. We have demonstrated the effectiveness of our management policy under two real-life workloads. Experimental results have shown that our management policy in comparison with baseline unmanaged execution saves up to 24% of energy with less than 4% performance overhead for our real-life workloads [28], [27], [26]. This work is done under the Large Scale Initiative Hemera project (Joint PhD between Avalon and IRIT (Toulouse) with J.-M. Pierson, P. Stolf and G. Da Costa).

6.5.2. Energy Considerations in Checkpointing and Fault Tolerance Protocols

Two key points should be taken into account in future exascale systems: fault tolerance and energy consumption. To address these challenges, we evaluated checkpointing and existing fault tolerance protocols from an energy point of view. We measured on a real testbed the power consumption of the main atomic operations found in these protocols: checkpointing, message logging and coordination. The results [16], [51] show that process coordination and RAM logging consume more power than checkpointing and HDD logging. However, the results we presented in Joules per Bytes for I/O operations, emphasize that checkpointing and HDD logging consume more energy than RAM logging because of the logging duration which is much more higher on HDD than on RAM. We have also shown that for identical nodes performing the same operation, the extra power cost due to this operation is the same. In general, we have learned that the power consumption of a node during a given operation remains constant during this operation. The power consumption of such a node is equal to its idle power consumption to which we add the extra power consumption due to the operation it is performing. Finally, we proposed to consider energy consumption as a criterion for the choice of fault tolerance protocols. In terms of energy consumption, we should promote message logging for applications exchanging small volumes of data and coordination for applications involving few processes. This work is a joint work with F. Cappello (Inria-UIUC-NCSA Joint Laboratory for Petascale Computing).

6.5.3. Towards a Smart and Energy-Aware Service-Oriented Manager for Extreme-Scale Applications

To address the issue of energy efficiency for exascale supercomputers, we proposed a smart and energy-aware service-oriented manager for exascale applications: SEASOMES [17]. This framework aggregates the various energy-efficient solutions to "consume less" energy and to "consume better". It involves both internal and external interactions with the various actors interfering directly or indirectly with the supercomputer. On the one hand, we recommended a more fine-grained collaboration between application and hardware resources in order to reduce energy consumption and provide sustainable exascale services. On the other hand, we suggested a cooperation between the user, the administrator, the resource manager and the energy supplier for the purpose of "consuming better".

6.6. Green-IT Innovation Analysis

Participant: Laurent Lefevre.

Green IT has recently appeared as a mandatory approach to take into account of energy efficiency in Information Technology. This research investigates the Green IT area and its opportunities for innovation. Main motivations for Green IT have been analyzed and we have proposed new definition of Green IT including social, environmental and economic concerns. We have proposed a new model of a virtuous circle that appears in Green IT: while Green IT has its own motivations, resulting research feeds other research field in a virtuous circle. Innovation in this particular sector paves the way for further innovation by means of original research not foreseen at first thoughts.

This analysis is joint work with IRIT (Toulouse - C. Herozog, J.-M. Pierson) [19].

6.7. Workflow Scheduling

Participants: Eddy Caron, Frédéric Desprez, Cristian Klein, Vincent Lanore, Sylvain Gault, Christian Pérez, Adrian Muresan, Frédéric Suter.

6.7.1. High-Level Waste Application Scheduling

Brought forward by EDF, a partner in the ANR COOP project, High-Level Waste is a multi-level application: It is composed of many moldable tasks, part of which are initially known. Some of these tasks may, with a certain probability, launch other tasks, which usually take longer. We have proposed several scheduling algorithms to optimize the performance of such applications, which are little studied in current literature. Experiments with simulations showed that considerable gains can be made, not only in terms of performance, but also performance portability. This work will be published in 2013 [31].

6.7.2. Elastic Scheduling for Functional Workflows

As a recent research direction we have focused on the development of an allocation strategy for budget-constrained workflow applications that target IaaS Cloud platforms. The workflow abstraction is very common amongst scientific applications. It is easy to find examples in any field from bioinformatics to geography. The reasons for the proliferation of workflow applications in science are various, from the building of applications on top of legacy code to modeling of applications that have an inherent workflow structure. The first workflow applications were composed of sequential tasks, but as computational units became more and more parallel, workflow applications have also evolved and are now formed of parallel tasks and, occasionally, parallel moldable tasks. The classic DAG structure of workflow applications has also changed as some applications need to perform refinement iteration, creating loop-like constructs.

We have considered a general model of workflow applications that permit non-deterministic transitions. We have elaborated two budget-constrained allocation strategies for this type of workflow. The problem is a bi-criteria optimization problem as we are optimizing both budget and workflow makespan [12].

For a practical validation of the work, we are currently working on the implementation of the budget-constrained scheduler as part of the Nimbus open source cloud platform. This is being tested with a cosmological simulation workflow application called *Ramses* (see Section 4.4). This is a parallel MPI application that, as part of this work, has been ported for execution on dynamic virtual platforms. This work has been done in the form of a two month internship at the Argonne National Laboratory, USA, under the guidance of Kate Keahey and has been accepted for poster presentation in the XSEDE 2012 conference.

6.7.3. Self-Healing of Operational Workflow Incidents on Distributed Computing Infrastructures

Distributed computing infrastructures are commonly used through scientific gateways, but operating these gateways requires important human intervention to handle operational incidents. We have designed a self-healing process that quantifies incident degrees of workflow activities from metrics measuring long-tail effect, application efficiency, data transfer issues, and site-specific problems. These metrics are simple enough to be computed online and they make little assumptions on the application or resource characteristics. From their degree, incidents are classified in levels and associated to sets of healing actions that are selected based on association rules modeling correlations between incident levels. We specifically study the long-tail effect issue, and propose a new algorithm to control task replication. The healing process is parametrized on real application traces acquired in production on the European Grid Infrastructure. Experimental results obtained in the Virtual Imaging Platform show that the proposed method speeds up execution up to a factor of 4, consumes up to 26% less resource time than a control execution and properly detects unrecoverable errors.

This work is done in collaboration with Tristan Glatard and Rafael Ferreira Da Silva from CREATIS (UMR5220).

6.7.4. Scheduling for MapReduce Based Applications

We have worked on scheduling algorithms for MapReduce applications in Grids and Clouds as we aim at providing resource-efficient and time-efficient scheduling algorithms. This work is mainly done within the scope of the Map-Reduce ANR project.

A deliverable presenting the heuristics for scheduling data transfers derived from a previous work by Berlinska and Drozdowsky has been written [50]. A section of a collaborative paper has been written and the paper has been presented at the ICA CON conference [9], [4]. The results of the aforementioned heuristics that has been previously implemented in a visualization / simulation tool, has been summarized in a paper accepted for RenPar. Moreover, these algorithms and heuristics have been implemented in the MapReduce framework HoMR.

6.8. Performance Evaluation and Modeling

Participants: Eddy Caron, Frédéric Desprez, Matthieu Imbert, Georges Markomanolis, Jonathan Rouzaud-Cornabas, Frédéric Suter.

6.8.1. Time-Independent Log Format

Simulation is a popular approach to obtain objective performance indicators of platforms that are not at one's disposal. It may for example help the dimensioning of compute clusters in large computing centers. In many cases, the execution of a distributed application does not behave as expected, it is thus necessary to understand what causes this strange behavior. Simulation provides the possibility to reproduce experiments under similar conditions. This is a suitable method for experimental validation of a parallel or distributed application.

The tracing instrumentation of a profiling tool is the ability to save all the information about the execution of an application at run-time. Every scientific application executed computed instructions. The originality of our approach is that we measure the completed instructions of the application and not its execution time. This means that if a distributed application is executed on N cores and we execute it again by mapping two processes per core then we need $N/2$ cores and more time for the execution time of the application. An execution trace of an instrumented application can be transformed into a corresponding list of actions. These actions can then be simulated by SimGrid. Moreover the SimGrid execution traces will contain almost the same data because the only change is the use of half cores but the same number of processes. This does not affect the number of the completed instructions so the simulation time does not get increased because of the overhead. The GRID'5000 platform is used for this work and the NAS Parallel Benchmarks are used to measure the performance of the clusters.

Our main contribution is to propose of a new execution log format that is time-independent. This means that we decouple the acquisition of the traces from the replay. Furthermore we implemented a trace replay tool which relies on top of fast, scalable and validated simulation kernel of SimGrid. We proved that this framework applies for some of the NAS Parallel Benchmarks and we can predict their performance with a good accuracy. Moreover we improved the accuracy of the performance's prediction by applying different instrumentation configurations according to the requirements of our framework. Some performance issues of the executed benchmarks were taken under consideration for more accurate predictions. Also the simulator was reimplemented in order to have more accurate results and take advantage of the last SimGrid's simulation techniques. Finally we did a survey on many different tracing tools with regards to the requirements of our methodology which includes all the latest provided tools from the community. For the extreme cases where we used many nodes by mapping a lot of processes per core, some issues were indicated that we are trying to solve in order to be able to apply our methodology with less overhead. Also we plan to predict the performance of more benchmarks.

6.8.2. Dynamic Network Forecasting

In distributed systems the knowledge of the network is mandatory to know the available connections and their performance. Indeed, to be able to efficiently schedule network transfers on computing platforms such as clusters, grids or clouds, accurate and timely predictions of network transfers completion times are needed. We designed a new metrology and performance prediction framework called Pilgrim which offers a service predicting the completion times of current and concurrent TCP transfers. This service uses SimGrid to simulate the network transfers. Ongoing work is to obtain experimental results comparing the predictions obtained from Pilgrim to the real transfer completion times.

6.8.3. Amazon EC2 simulation

During this year, we have developed an extension of SimGrid to simulate multi-platforms Clouds: SimGrid Cloud Broker (SGCB). It simulates the suite of services provided by Amazon AWS: EC2 for virtual machines, S3 for key-value storage and EBS for block storage. SGCB allows to easily evaluate different resource selection policy but also to simulate an entire application running on a set of resources that come from multiple Clouds. As the billing mechanism is a crucial feature of the Clouds, SGCB is able to simulate it. For this, we extended SimGrid in order to do the accounting of all virtual resources used. With this accounting, we are able to simulate the process of billing as Amazon does it. We are working to increase the accuracy of our performance models, and therefore the validity of the results for different use cases.

6.9. Cloud Resource Management

Participants: Eddy Caron, Frédéric Desprez, Arnaud Lefray, Jonathan Rouzaud-Cornabas, Julien Carpentier, Jean-Patrick Gelas, Laurent Lefevre, Maxime Morel, Olivier Mornard, Francois Rossigneux.

6.9.1. Resource Provisioning for Federations of Clouds

Since the visit of Jose Luis Lucas Simarro, we have established a collaboration with the Distributed Systems Architecture Research Group at Complutense University of Madrid (Spain) on resource brokering strategies for multiples Clouds. The purpose is to design new strategies that are able to migrate services from a Cloud to another one. VM migration is done to save money when the price of running a given VM change. Indeed, in modern Clouds such as Amazon EC2, Spot Instances have dynamic prices that change based on the law of supply and demand. Most of the current solutions only take into account the cost of computation when migrating services between Clouds. However, when a service is migrated, we need to pay network traffic between the two Clouds and the storage of the Virtual Machine image in both Clouds during the migration. We are studying through simulations different resource selection algorithms that take into account the cost of all resources: compute, storage, and network.

6.9.2. Energy Efficient Clouds

Within the projects CompatibleOne (Open Source Cloud Broker) and XLcloud (Energy Efficiency in Open-Stack based clouds), we explore the design of energy aware and energy efficient cloud infrastructures. Monitoring of physical and virtual resources is injected into cloud frameworks. Systems based on such metrics are designed in order to benefit from energy usage knowledge in virtual machines mapping and precise accounting [13].

6.9.3. User Isolation

Inter-VM and virtual network isolation is weak in terms of both security and performance. Accordingly, it can not guarantee performance, security and privacy requirements. This is a serious issue as most of clouds are multi-tenant and users do not trust each other. By improving the resource allocation process, we show how these issue can be solved and thus the overall security of the clouds improved. Moreover, we show how a Cloud Service Provider (CSP) can let the users express their security requirements. We show that isolation requirements have a cost for the Cloud Service Providers but they can bill requirements as an additional service. By doing so, they will have a new resource of income and the users trust in their platforms will increase as they can express security requirements.

6.9.4. Cloud Security

Mandatory Access Control is really poorly supported by Cloud environments. Our work proposes extensions of the OpenNebula Cloud in order to provide an advanced MAC protection of the virtual machines hosted by the different nodes of the Cloud. Thus, unique SELinux security labels are associated with the virtual machines and their resources. The instantiations and migrations of the virtual machines maintain those unique security labels. Moreover, PIGA-Virt provides a unified way to control the information flows within a virtual machine but also between multiple virtual machines. SELinux controls the direct flows. PIGA-Virt adds advanced controls. Thus, a PIGA protection rule can control several direct and indirect flows. The benchmarks of PIGA-Virt show that our Trusted OpenNebula Cloud is efficient regarding the quality of the protection.

This work is done in collaboration with Christian Toinard from LIFO/ENSI de Bourges.

6.10. Virtualizing Home Gateways at Large Scale

Participants: Jean-Patrick Gelas, Laurent Lefevre.

About 80-90% of the energy in today's wireline networks is consumed in the access network, with about 10 W per user being dissipated mostly by the customer premises equipment (CPE). Home gateway is a popular equipment deployed at the end of networks and supporting a set of heterogeneous services (from network to multimedia services). These gateways are difficult to manage for network operators and consume a lot of energy. This research explores the possibility to reduce the complexity of such equipment by moving services to some external dedicated and shared equipments. When combined to quasi passive CPE, this approach can reduce the energy consumption of wired networks infrastructures. This research is done within the GreenTouch initiative which aims to increase network energy efficiency by a factor of 1000 from current levels by 2015.

This work is done with collaboration with Addis Abeba University (Ethiopia) (M. Mulugeta and T. Assefa) [18].

6.11. Self-Adaptive Deployment

Participants: Eddy Caron, Maurice-Djibril Faye, Jonathan Rouzaud-Cornabas.

Software systems are increasingly expected to be self-adaptive. Such software systems have the capability to autonomously modify their behavior at run-time in response to changes in their environment. This capability may be included in the software systems at design time or later by external mechanisms. Therefore, along their development process multiple adaptation concerns must be considered, such as the response to changes in the utilization patterns, the need for alternative algorithms for implementing a function, or the diversity of the infrastructure. We have designed an architecture which aims to add self-adaptive capabilities to an existing middleware so that its deployment becomes self-adaptive. The framework uses external mechanisms for that purpose since this capability was not a native feature.

DANTE Team

6. New Results

6.1. Use of wireless sensor network for Assessing Interactions between Healthcare Workers and Patients under Airborne Precautions

Direct observation has been widely used to assess interactions between healthcare workers (HCWs) and patients but is time-consuming and feasible only over short periods. We used a Radio Frequency Identification Device (RFID) system to automatically measure HCW-patient interactions [14]. The RFID was well accepted by HCWs. This original technique holds promise for accurately and continuously measuring interactions between HCWs and patients, as a less resource-consuming substitute for direct observation. The results could be used to model the transmission of significant pathogens. HCW perceptions of interactions with patients accurately reflected reality.

6.2. Psychological Aspects of Social Communities

Social Network Analysis has often focused on the structure of the network without taking into account the characteristics of the individual involved. In this work [28], [8], we aim at identifying how individual differences in psychological traits affect the community structure of social networks. Instead of choosing to study only either structural or psychological properties of an individual, our aim is to exhibit in which way the psychological attributes of interacting individuals impacts the social network topology. Using psychological data from the myPersonality application and social data from Facebook, we confront the personality traits of the subjects to metrics obtained after applying the C3 community detection algorithm [41] to the social neighborhood of the subjects. We observe that introverts tend to have less communities and hide into large communities, whereas extroverts tend to act as bridges between more communities, which are on average smaller and of varying cohesion.

6.3. Community detection: dynamic, overlapping, fuzzy

Community, a notion transversal to all areas of Social Network Analysis, has drawn tremendous amount of attention across the sciences in the past decades. Numerous attempts to characterize both the sociological embodiment of the concept as well as its observable structural manifestation in the social network have to this date only converged in spirit. No formal consensus has been reached on the quantifiable aspects of community, despite it being deeply linked to topological and dynamic aspects of the underlying social network.

The DANTE team proceeded results on several aspects of community detection in large scale networks.

- Presenting a fresh approach to the evaluation of communities, we introduce and build upon the cohesion [8], a novel metric which captures the intrinsic quality, as a community, of a set of nodes in a network. The cohesion, defined in terms of social triads, was found to be highly correlated to the subjective perception of communitness through the use of a large-scale online experiment in which users were able to compute and rate the quality of their social groups on Facebook. The use of the cohesion proves invaluable in that it offers non-trivial insights on the network structure and its relation to the associated semantic. The use of the cohesion was used for example in order to study Agreement Groups in the United States Senate [35].
- Overlapping community detection is a popular topic in complex networks. As compared to disjoint community structure, overlapping community structure is more suitable to describe networks at a macroscopic level. Overlaps shared by communities play an important role in combining different communities. In this paper, two methods are proposed to detect overlapping community structure. One is called clique optimization, and the other is named fuzzy detection. Clique optimization aims at detecting granular overlaps. The clique optimization method is a fine grain scale approach. Each

granular overlap is a node connected to distinct communities and it is highly connected to each community. Fuzzy detection is at a coarser grain scale and aims at identifying modular overlaps. Modular overlaps represent groups of nodes that have high community membership degrees with several communities. A modular overlap is itself a possible cluster/sub-community [7], [38].

6.4. Structure of Changes in Dynamic Contact Networks

We present a methodology to investigate the structure of dynamic networks in terms of concentration of changes in the network. We handle dynamic networks as series of graphs on a set of nodes and consider the changes occurring between two consecutive graphs in the series. We apply our methodology to various dynamic contact networks coming from different contexts and we show that changes in these networks exhibit a non-trivial structure: they are not spread all over the network but are instead concentrated around a small fraction of nodes. We compare our observations on real-world networks to three classical dynamic network models and show that they do not capture this key property [31].

6.5. Dynamic Resource Management in Clouds: A Probabilistic Approach

Dynamic resource management has become an active area of research in the Cloud Computing paradigm. Cost of resources varies significantly depending on configuration for using them. Hence efficient management of resources is of prime interest to both Cloud Providers and Cloud Users. In this work we suggest a probabilistic resource provisioning approach that can be exploited as the input of a dynamic resource management scheme. Using a Video on Demand use case to justify our claims, we propose an analytical model inspired from standard models developed for epidemiology spreading, to represent sudden and intense workload variations. We show that the resulting model verifies a Large Deviation Principle that statistically characterises extreme rare events, such as the ones produced by 'buzz/flash crowd effects' that may cause workload overflow in the VoD context. This analysis provides valuable insight on expectable abnormal behaviours of systems. We exploit the information obtained using the Large Deviation Principle for the proposed Video on Demand use-case for defining policies (Service Level Agreements). We believe these policies for elastic resource provisioning and usage may be of some interest to all stakeholders in the emerging context of cloud networking [4], [24].

6.6. Classification of Content and Users in BitTorrent by Semi-supervised Learning Methods

P2P downloads still represent a large portion of today's Internet traffic. More than 100 million users operate BitTorrent and generate more than 30% of the total Internet traffic. Recently, a significant research effort has been done to develop tools for automatic classification of Internet traffic by application. The purpose of the present work is to provide a framework for sub-classification of P2P traffic generated by the BitTorrent protocol. The general intuition is that the users with similar interests download similar contents. This intuition can be rigorously formalised with the help of graph based semi-supervised learning approach. We have chosen to work with PageRank based semi-supervised learning method, which scales well with very large volumes of data. We provide recommendations for the choice of parameters in the PageRank based semi-supervised learning method. In particular, we show that it is advantageous to choose labelled points with large PageRank score.

This work was awarded best paper at the 3rd International Workshop on Traffic Analysis and Classification (in conjunction with the 8th International Wireless Communications and Mobile Computing Conference, 2012) [21] and led to a companion paper [22].

6.7. Large deviations estimates for the multiscale analysis of heart rate variability

In the realm of multiscale signal analysis, multifractal analysis provides with a natural and rich framework to measure the roughness of a time series. As such, it has drawn special attention of both mathematicians and practitioners, and led them to characterize relevant physiological factors impacting the heart rate variability. Notwithstanding these considerable progresses, multi-fractal analysis almost exclusively developed around the concept of Legendre singularity spectrum, for which efficient and elaborate estimators exist, but which are structurally blind to subtle features like non-concavity or, to a certain extent, non scaling of the distributions. Large deviations theory allows bypassing these limitations but it is only very recently that performing estimators were proposed to reliably compute the corresponding large deviations singularity spectrum. In this article, we illustrate the relevance of this approach, on both theoretical objects and on human heart rate signals from the Physionet public database. As conjectured, we verify that large deviations principles reveal significant information that otherwise remains hidden with classical approaches, and which can be reminiscent of some physiological characteristics. In particular we quantify the presence/absence of scale invariance of RR signals.

These results gather most achievements we carried out within the ANR project DMASC.

6.8. An Inexpensive Packet Capture Solution with Robust and Accurate Timestamping

The availability of inexpensive and reliable packet capture solutions is highly desirable for the management of future Internet infrastructures and practices. Currently, available solutions are either 1) based on GPS antennas and dedicated hardware and thus are expensive and difficult to deploy, or 2) based on commodity hardware and standard synchronization protocols and thus have inaccurate timestamps and cannot handle monitoring at high rate. In a series of ongoing works in collaboration with the Melbourne University (Australia), we proposed an architecture for a packet monitoring solution which combines inexpensive network cards capable of hardware timestamping, with RAD-clock, an open source software clock. In different papers, we presented the first implementation and evaluation of our approach, demonstrating a good compromise between affordability and accuracy [33], [36].

6.9. KBAC: Knowledge-Based Admission Control

Many methods have been proposed in the literature to perform admission control in order to provide a sufficient level of Quality of Service (QoS) to accepted flows. In this work, we introduce a novel data-driven method based on a timevarying model that we refer to as Knowledge-Based Admission Control solution (KBAC). Our KBAC solution consists of three main stages: (i) collect measurements on the on-going traffic over the communication link; (ii) maintain an up-to-date broad view of the link behavior, and feed it to a Knowledge Plane; (iii) model the observed link behavior by a mono-server queue whose parameters are set automatically and which predicts the expected QoS if a flow requesting admission were to be accepted. Our KBAC solution provides a probabilistic guarantee whose admission threshold is either expressed, as a bounded delay or as a bounded loss rate. We run extensive simulations to assess the behavior of our KBAC solution in the case of a delay threshold. The results show that our KBAC solution leads to a good trade-off between flow performance and resource utilization. This ability stems from the quick and automatic adjustment of its admission policy according to the actual variations on the traffic conditions [19].

6.10. Substitution Networks: Performance Collapse due to Overhead in Communication Times

A substitution network is a wireless solution whose purpose is to bring back connectivity or to provide additional bandwidth capacity to a network that just suffered a failure or a dramatic surge in its workload. We analyze the performance of the simplest possible multihop topology for a substitution network, i.e., the

multihop chain subject to traffic transmitted in both directions. Clearly, the potential capacity of a substitution network, whose technology should be embedded in mobile routers, is very likely to be far much smaller than the prior base network. We investigate the actual performance attained by such a substitution network under various conditions of the chain length and the carrier sensing range. Our results show that the capacity, viz. its maximum attainable throughput, reaches a peak at a given workload and then, for larger values of workload, decreases towards an asymptote which value can be drastically lower than the peak value. We give insights into this performance collapse and show the need for a suitable admission control [18].

6.11. Characterisation and Application of Idle Period Durations in IEEE 802.11 DCF-based Multihop Wireless Networks

Multihop wireless networks are used to provide internet connectivity to the users and the level of performance and quality expected by these users are increasing. In order to meet these performance and quality requirements, wireless communications should be enhanced. Previous works from the literature show that the performance and quality provided by an IEEE 802.11-based multihop wireless network are far from optimal and that there exist different ways to increase the efficiency and the quality of service of such a network. Some studies show that using the medium state as a parameter to tune the behaviour of an IEEE 802.11-based multihop network is an appropriate way to proceed. A station in a IEEE 802.11-based multihop wireless network senses the medium either busy or idle. The durations of idle periods and busy periods and their distributions have a clear impact on the network and nodes performance. The understanding of the relationship between these indicators, namely idle and busy periods, the network topology and the traffic, would give new insights to enhance the performance and quality of multihop wireless networks. Due to its multihop and distributed nature, the characterisation of idle period durations is difficult in such a network. This work explores the characterisation of idle period distribution by proposing a new analytical model and provides an application of this characterisation with the design of an adaptive backoff algorithm based on idle periods [30].

MESCAL Project-Team

6. New Results

6.1. Analysis and Control of Large Stochastic Systems

Perfect sampling 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. Even though, the general (non-monotone) case needs to consider the whole state space, we developed a new approach for the general case that only needs to consider two trajectories, an approach which is particularly effective when the state space can be partitioned into pieces where envelopes can be easily computed [8]. Importantly, we also showed that perfect sampling is possible in Jackson networks, even though the underlying Markov chain has a large or even infinite state space and illustrated the efficiency of our approach via numerical simulations [17]. In a similar vein, given that the analysis of a system's dynamics relies on the collection and the description of events, we developed in [37] a new approach to reduce the descriptiveness of a system by aggregating events' properties, such as their Shannon entropy, entropy gain, divergence etc. These measures were applied to the evaluation of geographic aggregations in the context of news analysis and they allowed us to determine which abstractions one should prefer depending on the task to perform.

In the study of Markov decision processes composed of a large number of objects, we showed that the optimal reward satisfies a Bellman equation, which converges to the solution of a continuous Hamilton-Jacobi-Bellman (HJB) equation based on the mean field approximation of the Markov decision process [10]. We also gave bounds on the difference of the rewards and an algorithm for deriving an approximating solution to the Markov decision process from a solution of the HJB equations. Furthermore, we also studied deterministic limits of Markov processes with discontinuous drifts and showed that under mild assumptions, the stochastic system is a constant-step stochastic approximation algorithm which converges to a differential inclusion obtained by convexifying the rescaled drift of the Markov chain [9].

Finally, in terms of performance evaluation and its applications, we also studied resource-aware business process models by defining a new framework that allows the generation of analytical models. We showed that the analysis of the generated SAN model provides several performance indices we showed that these indices can be easily calculated by a business specialist with no skills in stochastic modeling [7].

6.2. Game Theory and Applications

As far as results in pure game theory are concerned, we studied in [12] a general framework of systems wherein there exists a Pareto optimal allocation that is Pareto superior to an inefficient Nash equilibrium and defined a 'Nash proportionately fair' Pareto optima. In this context, we provided conditions for the existence of a Pareto-optimal allocation that is, truly or most closely, proportional to a Nash equilibrium – an approach with applications in non-cooperative flow-control problems in communication networks.

In a learning context, we also explored what happens beyond the standard first-order framework of continuous time game dynamics and introduced in [42] a class of higher order game dynamics, extending all first order imitative dynamics, and, in particular, the replicator dynamics to higher orders. In stark contrast to the first order case, we showed that weakly dominated strategies become eliminated in all n -th order payoff-monotonic dynamics for all $n > 1$ and strictly dominated strategies become extinct in n -th order dynamics n orders as fast as in first order. Finally, we also established a higher order analogue of the folk theorem of evolutionary game theory which shows that higher order accelerate the rate of convergence to equilibria in games.

In terms of applications, we also examined the distribution of traffic in networks whose users try to minimise their delays by adhering to a simple learning scheme inspired by the replicator dynamics of evolutionary game theory. A major challenge occurs in this context when the users' delays fluctuate unpredictably due to random external factors, but we showed that if users are not too greedy in their learning scheme, then the long-term averages of the users' traffic flows converge to the vicinity of an equilibrium [43].

6.3. Wireless networks

Power and energy considerations in wireless networks have brought to the forefront the need for efficient power allocation and handover policies.

In [13], we analyze the power allocation problem for orthogonal multiple access channels by means of a non-cooperative potential game in which each user distributes his power over the channels available to him. When the channels are static, we show that this game possesses a unique optimum point; moreover, if the network's users follow a distributed learning scheme based on the replicator dynamics of evolutionary game theory, then they converge to this optimum exponentially fast.

On the other hand, in case the network users have access to multiple-antenna technologies (as most smartphone users do nowadays), we also analyze the problem of finding the optimal signal covariance matrix for MIMO multiple access channels by using an approach based on "exponential learning" – a novel optimization method which applies more generally to (quasi-)convex problems defined over sets of positive-definite matrices (with or without trace constraints) [24]. Furthermore, by using a Riemannian-geometric approach, we devise a distributed optimization algorithm which converges to the optimum signal distribution exponentially fast: users attain an ϵ -neighborhood of the system's optimum configuration in time which is at most $\mathcal{O}(\log(1/\epsilon))$ (and, in practice, within only a few iterations, even for large numbers of users) [25].

In the context of heterogeneous wireless networks where vertical handovers are allowed, we also studied a control problem for a new joint admission and resource allocation controller. To account for multi-objective optimization, we considered the maximization of an objective subject to a set of constraints, and we turned this constrained problem into an unconstrained one that we solved numerically using the Semi-Markovian Decision Process (SMDP) framework [19].

6.4. Scheduling

The parallel computing platforms available today are increasingly larger, so it is necessary to develop efficient strategies providing safe and reliable completion for parallel applications. In [6], we proposed a performance model that expresses formally the checkpoint scheduling problem by exhibiting the tradeoff between the impact of the checkpoints operations and the lost computation due to failures. In particular, we proved that the checkpoint scheduling problem is NP-hard even in the simple case of uniform failure distribution and also presented a dynamic programming scheme for determining the optimal checkpointing times in all variants of the problem. On a similar issue, we proposed in [35] a fair scheduling algorithm that handles the problem of fair scheduling by adopting processor fair-share as a strategy for user fairness. Our approach showed that a parallel machine can give a similar type of performance guarantee as a round-robin scheduler, without requiring job preemption been required.

From a network calculus perspective, we presented in [16] a new formalism for data packetization in networks, the "packet curves". Indeed, a more precise knowledge of the packet characteristics can be efficiently exploited to get tighter performance bounds, for example for aggregation of flows, packet-based service policies and shared buffers; finally, we also gave a model for a wormhole switch and showed how our results can be used to get efficient delay bounds.

6.5. Multi-Core Systems

Modern multi-core platforms feature complex topologies with different cache levels and hierarchical memory subsystems, so thread and data placement become crucial to achieve good performance. In [14], we evaluate CPU and memory affinity strategies for numerical scientific multithreaded benchmarks on multi-core platforms and analyzed hardware performance event counters in order to acquire a better understanding of such impact. Likewise, thread mapping is an appealing approach to efficiently exploit the potential of modern chip-multiprocessors, so we proposed in [18] a dynamic thread mapping approach to automatically infer a suitable thread mapping strategy for transactional memory applications composed of multiple execution phases with potentially different transactional behavior in each phase. Our results showed that the proposed dynamic

approach presents performance improvements up to 31% compared to the best static solution. From an optimization perspective, the asymmetry in memory access latencies may reduce the overall performance of the system. Therefore, to achieve scalable performance in this environment, we exploited in [28] the machine architecture while taking into account the application communication patterns. Specifically, we introduced a topology-aware asymptotically optimal load balancing algorithm named HwTopoLB which combines the machine topology characteristics with the communication patterns of the application to equalize the application load on the available cores while reducing latencies. We also introduced in [27] a topology-aware load balancer called NucoLB that focuses on redistributing work while reducing communication costs among and within compute nodes, thus leading to performance improvements of up to 20% when compared to state-of-the-art load balancers.

6.6. Cloud Computing

Even though a new era of Cloud Computing has emerged, the characteristics of Cloud load in data centers is not perfectly clear. In [20], we characterized the job/task load and host load in a real-world production data center at Google Inc. by using a detailed trace of over 25 million tasks across over 12,500 hosts. We found that the Google data center exhibits finer resource allocation with respect to CPU and memory than that of Grid/HPC systems and Google jobs are always submitted with much higher frequency and they are much shorter than Grid jobs, leading to higher variance and noise. Moreover, as far as prediction is concerned, we designed in [21] a Bayes model to predict the mean load over a long-term time interval, as well as the mean load in consecutive future time intervals. Real-world experiments showed that our Bayes method achieved high accuracy with a mean squared error of 0.0014 and that it improves the load prediction accuracy by 5.6-50% compared to other state-of-the-art methods based on moving averages, auto-regression, and/or noise filters.

In a similar vein, the exploitation of Best Effort Distributed Computing Infrastructures (BE-DCIs) allows operators to maximize the utilization of the infrastructures, and users to access the unused resources at relatively low cost. Profiling the execution of Bag-of-Tasks (BoT) applications on several kinds of BE-DCIs demonstrates that their task completion rate drops near the end of the execution. In [33], we presented the SpeQuloS service which enhances the QoS of BoT applications executed on BE-DCIs by reducing the execution time, improving its stability, and reporting to users a predicted completion time. We presented the design and development of the framework and several strategies to decide when and how Cloud resources should be provisioned; moreover, performance evaluation using simulations showed that SpeQuloS fulfill its objectives in speeding up the execution of BoTs, in the best cases by a factor greater than 2, while offloading less than 2.5% of the workload to the Cloud. These topics were also further explored in the book chapter [30].

6.7. Experimentation and Visualization in Large Systems

Despite a widespread belief regarding the simulation of large-scale computing systems, we showed in [15] that achieving high scalability does not necessarily require to resort to overly simple models and ignore important phenomena. In fact, by relying on a modular and hierarchical platform representation while taking advantage of regularity when possible, we were able to model systems such as data and computing centers, peer-to-peer networks, grids, or clouds in a scalable way. Finally, in [34], we examined the ability to conduct consistent, controlled, and repeatable large-scale experiments in areas of computer science where availability, repeatability, and open sharing of electronic products are still difficult to achieve.

We also discussed in [22] the concept of the reconstructability of software environments and we proposed a tool for dealing with this problem. In a similar vein, we developed Expo [41], a tool for conducting experiments on distributed platforms. Our experiments confirmed that Expo is a promising tool to help the user with two primary concerns: how to perform a large scale experiment efficiently and easily, together with its reproducibility.

The exponential number of processes that are executed in high performance applications and the very detailed behavior that we can record about them lead to a trace size explosion both in space and time dimensions. Thus, 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. We dealt with this issue in [38] in two ways: first, by detailing data aggregation techniques that are fully configurable by the user to control the level of details in both space and time dimensions, and second, by presenting two visualization techniques that take advantage of the aggregated data to scale.

Furthermore, given that the performance of parallel and distributed applications is highly dependent on the characteristics of the execution environment, the network topology and characteristics directly impact data locality and movements as well as contention. Unfortunately few visualization available to the analyst are capable of accounting for such phenomena, so we proposed in [39] 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. Such visualization techniques enable us to explore and understand non-trivial behaviors that are impossible to grasp otherwise and the combination of multi-scale aggregation and dynamic graph layout allows us to scale the visualization seamlessly to large distributed systems.

MOAIS Project-Team

6. New Results

6.1. Work Stealing inside GPU

Graphics Processing units (GPU) have become a valuable support for High Performance Computing (HPC) applications. However, despite the many improvements of General Purpose GPUs, the current programming paradigms available, such as NVIDIA's CUDA, are still low-level and require strong programming effort, especially for irregular applications where dynamic load balancing is a key point to reach high performances. We have introduced a new hybrid programming scheme for general purpose graphics processors using two levels of parallelism. In the upper level, a program creates, in a lazy fashion, tasks to be scheduled on the different Streaming Multiprocessors (MP), as defined in the NVIDIA's architecture. We have embedded inside GPU a well-known work stealing algorithm to dynamically balance the workload. At lower level, tasks exploit each Streaming Processor (SP) following a data-parallel approach. Preliminary comparisons on data-parallel iteration over vectors show that this approach is competitive on regular workload over the standard CUDA library Thrust, based on a static scheduling. Nevertheless, our approach outperforms Thrust-based scheduling on irregular workloads.

6.2. XKaapi on top of Multi-CPU Multi-GPU

Most recent HPC platforms have heterogeneous nodes composed of a combination of multi-core CPUs and accelerators, like GPUs. Programming such nodes is typically based on a combination of OpenMP and CUDA/OpenCL codes; scheduling relies on a static partitioning and cost model. We have experiment XKaapi runtime system for multi-CPU and multi-GPU architectures, which supports a data-flow task model and a locality-aware work stealing scheduler. The XKaapi enables task multi-implementation on CPU or GPU and multi-level parallelism with different grain sizes. We demonstrate performance results on two dense linear algebra kernels, matrix product (GEMM) and Cholesky factorization (POTRF), to evaluate XKaapi on a heterogeneous architecture composed of two hexa-core CPUs and eight NVIDIA Fermi GPUs. Our conclusion is two-fold: First, fine grained parallelism and online scheduling achieve performance results as good as static strategies, and in most cases outperform them. This is due to an improved work stealing strategy that includes locality information; to a very light implementation of the tasks in XKaapi; and to an optimized search for ready tasks. Next, our XKaapi Cholesky is highly efficient on multi-CPU/multi-GPU due to its multi-level parallelism. Using eight NVIDIA Fermi GPUs and four CPUs, we measure up to 2.43 TFlop/s on double precision matrix product and 1.79 TFlop/s on Cholesky factorization; and respectively 5.09 TFlop/s and 3.92 TFlop/s in single precision. This is the first time that such a performance is obtained with more than four GPUs.

6.3. Formalizing the concept of cooperation

We study how to optimize scheduling problems for a large number of objectives, when multiple users are competing for common resources, with some appropriate notion of fairness between users. Formalizing the concept of cooperation in relation with multi-objective optimization, we can refine the classical methods in combinatorial optimization (that usually optimize one centralized objective) by introducing extra features (adding more objectives or constraints). The PhD thesis of Daniel Cordeiro [2] proposed various ways for handling this problem: multi-organization scheduling and its relaxed variant, impact of selfishness. In the same context, we investigated the field of Game Theory through the existence of Nash equilibria in some situations.

6.4. Fault-tolerance for large parallel systems

This PhD thesis of Slim Bouguerra [1] studied fault-tolerance issues for large parallel systems. We revisited, via a formal proof, the old well-known result which states that the optimal policy for exponential failure law is to put the check-points at periodic moments. We proposed new algorithms to handle check-points for any law in the input and variable check-point costs (JPDC paper).

PLANETE Project-Team

6. New Results

6.1. Towards Data-Centric Networking

Participants: Chadi Barakat, Damien Saucez, Jonathan Detchart, Mohamed Ali Kaafar, Ferdaouss Mattoussi, Marc Mendonca, Xuan-Nam Nguyen, Vincent Roca, Thierry Turletti.

- **DTN**

Delay Tolerant Networks (DTNs) stand for wireless networks where disconnections may occur frequently. In order to achieve data delivery in such challenging environments, researchers have proposed the use of store-carry-and-forward protocols: there, a node may store a message in its buffer and carry it along for long periods of time, until an appropriate forwarding opportunity arises. Multiple message replicas are often propagated to increase delivery probability. This combination of long-term storage and replication imposes a high storage and bandwidth overhead. Thus, efficient scheduling and drop policies are necessary to: (i) decide on the order by which messages should be replicated when contact durations are limited, and (ii) which messages should be discarded when nodes' buffers operate close to their capacity.

We worked on an optimal scheduling and drop policy that can optimize different performance metrics, such as the average delivery rate and the average delivery delay. First, we derived an optimal policy using global knowledge about the network, then we introduced a distributed algorithm that collects statistics about network history and uses appropriate estimators for the global knowledge required by the optimal policy, in practice. At the end, we are able to associate to each message inside the network a utility value that can be calculated locally, and that allows to compare it to other messages upon scheduling and buffer congestion. Our solution called HBSD (History Based Scheduling and Drop) integrates methods to reduce the overhead of the history-collection plane and to adapt to network conditions. The first version of HBSD and the theory behind have been published in 2008. A recent paper [27] provides an extension to a heterogenous mobility scenario in addition to refinements to the history collection algorithm. An implementation is proposed for the DTN2 architecture as an external router and experiments have been carried out by both real trace driven simulations and experiments over the SCORPION testbed at the University of California Santa Cruz. We refer to the web page of HBSD for more details http://planete.inria.fr/HBSD_DTN2/.

HBSD in its current version is for point-to-point communications. Another interesting schema is to consider one-to-many communications, where requesters for content express their interests to the network, which looks for the content on their behalf and delivers it back to them. Along the main ideas of HBSD, we worked on a content optimal-delivery algorithm, CODA, that distributes content to multiple receivers over a DTN. CODA assigns a utility to each content item published in the network; this value gauges the contribution of a single content replica to the network's overall delivery-rate. CODA performs buffer management by first calculating the delivery-rate utility of each cached content-replica and then discarding the least-useful item. When an application requests content, the node supporting the application will look for the content in its cache. It will immediately deliver it to the application if the content is stored in memory. In case the request cannot be satisfied immediately, the node will store the pending request in a table. When the node meets another device, it will send the list of all pending requests to its peer; the peer device will try to satisfy this list by sending the requester all the matching content stored in its own buffer. A meeting between a pair of devices might not last long enough for all requested content to be sent. We address this problem by sequencing transmissions of data in order of decreasing delivery-rate utility. A content item with few replicas in the network has a high delivery rate utility; these items must be transmitted first to avoid degrading the content delivery-rate metric. The node delivers the requested content to the application

as soon as it receives it in its buffer. We implement CODA over the CCNx protocol, which provides the basic tools for requesting, storing, and forwarding content. Detailed information on CODA and the implementation work carried out herein can be found in [76].

- **Naming and Routing in Content Centric Networks**

Content distribution prevails in today's Internet and content oriented networking proposes to access data directly by their content name instead of their location, changing so the way routing must be conceived. We proposed a routing mechanism that faces the new challenge of interconnecting content-oriented networks. Our solution relies on a naming resolution infrastructure that provides the binding between the content name and the content networks that can provide it. Content-oriented messages are sent encapsulated in IP packets between the content-oriented networks. In order to allow scalability and policy management, as well as traffic popularity independence, binding requests are always transmitted to the content owner. The content owner can then dynamically learn the caches in the network and adapt its binding to leverage the cache use.

The work done so far is related to routing between content-oriented networks. We are starting an activity on how to provide routing inside a content network. To that aim, we are investigating on the one hand probabilistic routing and, on the other hand, deterministic routing and possible extension to Bellman-Ford techniques. In addition to routing, we are investigating the problem of congestion in content-oriented networks. Indeed, in this new paradigm, congestion must be controlled on a per-hop basis, as opposed to the end-to-end congestion control that prevails today. We think that we can combine routing and congestion control to optimize resource consumption. Finally, we are studying the implications of using CCN from an economical perspective. See [100] for more details.

- **On the fairness of CCN**

Content-centric networking (CCN) is a new paradigm to better handle contents in the future Internet. Under the assumption that CCN networks will deploy a similar congestion control mechanism than in today's TCP/IP (i.e., AIMD), we built an analytical model of the bandwidth sharing in CCN based on the "square-root formula of TCP". With this model we can compare CCN download performance to what users get today. We consider different factors such as the way CCN routers are deployed, the popularity of contents, or the capacity of links and observe that when AIMD is used in a CCN network less popular content throughput is massively penalised whilst the individual gain for popular content is negligible. Finally, the main advantage of using CCN is the decrease of load at the server side. Our observations advocate the necessity to clearly define the notion of fairness in CCN and to design a proper congestion control to avoid less popular contents to become hardly accessible in tomorrow's Internet.

Our results [75] clearly point to a fairness issue if AIMD is used with CCN. Indeed, combining blindly AIMD and CCN can severely worsen the download throughput of less popular contents with respect to the today's Internet due to subtle interactions with in-network caching strategies. The way cache memories are distributed within chain topologies has been investigated too, showing that for small and heterogeneous cache spaces, placing the biggest caches close to clients improves performance due to a smaller RTT on average. On the other hand, CCN can significantly reduce the load at the server side independently of the cache allocation strategy. Our findings advocate the urge of clearly defining the notion of fairness in CCN and designing congestion control algorithms able to limit the unfairness observed between contents of different popularities. The work is currently used within the IRTF ICNRG research group in order to motivate and define an appropriate congestion control mechanism for information centric networks like CCN. Moreover, we are currently validating the analytical results with an implementation of CCN where we can evaluate how much our model

deviates from the reality when contents are of various size or small. The implementation will also be a support to test different congestion control mechanism.

- **CCN to enable profitable collaborative OTT services**

The ubiquity of broadband Internet and the proliferation of connected devices like laptops, tablets, or TV result in a high demand of multimedia content such as high definition video on demand (VOD) for which the Internet has been poorly designed with the Internet Protocol (IP). Information-Centric Networking and more precisely Content Centric Networking (CCN) overtake the limitation of IP by considering content as the essential element of the network instead of the topology. CCN and its content caching capabilities is particularly adapted to Over-The-Top (OTT) services like Netflix, Hulu, Xbox Live, or YouTube that distribute high-definition multimedia content to millions of consumers, independently of their location. However, bringing content as the most important component of the network implies fundamental changes in the Internet and the transition to a fully CCN Internet might take a long time. Despite this transition period where CCN and IP will co-exist, we have shown that OTT service providers and consumers have strong incentives for migrating to CCN. We also propose a transition mechanism based on the Locator/Identifier Separation Protocol (LISP) [28] that allows the provider to track the demands from its consumers even though they do not download the contents from another consumers instead of the producer itself.

CCN, compared to IP, provides better security and performance. This last point is very interesting for OTT service providers that deliver multimedia content where performance is a key factor for the adoption of the service by consumers. With CCN, the content can be retrieved from the caches in the different CCN islands, instead of always being delivered by the content publisher. As a result, content retrieval is faster for the consumer and the operational cost of the publisher is reduced. Moreover, as the content is cached by the consumers and because the consumer can provide the content to other consumers, the overall performance increases with the number of consumers instead of decreasing as it is the case in IP today where the content is delivered by the hosting server. This property is particularly interesting because it dampens the effect of flash crowds which are normally very costly for OTT service providers as they have to provision their servers and networks to support them. Using CCN with caching at the consumers has then a direct impact on the profit earned by the OTT service provider as its costs are reduced. However, to benefit from the caching capabilities of consumers, the producer must propose real incentives to its consumers to *collaborate* and cache the content. To understand how incentives can be provided, it is necessary to remember that content in OTT is provided either freely to the consumer or in exchange of a fee. When the content is provided freely, the incomes for the publisher are ensured by advertisements dispersed in the content (e.g., banner, commercial interruptions...). A consumer has incentives to collaborate with the system if it receives some sort of discount, expressed in advertisement reduction or fee reduction. On the one hand, the discount has a cost for the publisher as its revenues will be reduced. On the other hand, the collaboration from its consumers reduces its operational costs. Hence, the publisher must determine the optimal discount, such that it maximises its profit. The situation for the consumer is the exact opposite: its costs are increasing because it is providing content to other consumers but its revenues also increase as it receives a discount on its expenses. We have determined the conditions to respect when deploying OTT with loosely collaborative consumers [99]. We currently refine the results using game theory.

- **Software-Defined Networking in Heterogeneous Networked Environments**

Software-Defined Networking (SDN) has been proposed as a way to facilitate network evolution by allowing networks and their infrastructure to be programmable. In the context of the COMMUNITY associated team with University of California Santa Cruz (see URL <http://inrg.cse.ucsc.edu/community/>), we are studying the potential of SDN to facilitate the deployment and management of new architectures and services in heterogeneous environments. In particular, we focus on the fundamental issues related to enabling SDN in infrastructure-less/decentralized networked environments and we use OpenFlow as our target SDN platform. Our plan is to develop a hybrid SDN framework that strikes a balance between a completely decentralized approach like Active Networking and a centralized one such as OpenFlow~[58].

We are also currently evaluating the efficiency of SDN for optimizing caching in content-centric networks. CCN advocates in-network caching, i.e., to cache contents on the path from content providers to requesters. Although this on-path caching achieves good overall performance, we have shown that this strategy is far from being the optimal inside a domain. On this purpose, we proposed the notion of off-path caching by allowing deflection of the most popular traffic off the optimal path towards off-path caches available across the domain[100]. Off-path caching improves the global hit ratio and permits to reduce the peering links' bandwidth usage. We are now investigating whether SDN functionalities can be used to implement this optimal caching technique, in particular to identify of the most popular contents, and to configure deflection mechanisms within routers~[94].

- **Application-Level Forward Error Correction Codes (AL-FEC) and their Applications to Broadcast/Multicast Systems**

With the advent of broadcast/multicast systems (e.g., 3GPP MBMS services), large scale content broadcasting is becoming a key technology. This type of data distribution scheme largely relies on the use of Application Level Forward Error Correction codes (AL-FEC), not only to recover from erasures but also to improve the content broadcasting scheme itself (e.g., with FLUTE/ALC).

Our LDPC-Staircase codes, that offer a good balance in terms of performance, have been included as the primary AL-FEC solution for ISDB-Tmm (Integrated Services Digital Broadcasting, Terrestrial Mobile Multimedia), a Japanese standard for digital television (DTV) and digital radio, with a commercial service that started in April 2012. This is the first adoption of these codes in an international standard. These codes, along with our FLUTE/ALC software, are now part of the server and terminal protocol stack: <http://www.rapidtvnews.com/index.php/2012041721327/ntt-data-mse-and-expways-joint-solution-powers-japanese-mobile-tv-service.html>.

This success has been made possible, on the one hand, by major efforts in terms of standardization within IETF: the RFC 5170 (2008) defines the codes and their use in FLUTE/ALC, a protocol stack for massively scalable and reliable content delivery services, an active Internet-Draft published last year describes the use of these AL-FEC codes in FECFRAME, a framework for robust real-time streaming applications, and recent Internet-Drafts [91][92] define the GOE (Generalized Object Encoding) extension of LDPC-Staircase codes for UEP (Unequal Erasure Protection) and file bundle protection services.

This success has also been made possible, on the other hand, by our efforts in terms of design and evaluation of two efficient software codecs for LDPC-Staircase codes. One of them is distributed in open-source, as part of our OpenFEC project (<http://openfec.org>), a unique initiative that aims at promoting open and free AL-FEC solutions. The second one, a highly optimized version with improved decoding speed and reduced memory requirements, is commercialized through an industrial partner, Expway.

Since May 2012, along with the Expway French company, we are proposing the Reed-Solomon + LDPC-Staircase codes for the 3GPP-eMBMS call for technology, as a candidate for next generation AL-FEC codes for multimedia services. We have shown that these codes offer very good erasure

recovery capabilities, in line with 3GPP requirements, and extremely high decoding speeds, usually significantly faster than that of the other proposals. The final decision is expected for end of January 2013. In any case we have once again showed that these codes provide very good performance, often ahead of the competitors, and an excellent balance between several technical and non technical criteria.

Finally our activities in the context of the PhD of F. Mattoussi include the design, analysis and improvement of GLDPC-Staircase codes, a "Generalized" extension to LDPC-Staircase codes. We have shown in particular that these codes: (1) offer small rate capabilities, i.e. can produce a large number of repair symbols 'on-the-fly', when needed; (2) feature high erasure recovery capabilities, close to that of ideal codes. Therefore they offer a nice opportunity to extend the field of application of existing LDPC-Staircase codes (IETF RFC 5170), while keeping backward compatibility (i.e. LDPC-Staircase "codewords" can be decoded with a GPLDPC-Staircase codec). More information is available in [56][57][55].

- **Unequal Erasure Protection (UEP) and File bundle protection through the GOE (Generalized Object Encoding) scheme**

This activity has been initiated with the PostDoc work of Rodrigue IMAD. It focuses on Unequal Erasure Protection capabilities (UEP) (when a subset of an object has more importance than the remaining) and file bundle protection capabilities (e.g. when one wants to globally protect a large set of small objects).

After an in-depth understanding of the well-known PET (Priority Encoding Technique) scheme, and the UOD for RaptorQ (Universal Object Delivery) initiative of Qualcomm, which is a realization of the PET approach, we have designed the GOE FEC Scheme (Generalized Object Encoding) alternative. The idea, simple, is to decouple the FEC protection from the natural object boundaries, and to apply an independent FEC encoding to each "generalized object". The main difficulty is to find an appropriate signaling solution to synchronize the sender and receiver on the exact way FEC encoding is applied. In [91] we show this is feasible, while keeping a backward compatibility with receivers that do not support GOE FEC schemes. Two well known AL-FEC schemes have also been extended to support this new approach, with very minimal modifications, namely Reed-Solomon and LDPC-Staircase codes [92], [91].

During this work, we compared the GOE and UOD/PET schemes, both from an analytical point of view (we use an N-truncated negative binomial distribution to that purpose) and from an experimental, simulation based, point of view [64]. We have shown that the GOE approach, by the flexibility it offers, its simplicity, its backward compatibility and its good recovery capabilities (under finite or infinite length conditions), outperforms UOD/PET for practical realizations of UEP/file bundle protection systems. See also <http://www.ietf.org/proceedings/81/slides/rmt-2.pdf>.

- **Application-Level Forward Error Correction Codes (AL-FEC) and their Applications to Robust Streaming Systems**

AL-FEC codes are known to be useful to protect time-constrained flows. The goal of the IETF FECFRAME working group is to design a generic framework to enable various kinds of AL-FEC schemes to be integrated within RTP/UDP (or similar) data flows. Our contributions in the IETF context are three fold. First of all, we have contributed to the design and standardization of the FECFRAME framework, now published as a Standards Track RFC6363.

Secondly, we have proposed the use of Reed-Solomon codes (with and without RTP encapsulation of repair packets) and LDPC-Staircase codes within the FECFRAME framework: [85] for Reed-Solomon and [88] for LDPC-Staircase. Both documents are close to being published as RFCs.

Finally, in parallel, we have started an implementation of the FECFRAME framework in order to gain an in-depth understanding of the system. Previous results showed the benefits of LDPC-Staircase codes when dealing with high bit-rate real-time flows.

A second type of activity, in the context of robust streaming systems, consisted in the analysis of the Tetrys approach. Tetrys is a promising technique that features high reliability while being independent from RTT, and performs better than traditional block FEC techniques in a wide range of operational conditions.

- **A new File Delivery Application for Broadcast/Multicast Systems**

FLUTE [95] has long been the one and only official file delivery application on top of the ALC reliable multicast transport protocol. However FLUTE has several limitations (essentially because the object meta-data are transmitted independently of the objects themselves, in spite of their interdependency), features an intrinsic complexity, and is only available for ALC.

Therefore, we started the design of FCAST, a simple, lightweight file transfer application, that works both on top of both ALC and NORM [82]. This work is carried out as part of the IETF RMT Working Group, in collaboration with B. Adamson (NRL). This document has passed WG Last Call and is currently considered by IESG.

- **Security of the Broadcast/Multicast Systems**

Sooner or later, broadcasting systems will require security services. This is all the more true as heterogeneous broadcasting technologies are used, some of them being by nature open, such as WiFi networks. Therefore, one of the key security services is the authentication of the packet origin and the packet integrity check. To that purpose, we have specified the use of simple authentication and integrity schemes (i.e., group MAC and digital signatures) in the context of the ALC and NORM protocols and the standard is now published as IETF RFC 6584 [98].

- **High Performance Security Gateways for High Assurance Environments**

This work focuses on very high performance security gateways, compatible with 10Gbps or higher IPsec tunneling throughput, while offering a high assurance thanks in particular to a clear red/black flow separation. In this context we have studied last year the feasibility of high-bandwidth, secure communications on generic machines equipped with the latest CPUs and General-Purpose Graphical Processing Units (GPGPU).

The work carried out in 2011-2012 consisted in setting up and evaluating the high performance platform. This platform heavily relies on the Click modular TCP/IP protocol stack implementation, which turned out to be a key enabler both in terms of specialization of the stack and parallel processing. Our activities also consisted in analyzing the PMTU discovery aspect since it is a critical factor in achieving high bandwidths. To that goal we have designed a new approach for qualifying ICMP blackholes in the Internet, since PMTUD heavily relies on ICMP [51].

6.2. Network Security and Privacy

Participants: Claude Castelluccia, Gergely Acs, Mathieu Cunche, Daniele Perito, Lukasz Olejnik, Mohamed Ali Kaafar, Abdelberi Chaabane, Cédric Lauradoux, Minh-Dung Tran.

- *Private Big Data Publication* Public datasets are used in a variety of applications spanning from genome and web usage analysis to location-based and recommendation systems. Publishing such datasets is important since they can help us analyzing and understanding interesting patterns. For example, mobility trajectories have become widely collected in recent years and have opened the possibility to improve our understanding of large-scale social networks by investigating how people exchange information, interact, and develop social interactions. With billion of handsets in use worldwide, the quantity of mobility data is gigantic. When aggregated, they can help understand complex processes, such as the spread of viruses, and build better transportation systems, prevent traffic congestion. While the benefits provided by these datasets are indisputable, they unfortunately pose a considerable threat to individual privacy. In fact, mobility trajectories might be used by a malicious attacker to discover potential sensitive information about a user, such as his habits, religion or relationships. Because privacy is so important to people, companies and researchers are reluctant to publish datasets by fear of being held responsible for potential privacy breaches. As a result, only very few of them are actually released and available. This limits our ability to analyze such data to derive information that could benefit the general public. Here follows some recent results of our activities in this domain.

Privacy-Preserving Sequential Data Publication [41]: Sequential data is being increasingly used in a variety of applications, spanning from genome and web usage analysis to location-based recommendation systems. Publishing sequential data is of vital importance to the advancement of these applications since they can enable researchers to analyze and understand interesting sequential patterns. However, as shown by the re-identification attacks on the AOL and Netflix datasets, releasing sequential data may pose considerable threats to individual privacy. Recent research has indicated the failure of existing sanitization techniques to provide claimed privacy guarantees. It is therefore urgent to respond to this failure by developing new schemes with provable privacy guarantees. Differential privacy is one of the only models that can be used to provide such guarantees. Due to the inherent sequentiality and high-dimensionality, it is challenging to apply differential privacy to sequential data. In this work, we address this challenge by employing a variable-length n-gram model, which extracts the essential information of a sequential database in terms of a set of variable-length n-grams. Our approach makes use of a carefully designed exploration tree structure and a set of novel techniques based on the Markov assumption in order to lower the magnitude of added noise. The published n-grams are useful for many purposes. Furthermore, we develop a solution for generating a synthetic database, which enables a wider spectrum of data analysis tasks. Extensive experiments on real-life datasets demonstrate that our approach substantially outperforms the state-of-the-art techniques.

Private Histogram Publishing [33]:

Differential privacy can be used to release different types of data, and, in particular, histograms, which provide useful summaries of a dataset. Several differentially private histogram releasing schemes have been proposed recently. However, most of them directly add noise to the histogram counts, resulting in undesirable accuracy. In this work, we propose two sanitization techniques that exploit the inherent redundancy of real-life datasets in order to boost the accuracy of histograms. They lossily compress the data and sanitize the compressed data. Our first scheme is an optimization of the Fourier Perturbation Algorithm (FPA) presented in [13]. It improves the accuracy of the initial FPA by a factor of 10. The other scheme relies on clustering and exploits the redundancy between bins. Our extensive experimental evaluation over various real-life and synthetic datasets demonstrates that our techniques preserve very accurate distributions and considerably improve the accuracy of range queries over attributed histograms.

- *Privacy Issues on the Internet* Internet users are being increasingly tracked and profiled. Companies utilize profiling to provide customized, i.e. personalized services to their customers, and hence increase revenues.

Privacy issues of Targeted Advertising [37]: Behavioral advertising takes advantage from profiles of users' interests, characteristics (such as gender, age and ethnicity) and purchasing activities. For example, advertising or publishing companies use behavioral targeting to display advertisements that closely reflect users' interests (e.g. 'sports enthusiasts'). Typically, these interests are inferred from users' web browsing activities, which in turn allows building of users' profiles. It can be argued that customization resulting from profiling is also beneficial to users who receive useful information and relevant online ads in line with their interests. However, behavioral targeting is often perceived as a threat to privacy mainly because it heavily relies on users' personal information, collected by only a few companies. In this work, we show that behavioral advertising poses an additional privacy threat because targeted ads expose users' private data to any entity that has access to a small portion of these ads. More specifically, we show that an adversary who has access to a user's targeted ads can retrieve a large part of his interest profile. This constitutes a privacy breach because interest profiles often contain private and sensitive information.

On the Uniqueness of Web Browsing History Patterns [60]: We present the results of the first large-scale study of the uniqueness of Web browsing histories, gathered from a total of 368,284 Internet users who visited a history detection demonstration website. Our results show that for a majority of users (69%), the browsing history is unique and that users for whom we could detect at least 4 visited websites were uniquely identified by their histories in 97% of cases. We observe a high rate of stability in browser history fingerprints: for repeat visitors, 80% of fingerprints are identical over time, and differing ones were strongly correlated with original history contents, indicating static browsing preferences. We report a striking result that it is enough to test for a small number of pages in order to both enumerate users' interests and perform an efficient and unique behavioral fingerprint; we show that testing 50 web pages is enough to fingerprint 42% of users in our database, increasing to 70% with 500 web pages. Finally, we show that indirect history data, such as information about *categories* of visited websites can also be effective in fingerprinting users, and that similar fingerprinting can be performed by common script providers such as Google or Facebook.

- **Adaptive Password-Strength Meters from Markov Models [38]**

Passwords are a traditional and widespread method of authentication, both on the Internet and off-line. Passwords are portable, easy to understand for laypersons, and easy to implement for the operator. Thus, password-based authentication is likely to stay for the foreseeable future.

To ensure an acceptable level of security of user-chosen passwords, sites often use mechanisms to test the strength of a password (often called *pro-active password checkers*) and then reject weak passwords. Hopefully this ensures that passwords are reasonably strong on average and makes guessing passwords infeasible or at least too expensive for the adversary. Commonly used password checkers rely on rules such as requiring a number and a special character to be used. However, as we will show and also has been observed in previous work, the accuracy of such password checkers is low, which means that often insecure passwords are accepted and secure passwords are rejected. This adversely affects both security and usability.

In this work, we propose to use password strength meters based on Markov-models, which estimate the true strength of a password more accurately than rule-based strength meters. Roughly speaking, the Markov-model estimates the strength of a password by estimating the probability of the n -grams that compose said password. Best results can be obtained when the Markov-models are trained on the actual password database. We show, in this work, how to do so without sacrificing the security of the password database, even when the n -gram database is leaked.

We show how to build secure adaptive password strength meters, where security should hold even when the n -gram database leaks. This is similar to traditional password databases, where one tries

to minimize the effects of a database breach by hashing and salting the stored passwords. This is not a trivial task. One potential problem is that, particularly strong passwords, can be leaked entirely by an n -gram database (without noise added).

- **Fast Zero-Knowledge Authentication [47]** We explore new area/throughput trade-offs for the Girault, Poupard and Stern authentication protocol (GPS). This authentication protocol was selected in the NESSIE competition and is even part of the standard ISO/IEC 9798. The originality of our work comes from the fact that we exploit a fixed key to increase the throughput. It leads us to implement GPS using the Chapman constant multiplier. This parallel implementation is 40 times faster but 10 times bigger than the reference serial one. We propose to serialize this multiplier to reduce its area at the cost of lower throughput. Our hybrid Chapman's multiplier is 8 times faster but only twice bigger than the reference. Results presented here allow designers to adapt the performance of GPS authentication to their hardware resources. The complete GPS prover side is also integrated in the network stack of the PowWow sensor which contains an Actel IGLOO AGL250 FPGA as a proof of concept.

- **Energy Efficient Authentication Strategies for Network Coding [26]**

Recent advances in information theory and networking, e.g. aggregation, network coding or rateless codes, have significantly modified data dissemination in wireless networks. These new paradigms create new threats for security such as pollution attacks and denial of services (DoS). These attacks exploit the difficulty to authenticate data in such contexts. The particular case of xor network coding is considered herein. We investigate different strategies based on message authentication codes algorithms (MACs) to thwart these attacks. Yet, classical MAC designs are not compatible with the linear combination of network coding. Fortunately, MACs based on universal hash functions (UHF) match nicely the needs of network coding: some of these functions are linear $h(x_1 \oplus x_2) = h(x_1) \oplus h(x_2)$. To demonstrate their efficiency, we consider the case of wireless sensor networks (WSNs). Although these functions can drastically reduce the energy consumption of authentication (up to 68% gain over the classical designs is observed), they increase the threat of DoS. Indeed, an adversary can disrupt all communications by polluting few messages. To overcome this problem, a group testing algorithm is introduced for authentication resulting in a complexity linear in the number of attacks. The energy consumption is analyzed for cross-point and butterfly network topologies with respect to the possible attack scenarios. The results highlight the trade-offs between energy efficiency, authentication and the effective throughput for the different MAC modes.

- **Towards Stronger Jamming Model: Application to TH-UWB Radio [35]**

With the great expansion of wireless communications, jamming becomes a real threat. We propose a new model to evaluate the robustness of a communication system to jamming. The model results in more scenarios to be considered ranging from the favorable case to the worst case. The model is applied to a TH-UWB radio. The performance of such a radio in presence of the different jamming scenarios is analyzed. We introduce a mitigation solution based on stream cipher that restricts the jamming problem of the TH-UWB communication to the more favorable case while preserving confidentiality.

- **Privacy risks quantification in Online social networks**

In this project, we analyze the different capabilities of online social networks and aim to quantify the privacy risks users are undertaking in this context. Online Social Networks (OSNs) are a rich source of information about individuals. It may be difficult to justify the claim that the existence of public profiles breaches the privacy of their owners, as they are the ones who entered the data and made them publicly available in the first place. However, aggregation of multiple OSN public profiles is debatably a source of privacy loss, as profile owners may have expected each profile's information to stay within the boundaries of the OSN service in which it was created. First we present an empirical study of personal information revealed in public profiles of people who use multiple Online Social Networks (OSNs). This study aims to examine how users reveal their personal information across multiple OSNs. We consider the number of publicly available attributes in public

profiles, based on various demographics and show a correlation between the amount of information revealed in OSN profiles and specific occupations and the use of pseudonyms. Then, we measure the complementarity of information across OSNs and contrast it with our observations about users who share a larger amount of information. We also measure the consistency of information revelation patterns across OSNs, finding that users have preferred patterns when revealing information across OSNs. To evaluate the quality of aggregated profiles we introduce a consistency measure for attribute values, and show that aggregation also improves information granularity. Finally, we demonstrate how the availability of multiple OSN profiles can be exploited to improve the success of obtaining users' detailed contact information, by cross-linking with publicly available data sources such as online phone directories. This work has been published in ACM SIGCOMM WOSN [42].

In a second study, we examine the user tracking capabilities of the three major global Online Social Networks (OSNs). We study 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 from the audience. We also evaluate the tracking capabilities in practice and demonstrate by analyzing a real traffic traces that OSNs can reconstruct a significant portion of users web profile and browsing history. We finally provide insights into the relation between the browsing history characteristics and the OSN tracking potential, highlighting the high risk properties. This work has also been published in ACM SIGCOMM WOSN [40].

In a third study, we also analyzed the inference capabilities of third parties from seemingly harmless and unconsciously publicly shared data. Interests (or "likes") of users is one of the highly-available on-line information on the web. In this study, we show how these seemingly harmless interests (e.g., music interests) can leak privacy sensitive information about users. In particular, we infer their undisclosed (private) attributes using the public attributes of other users sharing similar interests. In order to compare user-defined interest names, we extract their semantics using an ontologized version of Wikipedia and measure their similarity by applying a statistical learning method. Besides self-declared interests in music, our technique does not rely on any further information about users such as friend relationships or group belongings. Our experiments, based on more than 104K public profiles collected from Facebook and more than 2000 private profiles provided by volunteers, show that our inference technique efficiently predicts attributes that are very often hidden by users. This is the first time that user interests are used for profiling, and more generally, semantics-driven inference of private data is addressed. Our work received many media attention and was published in the prestigious NDSS symposium [39].

- **On the Privacy threats of hidden information in Wireless communication**

Wi-Fi protocol has the potential to leak personal information. Wi-Fi capable devices commonly use active discovery mode to find the available Wi-Fi access points (APs). This mechanism includes broadcast of the AP names to which the mobile device has previously been connected to, in plain text, which may be easily observed and captured by any Wi-Fi device monitoring the control traffic. The combination of the AP names belonging to any mobile device can be considered as a Wi-Fi fingerprint, which can be used to identify the mobile device user. Our research investigates how it is possible to exploit these fingerprints to identify links between users i.e. owners of the mobile devices broadcasting such links. In this project, we have used an approach based on the similarity between the Wi-Fi fingerprints, which is equated to the likelihood of the corresponding users being linked. When computing the similarity between two Wi-Fi fingerprints, two dimensions need to be considered : (i) The number of network names in common. Indeed, sharing a network is an indication of the existence of a link, e.g. friends and family that share multiple Wi-Fi networks. (ii) The rarity of the network names in common. Some network names are very common and sharing them does not imply a link between the users. This is the case for public network names such as McDonalds Free Wi-Fi, or default network names such as NETGEAR and Linksys. On the other hand, uncommon network names such as Griffin Family Network or Orange-3EF50 are likely to

indicate a strong link between the users of these networks. Utilising a carefully designed similarity metric, we have been able to infer the existence of social links with a high confidence: 80% of the links were detected with an error rate of 7%. We show that through real-life experiments that owners of smartphones are particularly exposed to this threat, as indeed these devices are carried on persons throughout the day, connecting to multiple Wi-Fi networks and also broadcasting their connection history. There are a number of industry and research initiatives aiming to address Wi-Fi related privacy issues. The deployment of new technology i.e. privacy preserving discovery services, would necessitate software modifications in currently deployed APs and devices. The obvious solution to disable active discovery mode, comes at the expense of performance and usability, i.e. with an extended time duration for the Wi-Fi capable device to find and connect to an available AP. As a possible first step, users should be encouraged to remove the obsolete connection history entries, which may lower the similarity metric and thus reduce the ease of linkage. Our papers illustrating this study have been presented in the WoWMoM'12 conference [45] and in the IEEE MILCOM conference [43].

- **Information leakage in Ads networks**

In targeted (or behavioral) advertising, users' behaviors are tracked over time in order to customize served ads to their interests. This creates serious privacy concerns since for the purpose of profiling, private information is collected and centralized by a limited number of companies. Despite claims that this information is secure, there is a potential for this information to be leaked through the customized services these companies are offering. In this study, we show that targeted ads expose users' private data not only to ad providers but also to any entity that has access to users' ads. We propose a methodology to filter targeted ads and infer users' interests from them. We show that an adversary that has access to only a small number of websites containing Google ads can infer users' interests with an accuracy of more than 79% (Precision) and reconstruct as much as 58% of a Google Ads profile in general (Recall). This study is the first work that identifies and quantifies information leakage through ads served in targeted advertising. We published a paper illustrating these results in the prestigious Privacy Enhancing Technologies Symposium PETS 2012 [37].

- **Privacy in P2P file sharing systems**

In this study, we aim at characterizing anonymous file sharing systems from a privacy perspective. We concentrate on a recently deployed privacy-preserving file sharing system: OneSwarm. Our characterisation is based on measurement of several aspects of the OneSwarm system such as the nature of the shared and searched content and the geolocation and number of users. Our findings indicate that, as opposed to common belief, there is no significant difference in downloaded content between this system and the classical BitTorrent ecosystem. We also found that a majority of users appear to be located in countries where anti-piracy laws have been recently adopted and enforced (France, Sweden and U.S). Finally, we evaluate the level of privacy provided by OneSwarm, and show that, although the system has strong overall privacy, a collusion attack could potentially identify content providers. This work has been published in [46].

- **Privacy leakage on mobile devices: the Mobilities Inria-CNIL project**

This joint Inria-CNIL (the French data protection agency) project aims at assessing the privacy risks associated to the use of smartphones and tablets, in particular because of personal information leakage to remote third parties. Both applications and the base OS services are considered as potential source of information leakage. More precisely, the goals are to define a platform and a methodology to identify, measure, and see the evolution over the time of privacy risks.

If similar risks exist with a PC, the situation is more worrying with mobile terminals. The reasons are:

- the intrusive feature of these terminals that their owner continuously keep with them;
- the amount of personal information available on these terminals (mobile terminals aggregate personal information but also create them, for instance with geolocalisation information);

- the facility with which the owner can personalise its terminal with new applications;
- the financial incentives that lead companies to collect and use personal information;
- the fact that the terminal user has no tool (e.g. a "privacy" firewall) to control precisely what information is exchanged with whom. The permissions provided by Android is too coarse grained to be useful, and the new privacy dashboard of IOS 6 does not enable the user to have an idea of how personal information is used by an authorized application (a one time access to a personal information and local processing within the application can be acceptable, whereas the periodic transmission of this information to remote servers is not);

The final goals of the Mobilitics project are both to study the situation and trend, but also to make mobile terminal users aware of the situation, and to provide tools that may help them to better control the personal information flow of their terminal.

6.3. Formal and legal issues of privacy

Participants: Thibaud Antignac, Denis Butin, Daniel Le Métayer.

- **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 [36]. 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 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) 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 [53]. 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.

- **Privacy and discrimination**

Actually, 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 [52]. We 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 [29].

6.4. Network measurement, modeling and understanding

Participants: Chadi Barakat, Arnaud Legout, Ashwin Rao, Walid Dabbous, Tessema Mindaye, Mohamed Ali Kaafar, Dong Wang, Vincent Roca, Ludovic Jacquin, Byungchul Park.

The main objective of our work in this domain is a better monitoring of the Internet and a better understanding of its traffic. We work on new measurement techniques that scale with the fast increase in Internet traffic and growth of its size. We propose solutions for a fast and accurate identification of Internet traffic based on packet size statistics and host profiles. Within the ANR CMON project, we work on monitoring the quality of the Internet access by end-to-end probes, and on the detection and troubleshooting of network problems by collaboration among end users.

Next, is a sketch of our main contributions in this area.

- **Checking Traffic Differentiation at the Internet Access**

In the last few years, ISPs have been reported to discriminate against specific user traffic, especially if generated by bandwidth-hungry applications. The so-called network neutrality, advocating that an ISP should treat all incoming packets equally, has been a hot topic ever since. We propose Chkdif, a novel method to detect network neutrality violations that takes a radically different approach from existing work: it aims at both application and differentiation technique agnosticism. We achieve this in three steps. Firstly, we perform measurements with the user’s real traffic instead of using specific

application traces. Secondly, we do not assume that discrimination takes place on any particular packet field, which requires us to preserve the integrity of all the traffic we intend to test. Thirdly, we detect differentiation by comparing the performance of a traffic flow against that of all other traffic flows from the same user, considered as a whole.

Chkdiff is based on the following key ideas:

Idea 1: Use real user traffic. We want to test the existence of traffic discrimination for the exact set of applications run by the end user. Hence, we only consider user-generated traffic.

Idea 2: Leave user traffic unchanged, or almost. All methods performing active measurements send probes made of real application packets and of packets that are similar, but slightly modified, so that they do not get discriminated along their path. This is quite an assumption, as we do not know exactly what ISPs do behind the scenes. In the extreme case, ISPs could even white-list traffic generated by differentiation detecting tools. It is therefore crucial to preserve as much of the original packets as possible, as well as their original per-flow order. We will see that the modifications introduced by our tool affect only the ordering of packets, their TTL value or their IP identification field.

Idea 3: Baseline is the entire traffic performance. Since we do not want to make any hypothesis in advance on what kind of mechanisms - if any - are deployed, we claim that the performance of each single non-differentiated flow should present the same behaviour as that of the rest of our traffic as a whole. Differentiated flows, on the other hand, should stand out when compared to all other flows grouped together, where a large fraction of non-differentiated flows should mitigate the impact of differentiated ones.

Chkdiff is currently the subject of a collaboration with I3S around the PhD thesis of Riccardo Ravaioli (funded by the Labex UCN@Sophia). A first description of the tool is presented in [63].

- **Lightweight Enhanced Monitoring for High-Speed Networks**

Within the collaboration with Politecnico di Bari, we worked on LEMON, a lightweight enhanced monitoring algorithm based on packet sampling. This solution targets a pre-assigned accuracy on bitrate estimates, for each monitored flow at a router interface. To this end, LEMON takes into account some basic properties of the flows, which can be easily inferred from a sampled stream, and exploits them to dynamically adapt the monitoring time-window on a per-flow basis. Its effectiveness is tested using real packet traces. Experimental results show that LEMON is able to finely tune, in real-time, the monitoring window associated to each flow and its communication overhead can be kept low enough by choosing an appropriate aggregation policy in message exporting. Moreover, compared to a classic fixed-scale monitoring approach, it is able to better satisfy the accuracy requirements of bitrate estimates. Finally, LEMON incurs a low processing overhead, which can be easily sustained by currently deployed routers, such as a CISCO 12000 device. This work is currently under submission.

- **The Complete Picture of the Twitter Social Graph**

In this work [49], we collected the entire Twitter social graph that consists of 537 million Twitter accounts connected by 23.95 billion links, and performed a preliminary analysis of the collected data. In order to collect the social graph, we implemented a distributed crawler on the PlanetLab infrastructure that collected all information in 4 months. Our preliminary analysis already revealed some interesting properties. Whereas there are 537 million Twitter accounts, only 268 million already sent at least one tweet and no more than 54 million have been recently active. In addition, 40% of the accounts are not followed by anybody and 25% do not follow anybody. Finally, we found that the Twitter policies, but also social conventions (like the followback convention) have a huge impact on the structure of the Twitter social graph.

- **Meddle: Middleboxes for Increased Transparency and Control of Mobile Traffic**

Mobile networks are the most popular, fastest growing and least understood systems in today's Internet ecosystem. Despite a large collection of privacy, policy and performance issues in mobile networks users and researchers are faced with few options to characterize and address them. In this work [62] we designed Meddle, a framework aimed at enhancing transparency in mobile networks and providing a platform that enables users (and researchers) control mobile traffic. In the mobile environment, users are forced to interact with a single operating system tied to their device, generally run closedsource apps that routinely violate user privacy, and subscribe to network providers that can (and do) transparently modify, block or otherwise interfere with network traffic. Researchers face a similar set of challenges for characterizing and experimenting with mobile systems. To characterize mobile traffic and design new protocols and services that are better tailored to the mobile environment, we would like a framework that allows us to intercept and potentially modify traffic generated by mobile devices as they move with users, regardless of the device, OS, wireless technology, or carrier. However, implementing this functionality is difficult on mobile devices because it requires warrantyvoiding techniques such as jail breaking to access and manipulate traffic at the network layer. Even when using such an approach, carriers may manipulate traffic once it leaves the mobile device, thus rendering some research impractical. Furthermore, researchers generally have no ability to deploy solutions and services such as prefetching and security filters, that should be implemented in the network. In this work, we designed Meddle, a framework that combines virtual private networks (VPNs) with middleboxes to provide an experimental platform that aligns the interests of users and researchers.

- **Mobile users' behavior modeling in Video on Demand systems and its implication on user privacy and caching strategies**

In this project, we examine mobile users' behavior and their corresponding video viewing patterns from logs extracted from the servers of a large scale VoD system. We focus on the analysis of the main discrepancies that might exist when users access the VoD system catalog from WiFi or 3G connections. We also study factors that might impact mobile users' interests and video popularity. The users' behavior exhibits strong daily and weekly patterns, with mobile users' interests being surprisingly spread across almost all categories and video lengths, independently of the connection type. However, by examining the activity of users individually, we observed a concentration of interests and peculiar access patterns, which allows to classify the users and thus better predict their behavior. We also find the skewed video popularity distribution and demonstrate that the popularity of a video can be predicted using its very early popularity level. We then analyzed the sources of video viewing and found that even if search engines are the dominant sources for a majority of videos, they represent less than 10% (resp. 20%) of the sources for the highly popular videos in 3G (resp. WiFi) network. We also report that both the type of connection and the type of mobile device used have an impact on the viewing time and the source of viewing. Using our findings, we provide insights and recommendations that can be used to design intelligent mobile VoD systems and help in improving personalized services on these platforms. This work has been published in IMC 2012 [54].

- **Explicative models for Information Spreading on the web from a user profiling perspective**

Microblog services offer a unique approach to online information sharing allowing microblog users to forward messages to others. We study the process of information diffusion in a microblog service developing Galton-Watson with Killing (GWK) model, which has many implications ranging from privacy protection to experiments validation and benchmarking. We describe an information propagation as a discrete GWK process based on Galton-Watson model which models the evolution of family names. Our model explains the interaction between the topology of the social graph and the intrinsic interest of the message. We validate our models on dataset collected from Sina Weibo and Twitter microblogs. Sina Weibo is a Chinese microblog web service which reached over 100 million users as for January 2011. Our Sina Weibo dataset contains over 261 thousand tweets which have retweets and 2 million retweets from 500 thousand users. Twitter dataset contains over 1.1 million tweets which have retweets and 3.3 million retweets from 4.3 million users. The results of the validation show that our proposed GWK model fits the information diffusion of microblog service very well in terms of the number of message receivers. We show that our model can be used in generating tweets load and also analyze the relationships between parameters of our model and popularity of the diffused information. Our work is the first to give a systemic and comprehensive analysis for the information diffusion on microblog services, to be used in tweets-like load generators while still guaranteeing popularity distribution characteristics. Our paper illustrating this study will be presented in IEEE Infocom 2013 [69].

- **Tracking ICMP black holes at an Internet Scale**

ICMP is a key protocol to exchange control and error messages over the Internet. An appropriate ICMP's processing throughout a path is therefore a key requirement both for troubleshooting operations (e.g. debugging routing problems) and for several functionalities (e.g. Path Maximum Transmission Unit Discovery, PMTUD). Unfortunately it is common to see ICMP malfunctions, thereby causing various levels of problems. In our study, we first introduce a taxonomy of the way routers process ICMP, which is of great help to understand for instance certain traceroute outputs. Secondly we introduce IBTrack, a tool that any user can use to automatically characterize ICMP issues within the Internet, without requiring any additional in-network assistance (e.g. there is no vantage point). Finally we validate our IBTrack tool with large scale experiments and we take advantage of this opportunity to provide some statistics on how ICMP is managed by Internet routers. This work has been presented in IEEE Globecom [51].

6.5. Experimental Environment for Future Internet Architecture

Participants: Walid Dabbous, Thierry Parmentelat, Frédéric Urbani, Daniel Camara, Alina Quereilhac, Shafqat Ur-Rehman, Mohamed Larabi, Thierry Turlitti, Julien Tribino.

- **SFA Federation of experimental testbeds**

We are now involved in the NOVI (E.U. STREP) project, the F-Lab (French A.N.R.) project, the FED4FIRE (E.U. IP) project and have the lead of the "Control Plane Extensions" WorkPackage of OpenLab (E.U. IP) project. Within these frameworks, as part of the co-development agreement between the Planète team and Princeton University, we have made a great deal of contributions into one of the most visible and renown implementations of the Testbed-Federation architecture known as SFA for Slice-based Federation Architecture. As a sequel of former activities we also keep a low-noise maintenance activity of the PlanetLab software, which has been running in particular on the PlanetLab global testbed since 2004, with an ad-hoc federated model in place between PlanetLab Central (hosted by Princeton University) and PlanetLab Europe (hosted at Inria) since 2007.

During 2012, we have focused on the maturation of the SFA specifications and the SfaWrap codebase, with several objectives in mind. Firstly, we have contributed within the GENI (N.S.F.) project to the specifications of the Version 3 of the AM-API (Aggregate Manager API), which defines the primitives that a testbed management infrastructure has to provide in order to be SFA-compliant.

Secondly, knowing that our former SFA implementation was targeting PlanetLab testbeds only, we needed on the one hand, to make generic this SFA implementation, by completely redesign and refactor its codebase, and on the other hand, we needed to support all the resources allocation strategies supported by the testbeds, namely the allocation of both 'shared' and 'exclusive' resources. As a result of this redesign and development effort, our new SFA implementation is now disseminated and started to be known, under the name of SfaWrap, and we believe that it can be used as a production-grade alternative to quickly add SFA compatibility on top of many heterogeneous testbed management frameworks.

Finally, in order to allow the community of networking researchers to execute cross-testbed experiments, involving heterogeneous resources, Planète team has been instrumental in federating a set of well-known testbeds through the SfaWrap, namely PlanetLab Europe, Senslab - developed in other Inria Project-teams -, FEDERICA, the outcome of another E.U.-funded project and more recently NITOS, an OMF-enabled wireless testbed. See [96] and [97] for more details.

- **Content Centric Networks Simulation**

We worked this year on the extension of the DCE framework for ns-3 in order to run CCN implementation under the ns-3 simulator. DCE stands for Direct Code Execution, its goal is to execute unmodified C/C++ binaries under ns-3 network simulator. With this tool researchers and developers can use the same code to do simulation and real experiments. DCE operation principle is to catch the standard systems calls done by the real application in the experiment and to emulate them within the ns-3 virtual network topology. Concerning CCN we use the PARC implementation named CCNx which is a well working open source software reference implementation of Content Centric Network protocol. As promised by DCE this integration of CCNx requires no modification of its code, it requires 'only' working on adding the system calls used by CCN that are not already supported by DCE. The advantage of this approach is that the integration work of CCN advanced DCE and will be useful in others completely different experiments. Another great advantage is that every evolution of the CCNx implementation is very easy to integrate, all what is needed is to compile the new source code. The next steps will be naturally to use DCE/ns-3 to evaluation CCN protocols in specific scenarios, to improve the coverage of systems calls supported by DCE, and to improve the DCE scheduler to be more realistic and to take into account CPU time spent in router queues. This work is done in the context of the ANR CONNECT project and is currently under submission.

- **ns-3 Module store**

Bake is an integration tool which is used by software developers to automate the reproducible build of a number of projects which depend on each other and which might be developed, and hosted by unrelated parties. This software is being developed with the participation of the Planète group and is intended to be the automatic building tool adopted by the ns-3 project.

The client version of Bake is already working and the Planète group had a significant participation in its development. The contributions were in the context the addition of new functionalities, bug fixing and in the development of the regression tests. We are now starting the development of the ns-3 modules repository, which is a web portal to store the meta-information of the available modules. In the present state we have already designed and implemented the portal data basis and the main interface. It is already possible to register new modules and browse among the already registered ones.

The web portal has to be finished, notably the part that will create the xml file that will be used to feed the bake's client. We also need to add new functionalities to the client part, to enable incremental build over partially deployed environments. As it is today, bake does not enable the user to add just one new module to an already deployed version of the ns-3 simulator. This work is done in the context of the ADT MobSim in collaboration with Hipercom and Swing Inria project-teams. For more details see the Bake web page <http://planete.inria.fr/software/bake/index.html>

- **The ns-3 consortium**

We have founded last year a consortium between Inria and University of Washington. The goals of this consortium are to (1) provide a point of contact between industrial members and the ns-3 project, to enable them to provide suggestions and feedback about technical aspects, (2) guarantee maintenance of ns-3's core, organize public events in relation to ns-3, such as users' day and workshops and (3) provide a public face that is not directly a part of Inria or NSF by managing the <http://www.nsnam.org> web site.

- **Automated Deployment and Customization of Routing Overlays Across Heterogeneous Experimentation Platforms**

During the last decades, many institutions and companies around the world have invested great effort into building new network experimentation platforms. These platforms range from simulators, to emulators and live testbeds, and provide very heterogeneous ways to access resources and to run experiments.

Currently, a growing concern among platform owners is how to encourage researchers from different platform communities to take advantage of the resources they offer. However, one important aspect that needs to be overcome in order to appeal researchers to use as many experimentation platforms as necessary to best validate their results, is to decrease the inherent complexity to run experiments in different platforms. Even more so, to decrease the complexity of mixing resources from different platforms on a same experiment, to achieve the combination of resources best suited to the experiment needs.

To address this concern, we developed the Network Experiment Programming Interface (NEPI) whose goal is to make easier the use of different experimentation platforms, and switch among them easily. The development of NEPI started in 2009 with the implementation of the core API, an address allocator, a routing table configurator, but also a prototype ns-3 backend driven by a simple graphical user interface based on QT. On 2010 we validated and evolved the core API with the addition of a new backend based on linux network namespace containers and stabilized the existing ns-3 backend.

During 2011, we enhanced the design of NEPI and provided experiment validation, distributed experiment control, and failure recovery functionalities. In particular, we enforced separation between experiment design and execution stages, with off-line experiment validation. We also introduced a hierarchical distributed monitoring scheme to control experiment execution. We implemented a stateless message-based communication scheme, and added failure recovery mechanisms to improve robustness. Also on 2011, we started work on a prototype PlanetLab backend.

Last year, we extended NEPI to provide automated deployment and customization of routing overlays using resources from heterogeneous experimentation platforms. The main contribution of this work is to enable researchers to easily integrate different resources, such as simulated, emulated or physical nodes, on a same experiment, using a network overlay, thus addressing one of the main concerns previously mentioned.

We started by adding support to easily build routing overlays on PlanetLab, and providing the ability to customize network traffic by adding user defined filters to packets traversing the overlay tunnels [48]. We then improved this work by adding the ability to include simulated nodes from the ns-3 backend and emulated nodes from the linux containers backend into a single overlay network. We demonstrated the use of NEPI to build adn control routing overlays which incorporate resources from different on the ns-3 2012 community workshop [74].

- **Content Centric Networks Live Experimentation**

Realistic experimentation on top of Internet-like environments is key to evaluate the feasibility of world wide deployment of CCNx, and to assess the impact of existing Internet traffic conditions on CCN traffic. However, deploying live experiments on the Internet is a difficult and error prone task, specially when performed manually.

To address this issue, during the last year, we extended NEPI, a framework for managing network experiments, to support easy design, and automated deployment and control, of CCNx experiments on the PlanetLab testbed. Among other features, NEPI now enables the deployment of user modified CCNx sources on arbitrary PlanetLab nodes, and the creation of tunnels to enable the use of multicast FIB entries between CCNx daemons over the Internet. By supporting easy CCNx experimentation on PlanetLab, NEPI can help to explore the co-existence of CCN and TCP/IP architecture.

This work was presented as a poster and a demo at CCNxCon 2012, the CCNx <http://www.ccnx.org/> community meeting [73]. The work had a very good reception and gained NEPI some new users.

An online tutorial and demo were also made available at NEPI's web page <http://nepi.inria.fr/wiki/nepi/CCNxOnPlanetLabEurope>, for dissemination purposes.

- **Smooth-transition: a new methodology for dealing with various network experiment environments**

The smooth-transition is a new methodology, which supports various network experiment environments covering from pure simulation through realistic emulation consistently. The reproducibility in experimental network research is getting important feature for iterative experiments in short-term and long-term period. The main idea of this concept is providing the reproducibility in a broader sense. So far, we had to implement different experiments by different environment, such as simulation, application-level emulation, and link-level emulation. Whereas the smooth-transition is able to keep the context of the experiments started from a pure simulation up to a realistic emulation gradually. That means the user does not need to waste time any more for learning and following a lot of documents and manuals from each different environment. Moreover, anyone can easily start to use the testbed and to develop inside (i.e. protocol stack). Because NS3 which is the most popular and powerful network simulator has been used in this concept as an experiment engine.

The smooth-transition employees Network Experiment Programming Interface (NEPI) to conduct all functions, such as composing scenario, node deployment, experiment control, and resource management. The core of building this concept is NS3 which has Emulation (EMU) and Direct Code Execution (DCE) modules. EMU supports to use real network devices instead of NS3 MAC and PHY layer implementations. DCE is able to launch real application on top of NS3 protocol stacks. Furthermore, real Linux kernel (currently, net-next 2.6 is available) can replace NS3 Internet protocols by its advanced mode. This concept needs back-end system covering all experiment nodes. Control and Management Framework (OMF) plays an important role as a software framework to control and manage an wireless network testbed, and all messages are exchanged by Extensible Messaging and Presence Protocol (XMPP). Nitos scheduler has been adopted as a reservation system <http://nitlab.inf.uth.gr/NITlab/index.php/scheduler>. The user can reserve a time slot, nodes,

and wireless channels through its web page. In addition, SFA supports that the testbed is federated with other ones of outside.

The testbed provides PCAP files as a common outcome, and this file contains captured in and out packets. However, the file size is easily over gigabytes, then it makes a very long delay to process dozens of that files. To reduce the processing time efficiently, we are using an indexing scheme for fast collecting desired packets by filtering. In particular, this scheme is very useful to find packets occurred rarely, when an detailed analysis is required for an network event, such as retransmission, intrusion detection, and node association/disassociation. The indexing information is stored in a database file, and it does not need to be modified after making the file. The size of the file is very small compared with the PCAP file, so it provides fast packet filtering permanently, even after leaving the testbed. This work, post-processing of PCAP files, is in a collaboration with Diego Dujovne and Luciano Ahumada from the Universidad Diego Portales of Chili. Especially, YoungHwan Kim, a postdoc of the Planète group, has been currently dispatched for this collaboration for fourteen weeks (September 15 2012 ~ January 26 2013) in Santiago, Chile.

- **The FIT experimental platform**

We have started, since 2011, the procedure of building a new experimental platform at Sophia-Antipolis, in the context of the FIT Equipment of Excellence project. This platform has two main goals : the first one is to enable highly controllable experiments due to its anechoic environment. These experiments can be either hybrid-experiments (as NEPI will be deployed) or federated-experiments through several testbeds. The second goal is to make resource consuming experiments (like CCNx) possible due to some powerful servers that will be installed and connected to the PlanetLab testbed. During 2012, the specifications has been defined and the procedure will continue during the next year.

- **Network Simulations on a Grid**

We studied an hybrid approach for the evaluation of networking protocols based on the ns-3 network simulator and a Grid testbed. We analyzed the performance of the approach using a simple use case. Our evaluation shows that the scalability of our approach is mainly limited by the processor speed and memory capacities of the simulation node. We showed that by exploiting the emulation capacity of ns-3, it is possible to map complex network scenarios on grid nodes. We also proposed a basic mapping algorithm to distribute a network scenario on several node [32].

ROMA Team

5. New Results

5.1. Unified model for assessing checkpointing protocols at extreme-scale

In this work [38], we defined 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 identified a set of crucial parameters, instantiated them and compared the expected efficiency of the fault tolerant protocols, for a given application/platform pair. We then proposed 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 outlined comparative behaviors of checkpoint strategies at very large scale, thereby providing insight that is hardly accessible to direct experimentation.

5.2. Impact of fault prediction on checkpointing strategies

We dealt [34] with the impact of fault prediction techniques on checkpointing strategies. We extended the classical analysis of Young and Daly in the presence of a fault prediction system, which is characterized by its recall and its precision, and which provides either exact or window-based time predictions. We succeeded in deriving the optimal value of the checkpointing period (thereby minimizing the waste of resource usage due to checkpoint overhead) in all scenarios. These results allow to analytically assess the key parameters that impact the performance of fault predictors at very large scale. In addition, the results of this analytical evaluation were nicely corroborated by a comprehensive set of simulations, thereby demonstrating the validity of the model and the accuracy of the results.

5.3. Combining process replication and checkpointing for resilience on exascale systems

Processor failures in post-petascale settings are common occurrences. The traditional fault-tolerance solution, checkpoint-rollback, severely limits parallel efficiency. One solution is to replicate application processes so that a processor failure does not necessarily imply an application failure. Process replication, combined with checkpoint-rollback, has been recently advocated by Ferreira et al. [52]. We first identified [41] an incorrect analogy made in their work between process replication and the birthday problem, and derived correct values for the Mean Number of Failures To Interruption and Mean Time To Interruption for exponentially distributed failures. We then extended these results to arbitrary failure distributions, including closed-form solutions for Weibull distributions. Finally, we evaluated process replication using both synthetic and real-world failure traces. Our main findings are: (i) replication is less beneficial than claimed by Ferreira et al.; (ii) although the choice of the checkpointing period can have a high impact on application execution in the no-replication case, with process replication this choice is no longer critical.

5.4. On the complexity of scheduling checkpoints for computational workflows

This work [22] dealt with the complexity of scheduling computational workflows in the presence of Exponential failures. When such a failure occurs, rollback and recovery is used so that the execution can resume from the last checkpointed state. The goal is to minimize the expected execution time, and we have to decide in which order to execute the tasks, and whether to checkpoint or not after the completion of each given task. We showed that this scheduling problem is strongly NP-complete, and proposed a (polynomial-time) dynamic programming algorithm for the case where the application graph is a linear chain. These results laid the theoretical foundations of the problem, and constituted a prerequisite before discussing scheduling strategies for arbitrary DAGS of moldable tasks subject to general failure distributions.

5.5. Scheduling tree-shaped task graphs to minimize memory and makespan

We [44] investigated 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. We extended 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 studied the computational complexity of this problem and provided an inapproximability result even for unit weight trees. We proposed several heuristics, each with a different optimization focus, and we analyzed them in an extensive experimental evaluation using realistic trees.

5.6. Memory allocation for different classes of DAGs

In this work, we studied the complexity of traversing workflows whose tasks require large I/O files. Such workflows arise in many scientific fields, such as image processing, genomics or geophysical simulations. They usually exhibit some regularity, and most of them can be modeled as Series-Parallel Graph. We target a classical two-level memory system, where the main memory is faster but smaller than the secondary memory. A task in the workflow can be processed if all its predecessors have been processed, and if its input and output files fit in the currently available main memory. The amount of available memory at a given time depends upon the ordering in which the tasks are executed. We focus on the problem of minimizing the amount of main memory needed to process the whole DAG.

We first concentrate on the parallel composition of task chains, or fork-join graphs. We adapt an algorithm designed for trees by Liu [54]. We prove that an optimal schedule for fork-join can be split in two optimal tree schedules, which are obtained using Liu's algorithm. We then move to Series-Parallel graphs and propose a recursive adaptation of the previous algorithm, which consists in serializing every parallel compositions, starting from the innermost, using the fork-join algorithm. Simulations show that this algorithm always reach the optimal performance, and we provide a sketch of the optimality proof. We also study compositions of complete bipartite graphs, which are another important class of DAGs arising in scientific workflows. We propose an optimal algorithm for a class of compositions which we name tower of complete bipartite graphs.

5.7. Scheduling non-linear divisible loads

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. We show [35] 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.

5.8. Energy-aware scheduling under reliability and makespan constraints

We consider [13] a task graph mapped on a set of homogeneous processors. We aim at minimizing the energy consumption while enforcing two constraints: a prescribed bound on the execution time (or makespan), and a reliability threshold. Dynamic voltage and frequency scaling (DVFS) is an approach frequently used to reduce the energy consumption of a schedule, but slowing down the execution of a task to save energy is decreasing the reliability of the execution.

In this work, to improve the reliability of a schedule while reducing the energy consumption, we allow for the re-execution of some tasks. We assess the complexity of the tri-criteria scheduling problem (makespan, reliability, energy) of deciding which task to re-execute, and at which speed each execution of a task should be done, with two different speed models: either processors can have arbitrary speeds (continuous model), or a processor can run at a finite number of different speeds and change its speed during a computation (VDD model). We propose several novel tri-criteria scheduling heuristics under the continuous speed model, and we evaluate them through a set of simulations. The two best heuristics turn out to be very efficient and complementary.

5.9. Approximation algorithms for energy, reliability and makespan optimization problems

We consider [32] the problem of scheduling an application on a parallel computational platform. The application is a particular task graph, either a linear chain of tasks, or a set of independent tasks. The platform is made of identical processors, whose speed can be dynamically modified. It is also subject to failures: if a processor is slowed down to decrease the energy consumption, it has a higher chance to fail. Therefore, the scheduling problem requires to re-execute or replicate tasks (i.e., execute twice a same task, either on the same processor, or on two distinct processors), in order to increase the reliability. It is a tri-criteria problem: the goal is to minimize the energy consumption, while enforcing a bound on the total execution time (the makespan), and a constraint on the reliability of each task.

Our main contribution is to propose approximation algorithms for these particular classes of task graphs. For linear chains, we design a fully polynomial time approximation scheme. However, we show that there exists no constant factor approximation algorithm for independent tasks, unless $P=NP$, and we are able in this case to propose an approximation algorithm with a relaxation on the makespan constraint.

5.10. Optimal algorithms and approximation algorithms for replica placement with distance constraints in tree networks

We study [16] the problem of replica placement in tree networks subject to server capacity and distance constraints. The client requests are known beforehand, while the number and location of the servers are to be determined. The Single policy enforces that all requests of a client are served by a single server in the tree, while in the Multiple policy, the requests of a given client can be processed by multiple servers, thus distributing the processing of requests over the platform. For the Single policy, we prove that all instances of the problem are NP-hard, and we propose approximation algorithms. The problem with the Multiple policy was known to be NP-hard with distance constraints, but we provide a polynomial time optimal algorithm to solve the problem in the particular case of binary trees when no request exceeds the server capacity.

5.11. Throughput optimization for pipeline workflow scheduling with setup times

We tackle [15] pipeline workflow applications that are executed on a distributed platform with setup times. In such applications, several computation stages are interconnected as a linear application graph, and each stage holds a buffer of limited size where intermediate results are stored and a processor setup time occurs when passing from one stage to another. The considered stage/processor mapping strategy is based on interval

mappings, where an interval of consecutive stages is performed by the same processor and the objective is the throughput optimization. Typical examples for this kind of applications are streaming applications such as audio and video coding or decoding, image processing using co-processing devices as FPGA. Even when neglecting setup times, the problem is NP-hard on heterogeneous platforms and we therefore restrict to homogeneous resources. We provide an optimal algorithm for constellations with identical buffer capacities. When buffer sizes are not fixed, we deal with the problem of allocating the buffers in shared memory and present a $b/(b + 1)$ -approximation algorithm.

5.12. Semi-matching algorithms for scheduling parallel tasks under resource constraints

We study [37] 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. While the problem was known to be NP-complete for bipartite graphs, but solvable in polynomial time for unweighted graphs (i.e., unit 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 on these linear algorithms, they return solutions close to the optimal (or a known lower bound) in average.

5.13. A Symmetry preserving algorithm for matrix scaling

We present an iterative algorithm which asymptotically scales the ∞ -norm of each row and each column of a matrix to one. This scaling algorithm preserves symmetry of the original matrix and shows fast linear convergence with an asymptotic rate of $1/2$. We discuss extensions of the algorithm to the one-norm, and by inference to other norms. For the 1-norm case, we show again that convergence is linear, with the rate dependent on the spectrum of the scaled matrix. We demonstrate experimentally that the scaling algorithm improves the conditioning of the matrix and that it helps direct solvers by reducing the need for pivoting. In particular, for symmetric matrices the theoretical and experimental results highlight the potential of the proposed algorithm over existing alternatives. This work resulted in an improved version [43] of an earlier technical report [55].

5.14. On shared-memory parallelization of a sparse matrix scaling algorithm

We discuss [25] efficient shared memory parallelization of sparse matrix computations whose main traits resemble to those of the sparse matrix-vector multiply operation. Such computations are difficult to parallelize because of the relatively small computational granularity characterized by small number of operations per each data access. Our main application is a sparse matrix scaling algorithm which is more memory bound than the sparse matrix vector multiplication operation. We take the application and parallelize it using the standard OpenMP programming principles. Apart from the common race condition avoiding constructs, we do not reorganize the algorithm. Rather, we identify associated performance metrics and describe models to optimize them. By using these models, we implement parallel matrix scaling algorithms for two well-known sparse matrix storage formats. Experimental results show that simple parallelization attempts which leave data/work partitioning to the runtime scheduler can suffer from the overhead of avoiding race conditions especially when the number of threads increases. The proposed algorithms perform better than these algorithms by optimizing the identified performance metrics and reducing the overhead.

5.15. Investigations on push-relabel based algorithms for the maximum transversal problem

In a technical report [42], 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 augmenting path-based algorithms and the recently proposed pseudoflow approach. 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. We finalized this work by reporting the most important results in a journal article [9].

5.16. On optimal and balanced sparse matrix partitioning problems

We investigate [20] one dimensional partitioning of sparse matrices under a given ordering of the rows/columns. The partitioning constraint is to have load balance across processors when different parts are assigned to different processors. The load is defined as the number of rows, or columns, or the nonzeros assigned to a processor. The partitioning objective is to optimize different functions, including the well-known total communication volume arising in a distributed memory implementation of parallel sparse matrix-vector multiplication operations. The difference between our problem in this work and the general sparse matrix partitioning problem is that the parts should correspond to disjoint intervals of the given order. Whereas the partitioning problem without the interval constraint corresponds to the NP-complete hypergraph partitioning problem, the restricted problem corresponds to a polynomial-time solvable variant of the hypergraph partitioning problem. We adapt an existing dynamic programming algorithm designed for graphs to solve two related partitioning problems in graphs. We then propose graph models for a given hypergraph and a partitioning objective function so that the standard cutsize definition in the graph model exactly corresponds to the hypergraph partitioning objective function. In extensive experiments, we show that our proposed algorithm is helpful in practice. It even demonstrates performance superior to the standard hypergraph partitioners when the number of parts is high.

5.17. 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 [50], [51]. The construction algorithm has a worst-case time complexity of $\Theta(mn)$ for an $n \times n$ unsymmetric matrix having m off-diagonal nonzeros. We proposed [53] another algorithm that has a worst-case time complexity of $\mathcal{O}(m \log n)$. During this reporting period, we compared the two algorithms experimentally and showed that both algorithms are efficient in general. The known algorithm [51] 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.

5.18. Introduction of shared memory parallelism in a distributed-memory sparse multifrontal solver

We study the adaptation of a parallel distributed-memory solver, MUMPS, into a shared-memory code, targetting multicore architectures. An advantage of adapting the code rather than starting with a new design is to fully benefit from its numerical kernels and functionalities. We show how one can take advantage of OpenMP directives and of existing libraries optimized for shared-memory environments, in our case BLAS libraries [48]. We have also started to study approaches that take advantage of the specificities of NUMA architectures.

5.19. Improving multifrontal methods by means of low-Rank representations

Matrices coming from elliptic PDEs have been shown to have a low-rank property. Although the dense internal datastructures involved in a multifrontal method, the so-called frontal matrices or fronts, are full-rank, their off-diagonal blocks can then be approximated by low-rank products. We have studied a low-rank format called Block Low Rank and explained how it can be used to reduce the memory footprint and complexity of both the factorization and solve phases, depending on the way variables are grouped. The proposed approach can be used either to accelerate the factorization and solution phases or to build a preconditioner [47]. We have started the development of a version of MUMPS that exploits such properties. This work is in collaboration with EDF (contract funding for the Ph.D. thesis of C. Weisbecker at INPT) and C. Ashcraft (LSTC).

5.20. Parallel computation of inverse entries of a sparse matrix

We have worked on the parallel computation of several entries [31] of the inverse of a large sparse matrix. We assume that the matrix has already been factorized by a direct method and that the factors are distributed. Entries are efficiently computed by exploiting sparsity of the right-hand sides and the solution vectors in the triangular solution phase. We demonstrate that in this setting, parallelism and computational efficiency are two contrasting objectives. We develop an efficient approach and show its efficacy by runs using the MUMPS code that implements a parallel multifrontal method.

5.21. Robust memory-aware mappings for parallel multifrontal factorization

We have studied the memory scalability of the parallel multifrontal factorization of sparse matrices. In particular, we are interested in controlling the active memory specific to the multifrontal factorization. We illustrate why commonly used mapping strategies (e.g. proportional mapping) cannot achieve a high memory efficiency. We propose a class of “memory-aware” algorithms that aim at maximizing performance under given memory constraints, and explain why they provide reliable memory estimates, thus a more robust solver. We study these issues in the context of the MUMPS solver, in which new experimental static scheduling strategies have been implemented and experimented on large matrices [46].

SARDES Project-Team

5. New Results

5.1. Languages and Foundations: Process algebra

Participants: Damien Pous, Jean-Bernard Stefani, Barbara Petit.

The goal of this work is to study process algebraic foundations for component-based distributed programming. Most of this work takes place in the context of the ANR PiCoq and Rever projects.

To develop composable abstractions for programming dependable systems, we investigate concurrent reversible models of computation, where arbitrary executions can be reversed, step by step, in a causally consistent way. This year we have continued the study of primitives for controlling reversibility in a higher-order variant of the π -calculus. We have shown that the combination of a basic notion of message alternative coupled with a rollback primitive that respects causal consistency provides enough expressive power to encode various rollback policies. We have also started to study the expressive power of these primitives with respect to transactional constructs. In particular, we have shown that our primitives allow for a faithful encoding of a notion of communicating transaction proposed by Hennessy et al, while avoiding spurious rollbacks which mar Hennessy's approach. This work has been submitted for publication. A digest of our main ideas on controlling reversibility has appeared in [25].

We have also started a study on the cost of making a concurrent programming language reversible. More specifically, we have started from an abstract machine for a fragment of the Oz programming language and made it reversible. We have shown that the overhead of the reversible machine with respect to the original one in terms of space is at most linear in the number of execution steps, and that this bound is tight since some programs cannot be made reversible without storing a commensurate amount of information. This work has been published in [26].

5.2. Control for adaptive systems: Discrete control for adaptive and reconfigurable systems

Participants: Eric Rutten, Noël De Palma, Olivier Gruber, Fabienne Boyer, Xin An, Soguy Mak-Kare Gueye.

The goal of this work is to apply control techniques based on the behavioral model of reactive automata and the algorithmic techniques of discrete controller synthesis. We adopt the synchronous approach to reactive systems, and use an associated effective controller synthesis tool, Sigali, developed at Inria Rennes. We are exploring several target application domains, where we expect to find commonalities in the control problems, and variations in the definitions of configurations, and in the criteria motivating adaptation.

This year, we have started investigating the application of discrete controller synthesis to various problems in computer systems management and administration. The increasing complexity of computer systems has led to the automation of administration functions, in the form of autonomic managers. One important aspect requiring such management is the issue of energy consumption of computing systems, in the perspective of green computing. As these managers address each a specific aspect, there is a need for using several managers to cover all the domains of administration. However, coordinating them is necessary for proper and effective global administration. Such coordination is a problem of synchronization and logical control of administration operations that can be applied by autonomous managers on the managed system at a given time in response to events observed on the state of this system. We therefore propose to investigate the use of reactive models with events and states, and discrete control techniques to solve this problem. In [20], [21], [31], [30], we illustrate this approach by integrating a controller obtained by synchronous programming, based on Discrete Controller Synthesis, in an autonomic system administration infrastructure. The role of this controller is to orchestrate the execution of reconfiguration operations of all administration policies to satisfy properties of logical consistency. We have applied this approach to coordinate energy-aware managers for self-optimization, self-regulation of processor frequency and self-repair.

5.3. System support: System support for multicore machines

Participants: Vivien Quéma, Renaud Lachaize, Baptiste Lepers.

Multicore machines with Non-Uniform Memory Accesses (NUMA) are becoming commodity platforms. Efficiently exploiting their resources remains an open research problem. This line of work investigates system support to tackle various issues related to efficient resource management and programming support.

One of the key concerns in efficiently exploiting multicore NUMA architectures is to limit as much as possible the number of remote memory accesses (i.e., main memory accesses performed from a core to a memory bank that is not directly attached to it). However, in many cases, existing profilers do not provide enough information to help programmers achieve this goal. We have developed MemProf [24], the first profiler that allows programmers to choose and implement efficient application-level optimizations for NUMA systems. MemProf achieves this goal by allowing programmers to (i) precisely understand which memory objects are accessed remotely in memory, and (ii) building temporal flows of interactions between threads and objects. We evaluated MemProf using four applications (FaceRec, Streamcluster, Psearchy, and Apache) on three different machines. In each case, we showed how MemProf helped us choose and implement efficient optimizations, unlike existing profilers. These optimizations provide significant performance gains on the studied applications (up to 161%), while requiring very lightweight modifications (10 lines of code or less).

State-machine replication is a well-known fault-tolerance technique. Unfortunately existing state-machine replication schemes do not scale well on multicore machines. In collaboration with U. Texas at Austin (L. Alvisi), we have developed a new state-machine replication scheme [23], that departs from the standard agree-execute architecture of existing schemes, in favor of a more optimistic, and less deterministic, execute-verify replication scheme, which yields much better scalability. We have evaluated Eve's throughput gain compared with traditional sequential execution approaches, as well as Eve's overheads compared to unreplicated multithreaded execution and to alternative replication approaches.

5.4. System support: Performance and dependability benchmarking

Participants: Amit Sangroya, Damian Serrano-Garcia, Sara Bouchenak [correspondant].

MapReduce is a popular programming model for distributed data processing. Extensive research has been conducted on the reliability of MapReduce, ranging from adaptive and on-demand fault-tolerance to new fault-tolerance models. However, realistic benchmarks are still missing to analyze and compare the effectiveness of these proposals. To date, most MapReduce fault-tolerance solutions have been evaluated using micro benchmarks in an ad-hoc and overly simplified setting, which may not be representative of real-world applications. To remedy this situation, we have developed MRBS, a comprehensive benchmark suite for evaluating the dependability of MapReduce systems. MRBS includes five benchmarks covering several application domains and a wide range of execution scenarios such as data-intensive vs. compute-intensive applications, or batch applications vs. online interactive applications. MRBS allows to inject various types of faults at different rates. It also considers different application workloads and data loads, and produces extensive reliability, availability and performance statistics. We have shown the use of MRBS with Hadoop clusters running on Amazon EC2, and on a private cloud [29], [28].

SOCRATE Team

6. New Results

6.1. Flexible Radio Node

6.1.1. Radio wave propagation

The MR-FDPF (Multi-Resolution Frequency Domain Partial Flow) method is proven to be a fast and efficient method to simulate radio wave propagation. It is a deterministic model which can provide an accurate radio coverage prediction. In reality, radio channels have the nature of randomness due to e.g. moving people or air flow. Thus they can not be rigorously simulated by a pure deterministic model. However, it is believed that some statistics can be extracted from deterministic models and these statistics can be very useful to describe radio channels in reality. In [20], large scale fading statistical characteristics are extracted based on the MR-FDPF method. They are validated by comparison to both the theoretical result and measurement. The match also demonstrates that MR-FDPF is capable of simulating large scale fading.

In [2] we study Realistic Prediction of Bit error rate (BER) and adaptive modulation and coding (AMC) for Indoor Wireless Transmissions. Bit error rate is an important parameter for evaluating the performance of wireless networks. In this letter, a realistic BER for indoor wireless transmissions is predicted. The prediction is based on a deterministic radio propagation model, the MR-FDPF model, which is capable of providing accurate fading statistics. The obtained BER map can be used in many cases, e.g., adaptive modulation and coding scheme or power allocation.

In [4], we propose a modification of the MR-FDPF method that allows simulating radio propagation channels in a frequency range. The performance of the proposed MR-FDPF implementation has been analyzed based on different realistic propagation scenarios. We also analyze the possibility of applying the multi-resolution frequency domain approach to the well-known transmission-line matrix method. The proposed multi-resolution frequency domain transmission-line matrix method provides a computationally efficient way of modeling radio wave propagation in three dimensional space at multiple frequencies.

In [3], we consider the performance of coded wireless communication systems experiencing non-frequency selective fading channels in shadowed environments. The quality of service (QoS) in a wireless network is dependent on the packet error outage (PEO). We address the problem of finding a tractable expression for the coded PEO over Nakagami-m channels with shadowing, considering multilevel modulations, various block, convolutional channel coding schemes and hard decision decoding. In order to obtain the coded PEO, an inversion of the coded packet error probability (PEP) w.r.t. the signal to noise ratio (SNR) is needed. To this end, we propose an invertible approximation for the coded PEP w.r.t. the uncoded bit error probability (BEP) in Nakagami-m fading channels which is accurate for all BEPs of interest. The BEP itself depends on the average SNR and we hence make use of previous results on the inversion of the uncoded BEP w.r.t. the SNR in Nakagami-m fading channels, holding for M-PSK and M-QAM signals. We were thus able to obtain a reliable closed form expression for the coded PEO in flat fading and shadowing channels

6.1.2. Power consumption

In [24], we propose the use of an existing opensource network simulator, WSNet, to evaluate the interest of using multi-mode relays in terms of energy consumption. We show that the combination of MIMO and multi-mode provides a solution to reduce global energy consumption, but that conclusions are really scenario-dependent. Moreover, we explain how a multi-mode MIMO terminal can improve these results using adaptive strategies.

the energy consumption in wireless sensor networks is studied. In order to minimize the consumed power at the analog and RF part, an energy recovering system combined with a wake-up radio is proposed for discussion. The proposed architecture has three activity levels : zero consumption, low and high energy consumption. In order to quantify the gain in terms of power consumption, a power consumption model state of the art is proposed. in [7] all radio channel models which can be used for MIMO heterogeneous network with small cells are described.

6.1.3. MIMO

In [28], we study MIMO and next generation system. For the past decade or more MIMO systems have been the subject of very intensive research. However in the past few years, these techniques have begun to be implemented in practice. In particular they have appeared in the standards for next generation systems such as LTE, 3GPP-LTE Advanced and WiMAX, as well as the latest versions of Wifi. This chapter, extracted from the book edited by the Cost Action 2100: “Pervasive Mobile and Ambient Wireless Communications”, brings together the MIMO systems used in next generation systems with other work on the implementation and simulation of these systems. It also describes advances in MIMO techniques in a number of areas. The first section is divided into two sub-sections dealing first with simulators and testbeds which are used in system-level simulators to evaluate overall system capacity, as discussed in later chapters of this book. Secondly the development of terminals for next generation MIMO systems is considered, especially considering the additional RF hardware required for MIMO. Section 7.2 then discusses especially precoding techniques used in many of the recent standards to implement MIMO. In particular precoding allows the implementation of closed loop or adaptive MIMO. In next generation systems there is also much increased attention on MU-MIMO and on multi-terminal MIMO in general, including so-called “network MIMO ” approaches, which appear in LTE as Coordinated Multiple Point: this is covered in Sect. 7.3. Various advanced MIMO transmission and detection approaches are covered in Sects. 7.4 to 7.6, including some interesting work on MIMO techniques involving continuous phase modulation, giving advantages in terms of peak-to-average power ratio.

6.2. Agile Radio Resource Sharing

6.2.1. Wireless Multi-hop Networks

In [6], we study energy-delay tradeoff in wireless multihop networks with unreliable links. Energy efficiency and transmission delay are very important parameters for wireless multihop networks. Numerous works that study energy efficiency and delay are based on the assumption of reliable links. However, the unreliability of channels is inevitable in wireless multihop networks. In addition, most of works focus on self-organization protocol design while keeping non-protocol system parameters fixed. While, very few works reveal the relationship between the network performance and these physical parameters, in other words, the best networks performance could be obtained by the physical parameters. This paper investigates the tradeoff between the energy consumption and the latency of communications in a wireless multihop network using a realistic unreliable link model. It provides a closed-form expression of the lower bound of the energy–delay tradeoff and of energy efficiency for different channel models (additive white Gaussian noise, Rayleigh fast fading and Rayleigh block-fading) in a linear network. These analytical results are also verified in 2-dimensional Poisson networks using simulations. The closed-form expression provides a framework to evaluate the energy–delay performance and to optimize the parameters in physical layer, MAC layer and routing layer from the viewpoint of cross-layer design during the planning phase of a network.

6.2.2. Relay and Cooperative Communications

In [16], we aim at characterizing the gain induced by using relay channels in a linear network under both capacity constraint and realistic energy model. We express a general model based on a convex optimization problem. Then, we use numerical tools to obtain results on the outer and inner bounds of the capacity of the full and half duplex relay channel. We then extend this study with more complex networks based on relay channels, especially networks formed by a linear chain of nodes. We describe the Pareto optimal solutions of the minimization problem with respect to the consumed energy and latency in such a linear network. From the simple case of the linear multi-hop network, we study the gains when implementing a linear chain of relay channels and compare these results to the simpler multi-hop transmission.

In [15], we present preliminary results on achievable rates in half-duplex cooperative multiple access channels (CMAC). We show that the upper bound on the capacity of the half-duplex CMAC can be solved using convex optimization techniques. Under a Gaussian model, we study the maximal achievable rate by every node in the network. We propose a number of scenarios, encompassing existing and theoretical cooperation schemes. Using these hypotheses, we evaluate the performance of both a non-cooperative concurrent access and simple cooperative multi-hop or relaying schemes with respect to the upper bound. The performance is compared for the various scenarios, and we provide analyses of specific cases in order to illustrate how our framework may be used to answer targeted questions about the capacity of CMACs.

In [31], we aim at obtaining usable bounds on the performance of CMAC under a Gaussian model. We first show that the problem can be transformed into a convex optimization problem which makes it easily solvable using numerical tools. We propose, as a line of study, to consider the maximal achievable common rate by every node in the network. We then proceed to express closed-form bounds on the capacity region of the CMAC in that common rate scenario. We study simple cooperation schemes based on existing results in relay channels and compare them to other medium sharing approaches. In the end, we show that using the relay-channel based protocols can be efficient for some parameters, but gets less interesting in the Gaussian case if the source-destination links are good enough.

In [30], we study the optimal power allocations in CMACs, where we aim at maximizing the rate achievable by both sources simultaneously rather than the sum of achievable rates. Separating our study between the coherent and non-coherent case, we obtain closed-form expressions for the optimal power allocations w.r.t. the outer bounds of the capacity region, as well as decode-and-forward and non-cooperative inner bounds. We point out during our resolution that the general CMAC model behaves as a multiple access relay channel (MARC), where a "virtual" relay node is introduced to represent the cooperation between the sources. This equivalent model simplifies the original power allocation problem. We finally show that the general cut-set outer bound on the capacity region of the equivalent MARC matches exactly the tightest known outer bound on the capacity region of the original CMAC.

In [17] we address the distributed power adaptation problem on the downlink for wireless cellular networks. As a consequence of uncoordinated local scheduling decisions in classical networks, the base stations produce mutual uncontrolled interference on their co-channel users. This interference is of a variable nature, and is hardly predictable, which leads to suboptimal scheduling and power control decisions. While some works propose to introduce cooperation between base stations, in this work we propose instead to introduce a model of power variations, called trajectories in the powers space, to help each base station to predict the variations of other base stations powers. The trajectories are then updated using a Model Predictive Control (MPC) to adapt transmit powers according to a trade-off between inertia (to being predictable) and adaptation to fit with capacity needs. A Kalman filter is used for the interference prediction. In addition, the channel gains are also predicted, in order to anticipate channel fading states. This scheme can be seen as a dynamic distributed uncoordinated power control for multichannel transmission that fits the concept of self-optimised and self-organised wireless networks. By using the finite horizon MPC, the transmit powers are smoothly adapted to progressively leave the current trajectory toward the optimal trajectory. We formulate the optimisation problem as the minimisation of the utility function of the difference between the target powers and MPC predicted power values. The presented simulation results show that in dynamic channel conditions, the benefit of our approach is the reduction of the interference fluctuations, and as a consequence a more accurate interference prediction, which can further lead to a more efficient distributed scheduling, as well as the reduction of the overall power consumption.

6.2.3. BAN

In [26] we present a simple Body Area Network (BAN) platform that was built to monitor the performance of a marathon athlete all along the race, meeting real-time and QoS constraints, under good transmission conditions. Data collected during the event (packet loss, signal strength) allowed us to obtain a primary knowledge about the behavior of the radio transmissions between the different links in the network. The results of this experiment and their important disparities observed between the links point out the need to improve the transmission strategy.

6.2.4. Network coding

One of the most powerful ways to achieve transmission reliability over wireless links is to employ efficient coding techniques. In [10] investigates the performance of a transmission over a relay channel where information is protected by two layers of coding. In the first layer, transmission reliability is ensured by fountain coding at the source. The second layer incorporates network coding at the relay node. Thus, fountain coded packets are re-encoded at the relay in order to increase packet diversity and reduce energy consumption. Performance of the transmission is measured by the total number of transmissions needed until the message is successfully decoded at the destination. We show through both analytical derivations and simulations that adding network coding capabilities at the relay optimizes system resource consumption. When the source uses a random linear fountain code, the proposed two layer encoding becomes more powerful as it reduces the transmission rate over the direct link between the source and the destination.

In [27] we study the deployment of fountain codes and network coding in a wireless sensor network (WSN). A WSN is composed of sensor nodes with restricted capacities: memory, energy and computational power. The nodes are usually randomly scattered across the monitored area and the environment may vary. In the presence of fading, outage and node failures, fountain codes are a promising solution to guaranty reliability and improve transmission robustness. The benefits of fountain codes are explored based on an event-driven WSNet simulator considering realistic implementation based on standard IEEE802.15.4. Fountain codes are rateless and capable of adapting their rate to the channel on the fly using a limited feedback channel. In this thesis, we highlight the benefits brought by fountain code in terms of energy consumption and transmission delay. In addition to the traditional transmission with fountain code, we propose in this thesis to study the network coding transmission scheme where nodes are allowed to process the information before forwarding it to their neighbors. By this means, we can say that packet diversity is exploited as each individual packet is unique and contains different representations of binary data. Redundancy is thus optimized since repetitions are avoided and replaced with diversified information. This can further lead to an overall improved performance in cooperative communication where nodes are allowed to assist in relaying packets from the source the destination. We highlight in this thesis the benefits of fountain code combined to network coding and show that it leads to a reduction in transmission delay and energy consumption. The latter is vital to the life duration of any wireless sensor network.

In [9] we tackle the problem of providing end to end reliable transmissions in a randomly deployed wireless sensor network. To this aim, we investigate the simultaneous use of gradient broadcast routing (for its inherent adaptability to any network topology and its changes), fountain codes (for their universal property) and intra-flow network coding (to introduce packet diversity in redundant copies). We present the impact of the proposed strategy on a realistic network. This work permits to highlight that, compared to basic gradient broadcast routing, the strategy not only improves the reliability and the delay in the network but also clearly increases its lifetime.

6.2.5. Vehicular networks

In [22] we study a hybrid propagation model For large-scale variations caused By vehicular traffic in small cells. we present a propagation model generating time series of large-scale power variations for small-cell radio links intersected by vehicular traffic. The model combines stochastic processing and geometric computation. For each road crossing a link, a two-state process parameterized by mobility statistics represents the obstruction status. When the status is set to obstructed, a fluctuation pattern is generated. Based on previously published measurements, both mobility statistics and time series results are validated through the comparison of respectively inter-obstruction duration distributions and outage probabilities. The proposed model avoids resource consuming iterative propagation prediction while providing realistic and frequency adaptive results.

In [21], we performed measurements of large-scale variations caused by vehicular traffic in small-cell. This paper presents and characterizes large-scale variations of received power generated by vehicular traffic crossing a radio link. Measurements in the 2 GHz band for several small-cell configurations involve various transmitter heights, link distances and urban densities. Observations showed that stronger losses up to 30 dB

are due to medium to high vehicles. Lower vehicles have a smaller impact in links perpendicular to traffic, but amplitude variations and duration can reach larger values when the receiver is at cell radius limits.

6.2.6. security

In [18] we study Security Embedding on ultra wideband impulse radio(UWB-IR) Physical Layer. The main goal of this work is to incorporate security in an existing ultra wideband (UWB) network. We present an embedding method where a tag is added at the physical layer and superimposed to the UWB-impulse radio signal. The tag should be added in a transparent way so that guaranteeing compatibility with existing receivers ignoring the presence of the tag. We discuss technical details of the new embedding method. In addition, we discuss embedding strength and we analyze robustness performance. We demonstrate that the proposed embedding technique meets all the system design constraints.

In [11] we study Jamming in time-hopping ultrawide band (TH-UWB) Radio. With the great expansion of wireless communications, jamming becomes a real threat. We propose a new model to evaluate the robustness of a communication system to jamming. The model results in more scenarios to be considered ranging from the favorable case to the worst case. The model is applied to a TH-UWB radio. The performance of such a radio in presence of the different jamming scenarios is analyzed. We introduce a mitigation solution based on stream cipher that restricts the jamming problem of the TH-UWB communication to the more favorable case while preserving confidentiality.

6.2.7. Network Information Theory

Fundamental performance limits of multi-hop wireless transmissions are being investigated in [33] from a multiobjective perspective where transmission decisions (i.e. relay selection, scheduling or routing decision) modify the trade-off between capacity, reliability, end-to-end delay or network-wide energy consumption. In our previous work presented in the Inria research report RR-7799, Pareto-optimal performance bounds and network parameters have been derived for a 1-relay and 2-relay network within a MultiObjective(MO) performance evaluation framework. We show in this report that these bounds are tight since they can be reached by simple practical coding strategies performed by the source and the relays. Such strategies constitute achievable lower MO performance bounds on the real MO performance limits. More precisely, we adopt a coding strategy where the source transmits a random linear fountain code which is coupled to a network coding strategy performed by the relays. Two different network coding strategies are investigated. Practical performance bounds for both strategies are compared to the theoretical bound. We show that the theoretical bound is tight: generational distance between the practical and theoretical bound for the best strategy is only of 0.0042

In [19] we revisit the problem of non-cooperativ association of mobiles to access points using game theory. We consider in this paper games related to the association problem of mobiles to an access point. It consists of deciding to which access point to connect. We consider the choice between two access points or more, where the access decisions may depend on the number of mobiles connected to each one of the access points. We obtain new results using elementary tools in congestion and crowding games.

In [23] we study stochastic analysis of energy savings with sleep mode in Orthogonal Frequency-Division Multiple Access (OFDMA) wireless networks. The issue of energy efficiency in OFDMA wireless networks is discussed in this paper. Our interest is focused on the promising concept of base station sleep mode, introduced recently as a key feature in order to dramatically reduce network energy consumption. The proposed technical approach fully exploits the properties of stochastic geometry, where the number of active cells is reduced in a way that the outage probability, or equivalently the signal to interference plus noise distribution, remains the same. The optimal energy efficiency gains are then specified with the help of a simplified but yet realistic base station power consumption model. Furthermore, the authors extend their initial work by studying a non-singular path loss model in order to verify the validity of the analysis and finally, the impact on the achieved user capacity is investigated. In this context, the significant contribution of this paper is the evaluation of the theoretically optimal energy savings of sleep mode, with respect to the decisive role that the base station power profile plays.

6.3. Software Radio Programming Model

6.3.1. *Virtual Radio Machine*

In [14] we present a survey of existing prototypes dedicated to software defined radio. We propose a classification related to the architectural organization of the prototypes and provide some conclusions about the most promising architectures. This study should be useful for cognitive radio testbed designers who have to choose between many possible computing platforms. We also introduce a new cognitive radio testbed currently under construction and explain how this study have influenced the test-bed designers choices.

6.3.2. *Embedded systems*

In [13], we explore new area/throughput trade-offs for the Girault, Poupard and Stern authentication protocol (GPS). This authentication protocol was selected in the NESSIE competition and is even part of the standard ISO/IEC 9798. The originality of our work comes from the fact that we exploit a fixed key to increase the throughput. It leads us to implement GPS using the Chapman constant multiplier. This parallel implementation is 40 times faster but 10 times bigger than the reference serial one. We propose to serialize this multiplier to reduce its area at the cost of lower throughput. Our hybrid Chapman's multiplier is 8 times faster but only twice bigger than the reference. Results presented here allow designers to adapt the performance of GPS authentication to their hardware resources. The complete GPS prover side is also integrated in the network stack of the POW-WOW sensor which contains an Actel IGLOO AGL250 FPGA as a proof of concept.

The people involved in this axes also published in the computer science field. For instance in [1] static vulnerability detection in java service-oriented components is studied. In [12] A lightweight Hash function family based on FCSRs is studied.

URBANET Team

6. New Results

6.1. Scalable protocols for capillary networks.

Participants: Ibrahim Amadou, Quentin Lampin, Bilel Romdhani, Alexandre Mouradian, Isabelle Augé-Blum, Fabrice Valois

6.1.1. Beacon-less and opportunistic routing.

During the thesis of Ibrahim Amadou [1], we were focused on the issues of energy in WSNs through energy-efficient routing and medium access control protocols. The contributions of research work can be summarized as follows. First, we were interested in the energy issues at the routing layer for multi-hop wireless sensor networks (WSNs). We proposed a mathematical framework to model and analyze the energy consumption of routing protocols in multi-hop WSNs by taking into account the protocol parameters, the traffic pattern and the network characteristics defined by the medium channel properties, the dynamic topology behavior, the network diameter and the node density. We showed that Beacon-less routing protocol is a good candidate for energy saving in WSNs.

We investigated the performance of some existing relay selection schemes which are used by Beacon-less routing protocols. Extensive simulations were realized in order to evaluate their performance locally in terms of packet delivery ratio, duplicated packet and delay. Then, we extended the work in multi-hop wireless networks and developed an optimal solution, Enhanced Nearest Forwarding within Radius, which tries to minimize the per-hop expected number of retransmissions in order to save energy.

We presented a new Beacon-less routing protocol called Pizza-Forwarding (PF) without any assumption on the radio environment: neither the radio range nor symmetric radio links nor radio properties (shadowing, etc.) are assumed or restricted. A classical greedy mode is proposed. To overcome the hole problem, packets are forwarded to an optimal node in the two hop neighbor following a reactive and optimized neighborhood discovery.

In order to save energy due to idle listening and overhearing, we proposed to combine PF's main concepts with an energy-efficient MAC protocol to provide a joint MAC/routing protocol suitable for a real radio environment. Performance results lead to conclude to the powerful behavior of PF-MAC.

In collaboration with Orange Labs, we designed QOR, an opportunistic routing protocol for wireless sensor networks [16]. QOR first builds a stable directed acyclic logical routing structure and a prefix-based addressing plan stemming from data sinks. This addressing plan is then used to define the potential forwarders set for each source and allows a strict scheduling and a unique selection of the forwarder for each transmission thanks to a cascading acknowledgment scheme. QOR is particularly suited for sensor networks that require high delivery ratio under severe energy constraints. Extensive simulations show the benefits of QOR over an implementation of the IETF routing protocol for Lossy and Low Power networks, RPL, tailored to provide high delivery ratios. Our case studies show that QOR saves up to 50% energy and reduces the end-to-end delay of a factor of 4 times while maintaining similar delivery ratios.

Most existing routing protocols designed for WSNs assume that links are symmetric, which is in contradiction with what is observed in the field. Indeed, many links in real-world WSNs are asymmetric. Asymmetric links can dramatically decrease the performance of routing algorithms not designed to cope with them. Quite naturally, most existing routing protocol implementations prune the asymmetric links to only use the symmetric ones. In our experience, asymmetric links are a valuable asset to improve network connectivity, capacity and overall performance [20],[2]. We therefore introduced AsymRP (Asymmetric Convergecast Routing Protocol) [21], a new routing protocol for collecting data in WSNs. AsymRP assumes 2-hop neighborhood knowledge and uses implicit and explicit acknowledgment. It takes advantage of asymmetric links to increase delivery ratio while lowering hop count and packet replication.

6.1.2. MAC and cross-layer mechanisms for QoS.

Protocols developed during the last years for Wireless Sensor Networks (WSNs) are mainly focused on energy-consumption optimization and autonomous mechanisms (e.g. self-organization, self-configuration, etc). Nevertheless, with new WSN applications appear new QoS requirements such as time constraints. Real-time applications require the packets to be delivered before a known time bound which depends on the application requirements. We particularly focused on applications which consist in alarms that are sent to the sink node (e.g. air pollution monitoring). We proposed the Real-Time X-layer Protocol (RTXP) [27], a real-time communication protocol that integrates mechanisms for both MAC and routing layers. Our proposal aims at guaranteeing an end-to-end constraint delay, while keeping good performances on other parameters, such as energy consumption. For this purpose the protocol relies on a hop-count-based Virtual Coordinate System (VCS) which classifies nodes having the same hop-count from the sink, allows forwarder selection, and gives to the nodes an unique identifier in a 2-hop neighborhood allowing deterministic medium access. This protocol has better performances than state-of-the-art protocols, in terms of time constraints and reliability, even with unreliable radio links.

In the ARESA2 project, but also in a joint collaboration with Orange Labs, we studied receiver initiated MAC protocol to compare their performance to the more classical receiver-based MAC one [17]. We proposed the Self Adapting Receiver Initiated MAC protocol (SARI-MAC), a novel asynchronous MAC protocol for energy constrained Wireless Sensor Networks. SARI-MAC self-adapts to the traffic load to meet specified Quality of Service requirements at the lowest energy cost possible. To do so, SARI-MAC relies on traffic estimation, duty-cycle adaptation and acknowledgment mechanisms. Our performance evaluation assesses that SARI-MAC meets given QoS requirements in a energy efficient manner and outperforms the state of the art protocol RI-MAC in a broad range of traffic scenarios.

For energy constrained wireless sensor networks, lifetime is a critical issue. Several medium access control protocols have been proposed to address this issue, often at the cost of poor network capacity. To address both capacity and energy issues, we proposed a novel medium sharing protocol for Wireless Sensor Networks named Cascading Tournament (CT-MAC) [15]. CT-MAC is a synchronous, localized, dynamic, joint contention/allocation protocol. Relying on cascading iterations of tournaments, CT-MAC allocates multiple time slots to nodes that compete for accessing the medium. CT-MAC offers an unprecedented trade-off between traffic delay, network capacity and energy efficiency and stands out as a solid candidate for energy constrained sensor networks that must support heterogeneous traffic loads. Our simulations show that CT-MAC significantly outperforms the state-of-the-art SCP- MAC protocol.

6.2. Characterizing urban capillary wireless networks.

Participants: Sandesh Uppoor, Diala Naboulsi, Rodrigue Domga Komguem, Anis Ouni, Alexandre Mouradian, Isabelle Augé-Blum, Hervé Rivano, Marco Fiore, Fabrice Valois

6.2.1. Properties of urban road traffic of interest to mobile networking.

The management of mobility is commonly regarded as one of the most critical issues in large-scale telecommunication networks. The problem is exacerbated when considering vehicular mobility, which is characterized by road-constrained movements, high speeds, sudden changes of movement direction and acceleration, and significant variations of these dynamics over daytime. The understanding of the properties of car movement patterns becomes then paramount to the design and evaluation of network solutions aimed at vehicular environments.

We first analyzed how the vehicular mobility in a large-scale urban region affects a cellular infrastructure intended to support on-board users. We studied the spatial and temporal distribution of traffic load induced by vehicular users, their spatial flows, their inter-arrival and residence times at cells [22].

We then studied the topological features of a network built on moving vehicles, considering the instantaneous connectivity of the system [28]. Our results evidence the spatial and temporal diversity of road traffic, stressing the importance of a correct modeling of road traffic towards the reliable performance evaluation of network

protocols. Additionally, the results outline how commonly adopted assumptions (e.g., Poisson user arrivals at the network base stations) do not hold under vehicular environments, and how the V2V-based network has low connectivity, availability, reliability and navigability properties.

6.2.2. *The limits of RSSI-based localization.*

Numerous localization protocols in Wireless Sensor Networks are based on Received Signal Strength Indicator. Because absolute positioning is not always available, localization based on RSSI is popular. More, no extra hardware is needed unlike solutions based on infra-red or ultrasonic. Moreover, the theory gives a RSSI as a function of distance. However, using RSSI as a distance metric involves errors in the measured values, resulting path-loss, fading, and shadowing effects. We did experimentation results from three large WSNs, each with up to 250 nodes [23]. Based on our findings from the 3 systems, the relation between RSSI and distance is investigated according to the topology properties and the radio environment. We underline the intrinsic limitations of RSSI as a distance metric, in terms of accuracy and stability. Contrary to what we assumed, collaborative localization protocol based on Spring-Relaxation algorithm can not smooth the distance-estimation errors obtained with RSSI measurements.

6.2.3. *Modeling and optimization of wireless networks.*

In critical real-time applications, when an event is detected, the Worst Case Traversal Time (WCTT) of the message must be bounded. However, despite this, real-time protocols for WSNs are rarely formally verified. The model checking of WSNs is a challenging problem for several reasons. First, WSNs are usually large scale so it induces state space explosion during the verification. Moreover, wireless communications produce a local broadcast behavior which means that a packet is received only by nodes which are in the radio range of the sender. Finally, the radio link is probabilistic. The modeling of those aspects of the wireless link in model checking is not straightforward and it has to be done in a way that mitigates the state space explosion problem. We are currently working on proposing a methodology adapted to WSNs, and based on Timed Automata (TA) and model-checking. First results are promising [19], but needed to be further investigated.

While the large variety of routing protocols (geographical, gradient, reactive, ...) proposed in the literature provide a set of pertinent solutions for optimizing the energy consumption for multi-hop wireless networks, they do not permit to know the conditions of use of these protocols based on parameters such as: the dynamics of topology, traffic pattern, the density and diameter of the network, the load, etc. In [12], we presented a theoretical model for evaluating the energy consumption for communication protocols taking into account both the dynamics of nodes and links, the properties of topology, the traffic pattern, the control/data packets and a realistic channel model. This model is applied successively to several protocols (GPSR, AODV, OLSR and PF) to highlight their optimum usage and it permits to conclude that Beacon-Less routing protocols are adapted for application with low traffic.

We continued developing optimization tools for building optimal solution to various problems of multi-hop wireless networks. Most of these contributions combine graph theoretical basis with Mixed Integer Linear Programming techniques, and are valuable for understanding the extremal behaviors of the systems and guide the development of efficient architectures and protocols. In this sense, we have considered a new edge coloring problem to model call scheduling optimization issues in wireless mesh networks: the proportional coloring [6]. It consists in finding a minimum cost edge coloring of a graph which preserves the proportion given by the weights associated to each of its edges. We show that deciding if a weighted graph admits a proportional coloring is pseudo-polynomial while determining its proportional chromatic index is NP-hard. We then give lower and upper bounds for this parameter that can be computed in pseudo-polynomial time. We finally identify a class of graphs and a class of weighted graphs for which the proportional chromatic index can be exactly determined.

Dealing with wireless mesh network, we have investigated the fundamental trade-off between transmitting energy consumption and network capacity [24]. The results on this trade-off have been computed using MILP models solved with column generation techniques. The main contribution relies in the ability to consider a realistic SINR model of the physical layer with a continuous power control and discrete transmission rate

selection at each node. In order to model these functionalities, a strong formulation (in the sense that the linear relaxation gives relevant lower bounds) of the rate selection is introduced.

The behavior of beaconless geographic forwarding protocols for wireless sensor networks has also been modeled [9]. A realistic physical layer is taken into account by combining MILP models with simulation based inputs on the number of required retransmissions for realizing a transmission. The model is then able to compute energy efficient routings and allows for understanding the most efficient relay selection schemes, denoted Furthest Forward within Reliable neighbors (FFRe).

6.3. Solutions for cellular networks.

Participants: Anis Ouni, Fabrice Valois, Hervé Rivano, Marco Fiore

6.3.1. Content downloading through a vehicular network.

We considered a system that leverages vehicle-to-infrastructure (V2I) and vehicle-to-vehicle (V2V) communication to transfer large contents to users on-board moving cars. This paradigm is intended to relieve the cellular infrastructure from the high load that such downloads would induce, once vehicles are widely equipped with infotainment devices.

We first characterized the theoretical performance limits of such a vehicular content downloading system by modeling the downloading process as an optimization problem, and maximizing the overall system throughput. Our approach allows us to investigate the impact of different factors, such as the roadside infrastructure deployment, the vehicle-to-vehicle relaying, and the penetration rate of the communication technology, even in presence of large instances of the problem [7]. We then evaluated practical protocols for vehicular downloading, devising solutions for the selection of relay vehicles and data chunks at the Road Side Units (RSUs), and evaluating them in real-world road topologies, under different infrastructure deployment strategies [8].

Our results show that V2V transfers can significantly increase the download rate of vehicular users in urban/suburban environments, and that such a result holds throughout diverse mobility scenarios, RSU placements and network loads. Also, they highlight the existence of two operational regimes at different penetration rates and the importance of an efficient, yet 2-hop constrained, V2V relaying.

6.3.2. Toward green mesh and cellular networks.

On the one hand, a promising technique for minimizing the transmission power of cellular networks seems to be a dramatic densification of micro-cells coverage. On the other hand, increasing the number of micro-cells multiplies the energy consumed by the cells whatever their state, idle, transmitting or receiving. For a sustainable deployment of such micro-cell infrastructures and for a significant decrease of the overall energy consumption, an operator needs to be able to switch off cells when there are not absolutely needed. The densification of the cells induces the need for an autonomic control of the on/off state of cells. This has motivated a preliminary investigation on exploiting within the micro-cellular settings the manifold results of duty cycles for Wireless Sensor Networks where switching nodes on and off is done in a distributed or localized manner while coverage and connectivity properties are maintained [29].

Focusing on broadband wireless mesh networks based on OFDMA resource management, and considering a realistic SINR model of the physical layer with a continuous power control and discrete transmission rate selection at each node, we have investigated the trade-off between transmission energy consumption and network capacity [24]. Correlation between capacity and energy consumption is analyzed as well as the impact of physical layer parameters - SINR threshold and path-loss exponent. We highlight that there is no significant tradeoff between capacity and energy when the power consumption of idle nodes is important. We also show that both energy consumption and network capacity are very sensitive to the SINR threshold variation. We also highlight that power control and rate selection are not expandable to an optimal system configuration.

6.4. Miscellaneous security issues in capillary networks.

Participants: Ochirkhand Erdene-Ochir, Fabrice Valois, Marco Fiore

6.4.1. Resiliency in routing protocols.

Within the ARESA2 project, we defined the notion of resiliency for routing protocols in wireless sensor networks and we applied it to several routing strategies to provide an understandable taxonomy [3]. Efforts have been made to compare routing protocols according to their resiliency in wireless multi-hop sensor networks in the presence of packet dropping malicious insiders. In [13], we proposed a new taxonomy of routing protocols obtained by applying our resiliency metric. Several resiliency enhancing methods such as introducing a random behavior to the classical routing protocols and a new data replication method based on the distance information have been evaluated as well. Simulation results demonstrate the effectiveness of the proposed approach.

6.4.2. Verifying the positions announced by mobile nodes.

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.

We proposed a fully-distributed cooperative solution that is robust against independent and colluding 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 [5].

A centralized solution was also developed, that 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 [18].

E-MOTION Project-Team

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. Problem addressed and background

The general framework based on the new concept of continuous symmetry developed during the last two years (see [67] for a detailed description of this framework) has been extensively applied to investigate the visual inertial structure from motion problem. This problem was already considered in 2011. During 2012 more general results have been found. Special attention has been devoted to identify the conditions under which the problem has a finite number of solutions. Specifically, it has been shown that the problem can have a unique solution, two distinct solutions and infinite solutions depending on the trajectory, on the number of point-features and on their layout and on the number of camera images. The investigation has also performed in the case when the inertial data are biased, showing that, in this latter case, more images and more restrictive conditions on the trajectory are required for the problem resolvability.

5.1.1.2. Theoretical results

The new results have been published on the journal of Transaction on Robotics [68], in a technical report [43] and submitted to the International Journal of Computer Vision. We have also considered the case of structured light. Specifically, we have considered a sensor assembling (from now on *aerial vehicle*) consisting of a monocular camera and inertial sensors. Additionally, a laser pointer is mounted on the aerial vehicle and it produces a laser spot. The laser spot is observed by the monocular camera and it is the unique point feature used in the proposed approach. We focus our attention to the case when the aerial vehicle moves in proximity of a planar surface and in particular when the laser spot belongs to this surface. We introduced two novel contributions. The former is the analytical derivation of all the observable modes, i.e., all the physical quantities that can be determined by only using the inertial data and the camera observations of the laser spot during a short time-interval. This derivation was based on the framework introduced in [67]. Specifically, it is shown that the observable modes are: the distance of the vehicle from the planar surface; the component of the vehicle speed, which is orthogonal to the planar surface; the relative orientation of the vehicle with respect to the planar surface; the orientation of the planar surface with respect to the gravity. The second contribution is the introduction of a simple recursive method to perform the estimation of all the aforementioned observable modes. This method is based on a local decomposition of the original system, which separates the observable modes from the rest of the system. The method has been validated by using synthetic data. Additionally, preliminary tests with real data are provided and more complete experiments are in progress. The presented approach can be integrated in the framework of autonomous take-off and landing, safe touch-down and low altitude manoeuvres even in dark or featureless environment. These results have been published in the iros conference [31]

5.1.1.3. Experimental results

In parallel to this theoretical activity an experimental activity has been carried out in order to deploy our technologies to industrial partners. To this regard, we had a collaboration with the company Delta Drone in Grenoble. In this framework we introduced a new method to localize a micro aerial vehicle (MAV) in GPS denied environments and without the usage of any known pattern. 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. We have implemented this method on the platform available in our lab. This is a *Pelican* from *Ascending Technologies* equipped with an Intel Atom processor board (1.6 GHz, 1 GB RAM) (figure 1).



Figure 1. AscTec Pelican quadcopter equipped with a monocular camera.

Our sensor suite consists of an Inertial Measurement Unit (3-Axis Gyro, 3-Axis Accelerometer) belonging to the Flight Control Unit (FCU) “AscTec Autopilot”, and a monocular camera (*Matrix Vision mvBlueFOX*, $FOV: 130\text{ deg}$). The camera is calibrated using the Camera Calibration Toolbox for Matlab by J.Y. Bouguet at caltech. The calibration between IMU and camera has been performed using the Inertial Measurement Unit and Camera Calibration Toolbox in [66]. The IMU provides measurements update at a rate of $100Hz$, while the camera framerate is $10Hz$. The Low Level Processor (LLP) of our Pelican is flashed with the *2012 LLP Firmware* and performs attitude data fusion and attitude control. We flashed the High Level Processor (HLP) with the *asctec_hl_firmware* [48]. The onboard computer runs linux 10.04 and ROS (Robot Operating System). We implemented our method using ROS as a middleware for communication and monitoring. The HLP communicates with the onboard computer through a FCU-ROS node. The communication between the camera and the onboard computer is achieved by a ROS node as well. The presented algorithms are running online and onboard at $10Hz$.

The scenario setup is shown in figure 3. Since our lab is not yet equipped with a Motion Capture System, we used an ARToolKit Marker with the only aim of having a ground truth to evaluate the performance of our approach. The estimation of the camera pose provided by the marker is not used to perform the estimation. The marker is positioned such that its reference frame is coincident with the configuration shown in figure 3. The three features considered are the center of the three little balls in figure 3. The use of three blob markers instead of natural features is only related to the need to get a ground truth. The information related to the pattern composed by the 3 features ($D = 0.25m$, $\gamma_1 = 60deg$, $\gamma_2 = 120deg$) is only used to evaluate the performance of our approach. The algorithm does not require any information about the features configuration.

Figure 4 and 5 show respectively the position and the attitude by using the proposed approach. The estimated values are compared with the ground truth obtained with the ARToolkit marker. From figure 4 we see that the difference between our estimates and the ground truth values is of the order of $2cm$ for x and y and less than $0.5cm$ for z . From figure 5 we see that the difference between our estimates and the ground truth values is of the order of $2deg$ for *Pitch* and less than $0.5deg$ for *Roll* and *Yaw*.

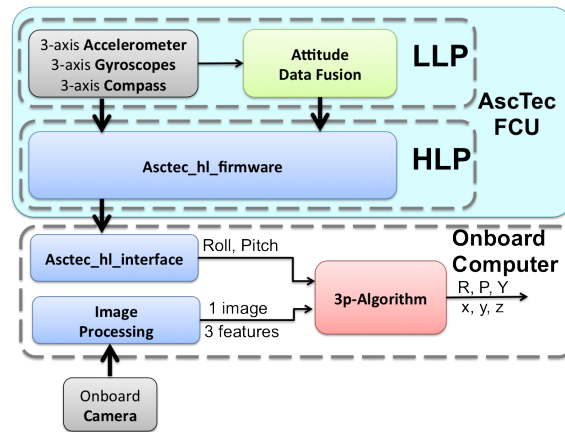


Figure 2. Our Pelican quadcopter: a system overview

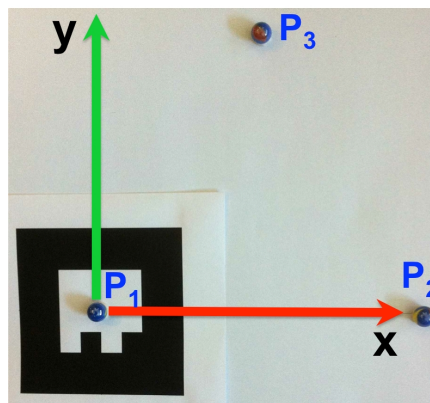


Figure 3. Scenario: The AR Marker and the 3 balls are used only with the aim to get a rough ground truth. The AR Marker provides the camera 6DOF pose in a global reference frame according to our conventions.

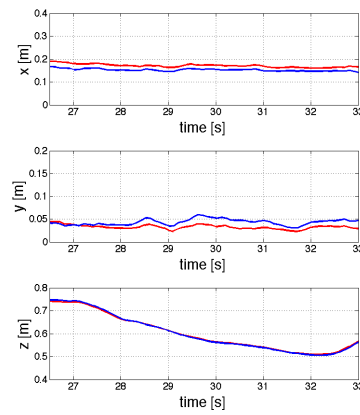


Figure 4. Estimated position, respectively x , y , z . The red lines represent the estimated values with the 3p-Algorithm, the blue ones represent a rough ground truth (from ARToolkit Marker).

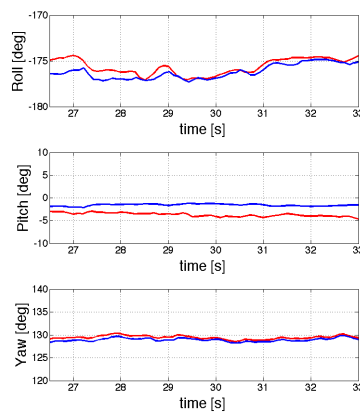


Figure 5. Estimated attitude, respectively Roll, Pitch, Yaw. The red lines represent the estimated values with the 3p-Algorithm, the blue ones represent a rough ground truth (from ARToolkit Marker).

We believe that the main source of error is due to the distortion of the lens, which is not fully compensated by the calibration. Note that this distortion also affects our ground truth. We plan to test our approach in an environment equipped with a Motion Capture System.

This method is currently under evaluation to be patented.

5.1.2. Visual recognition for intelligent vehicles

Participants: Alexandros Makris, Mathias Perrollaz, Christian Laugier.

We developed a generic object class recognition method. The state-of-the-art visual object class recognition systems operate with local descriptors and codebook representation of the objects. Various local features (e.g., gradient maps, edges) are used to create the descriptors. Then kernel based classifiers are commonly employed to classify the detected features in one of several object classes [50] [54]. The recognition of vehicles or pedestrians from sensors mounted on a moving platform is achieved by different approaches using various types of sensors, e.g., stereo camera, laser [61] [52]. The approaches that perform data fusion from various sensors have proven to be the more robust in a variety of road conditions [76].

Our work focuses on the development of an object class recognition system which follows the part based detection approach [65]. The system fuses intensity and depth information in a probabilistic framework. To train the system for a specific object class, a database of annotated with bounding boxes images of the class objects is required. Therefore, extending the system to recognize different object classes is straightforward. We apply our method to the problem of detecting vehicles by means of on-board sensors. Initially, depth information is used to find regions of interest. Additionally, the depth of each local feature is used to weight its contribution to the posterior of the object position in the corresponding scale. The votes are then accumulated in a 3d space-scale space and the possible detections are the local maxima in that space.

The novelty of our approach is the fusion of depth and intensity information to form a probabilistic part-based detector. Using depth information is beneficial for the robustness of the approach, because we avoid including many noisy detections resulting from false matches between features of different scales. The method is tested with stereo video sequences captured in an urban environment. Figure 6 shows some example detections. The proposed method detects cars in various scales, in cases with partial occlusions, and under significant background clutter.



Figure 6. Car detection examples. The new weighting strategy allows to better detect the partially occluded objects.

In 2012, we worked on two particular improvements of the method. First, we modified the weighting strategy in order to increase the detection of partially occluded objects. This approach effectively improves the detection results. Second, we consider replacing the current depth descriptor, which only integrates depth information, with a more advanced depth descriptor (e.g., the NARF descriptor). This work is still in progress, in collaboration with Dimitrios Kanoulas, PhD student in Northeastern University (USA).

In 2012, the full method for objects recognition has been submitted for publication in IEEE Transactions on Intelligent Transportation Systems.

5.1.3. Bayesian Motion Detection in Dynamic Environments

Participants: Qadeer Baig, Jander Perrollaz, Mathias Botelho, Christian Laugier.

5.1.3.1. Introduction

Bayesian Occupancy Filter (BOF) [51] is a grid based perception framework that we use for environment monitoring. In this representation this framework estimates the probability of occupancy as well as velocity of each cell of this grid using sensor data. Output of this framework is used by Fast Clustering Tracking Algorithm (FCTA) [69] to cluster objects and to track them. An important point is that BOF estimates cell velocities without motion information of the ego vehicle, so these are relative velocities. Since no motion information are used, the static objects observed from the moving ego vehicle are also tracked, this results into many false moving objects. Although many of these false positives can be removed by tuning parameters of FCTA, however, this usually is a time consuming task. We note that the number of false can be reduced as well as dependence on FCTA parameters can be relaxed if we can separate the input to BOF into static and dynamic parts. Adding these motion information with cells will allow BOF to calculate velocity information for moving cells only and FCTA will also ignore the static cells while clustering step resulting into faster calculations and better track. In this context we have developed a very fast motion detection technique to separate BOF input into static and dynamic parts. The integration of this module with BOF and FCTA has helped us remove about 78% of the false positives. This technique is summarized next.

5.1.3.2. Fast Motion Detection

In this section we summarize the technique that we have developed to find moving parts of the environment. This motion detection module is situated in the processing chain just before the BOF. The input to this module consists of an occupancy grid generated by the fusion module. And the output of this module is used by both BOF and FCTA modules.

The objective of this module is to separate the input occupancy grid into two parts: cells belonging to static objects and cells belonging to moving objects. The main idea of this separation between static and dynamic parts, consists of keeping a track of how many times a cell is observed as free and how many times it is observed as moving. However to realize this concept we must solve the localization problem. We solve this problem using velocity and rotation information given by MTi-G XSens unit. This allows us to map cells between two input grids OG_{t-1} and OG_t at time $t - 1$ and t as shown in figure 7 .

We use two sets of *Free* and *Occupied* counter arrays. One set is initialized from new input grid at time t whereas other set keeps updated counts until time $t - 1$. Then after above transformation between cells of grids OG_{t-1} and OG_t newly initialized set of arrays is updated from arrays at time $t - 1$, resulting in incremented counts for overlapping areas between two grids. Finally following decision function is used to separate cells of current input grid OG_t into static and dynamic parts and results are stored in a motion grid.

$$MotionGrid_t[i] = \begin{cases} 1, & OG_t[i] > 0.5 \text{ and} \\ & FreeCount_t[i] > 2 * OccupiedCount_t[i] \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

This technique being simple is quite robust and efficient and does not oblige us to solve the complete SLAM problem. This work is published as [19] and [20].

5.1.3.3. Integration within the BOF framework

We have updated the BOF implementation to take into account the motion detection results. The motion grid is used as an input for updating the BOF. If the input motion grid tells that a cell belongs to a static object, then during prediction and update cycles of BOF the cell's velocity distribution over the velocity range is set to uniform for all discrete velocity values. This essentially means that no velocity information for a given cell

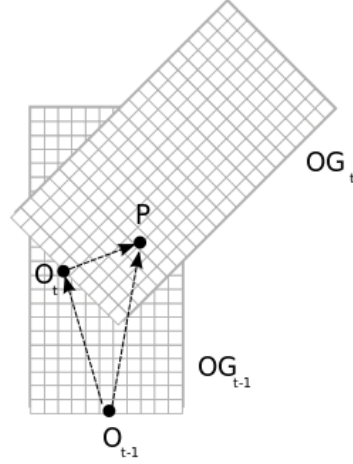


Figure 7. Position of the grid at time instants $t - 1$ and t . Vehicle undergoes a motion of $u_t = (\nu_t, \omega_t)$ to move from O_{t-1} to O_t . We need to find the position of point P of grid OG_{t-1} in grid OG_t .

is available and the cell is labeled as static in the current BOF implementation. However, if the cell has been detected as belonging to a moving object, then the velocity distribution prediction and the update cycle are carried out normally. In formal terms this change in the parametric form of dynamic model can be stated as:

$$P(A_i^t | A_i^{t-1}) = \begin{cases} (1 - \epsilon)P(A_{A_i^{t-1}}^{t-1}) + \epsilon/\|A_i\| & \text{if } MotionGrid_t[i] > 0 \\ 1/\|A_i\| & \text{otherwise} \end{cases}$$

where A_i^t is the set of antecedents of cell i at time t and ϵ is a parameter of BOF, modelling the prediction error probability.

5.1.3.4. Integration with FCTA

We have also updated the FCTA implementation to take into account the motion detection results. The cells which do not possess the velocity information are now ignored during the clustering step. While generally most of the areas belong to static objects and are detected as static by the motion detection module, two main advantages are expected from this strategy: (i) the clustering stage of the algorithm is highly accelerated by the reduction of hypotheses, and (ii) the false moving clusters are ignored because they are not considered for clustering, even with the relaxed FCTA parameters.

5.1.3.5. Results

Some qualitative results of motion detection module are shown in figure 8, (rectangles around the objects are drawn manually to highlight them). As expected, the moving objects are properly detected. For example, figure 8 (left) shows the motion detection scenario of two cars, and the car moving around a roundabout has been successfully detected in figure 8 (right). Some noise is also visible on the results, mainly due to two causes: first, the uncertainty on the IMU measurements along with the circular motion model may result in some errors in the estimation of the motion; second, the decision function is too rough for taking correct decisions in every situation. The results would benefit from replacing this function by a probabilistic model.

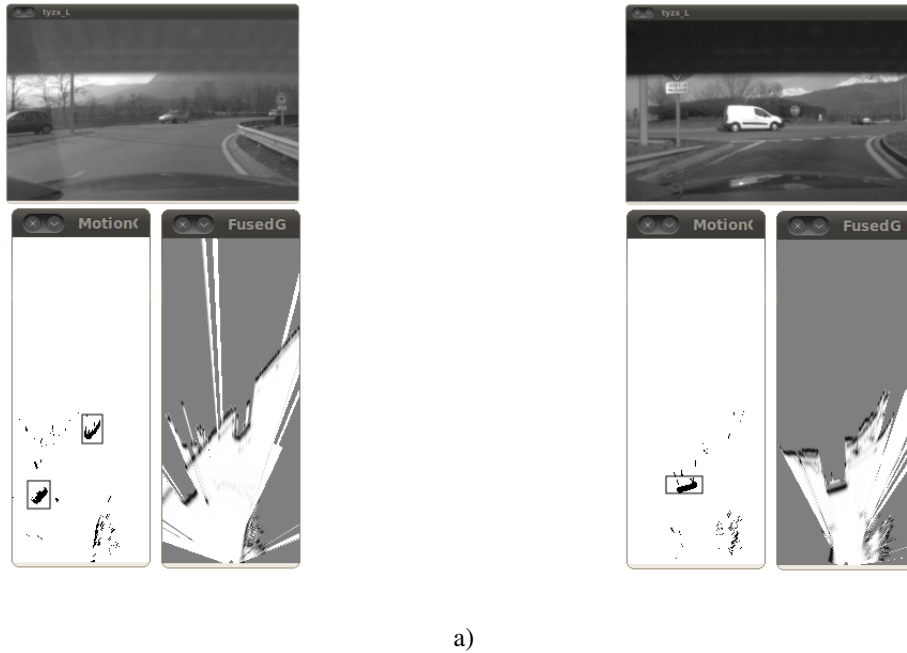


Figure 8. Left: Motion detection results of two cars. Top, scenario, bottom right input fused grid, bottom left resulting motion grid. Right: Motion detection results of a car on a roundabout. Some noise due to sensor uncertainty is also visible

The tracking results of FCTA are highly sensitive to its parameters values. There are less false positives when strict parameters (large thresholds) are used, however, a large number of the true tracks may be missed, resulting in numerous miss detections -note that since the focus of this work is to detect moving objects, we consider in this part that detections belonging to the static environment are false alarms-. The relaxed parameters (small thresholds) provide less miss detections, however, a large number of false tracks are detected. While finding the appropriate set of parameters can be a challenging task, our implementation of the motion detection module with relaxed parameters represents a trade-off.

The following statistics with a dataset duration of about 13 minutes give an insight into the improvements gained with this implementation. When the motion detection module is not used, 22303 tracks are detected. The activation of the motion detection module with all other parameters being equal provides to detect 4796 tracks. This example shows the advantage of the motion detection module because it allows us to remove most of the false tracks while leaving most of the true tracks. Some qualitative FCTA tracking results with and without motion detection module activated (with all other parameters being same) are shown in figures 9 and 10 . Red rectangles are the detected tracks by FCTA in the shown scenario. We clearly see that most of the false positives have been removed.

5.1.3.6. Conclusion

In this section we have presented a fast technique to find moving objects from laser data and its integration with Bayesian Occupancy Filter (BOF) and Fast Clustering-Tracking Algorithm (FCTA). We have seen that after this integration we were able to remove a significant number of false alarms, this has also relaxed the dependence of results on the FCTA parameters.

We plan to change the rather ad hoc decision module that is currently based on occupied and free counter values to a more formal probabilistic function that also takes into account the uncertainty effects on the neighboring cells to accommodate the localization errors. We are also working on extending the tracking module from single motion mode to multiple motion modes.



Figure 9. Tracking results of a car. Left, FCTA results without motion detection module activated. Right, same scenario but with motion detection module activated.

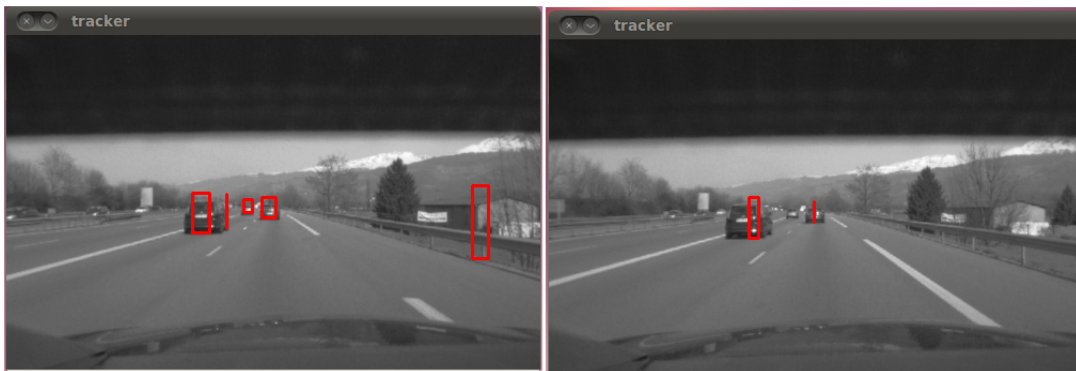


Figure 10. Tracking results of two cars on highway. Left, FCTA results without motion detection module activated. Right, same scenario but with motion detection module activated.

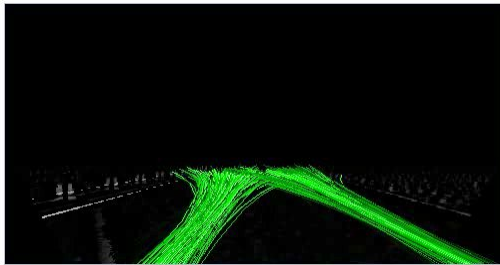
5.1.4. Vision-based Lane Tracker

Participants: Mathias Perrollaz, Amaury Nègre.

For perception in road structured environment the detection of the lane markers and its localization provide an interesting information to predict drivers behaviors and to evaluate collision risks. We currently develop a real time road lane detection and tracking application using camera's image information. The tracking application estimates simultaneously the road plane orientation, the lane curvature and the camera position by using a Monte-Carlo particle filter. With this method, the parameter distribution is represented by a set of particles (see Fig 11 .a) that are sequentially updated using the vehicle dynamic model, evaluated by a ridge extraction (Fig 11 .b) and sampled considering the evaluation result. The average of the particles, displayed on Fig 11 .c) provides a good estimation of the lane state.

To obtain real-time performance, we implemented the whole process on GPU using the nVidia Cuda toolkit.

The output of this application has been mainly used to predict lane change behaviour 5.2.1 and to risk estimation applications.



(a) Particles cloud



(b) Ridges extraction



(c) Lane state estimation

Figure 11. Visual Particle based lane tracking. (a) The Lane state is estimated by a particles set which is recursively updated, evaluated and resampled. (b) A ridge image is compute to estimate each particle. (c) The average of the particle state provides a good estimation of the lane.

5.1.5. Experimental platform for road perception

Participants: Nicolas Vignard, Mathias Perrollaz, Amaury Nègre.

5.1.5.1. Experimental platform material description

Our experimental platform is a Lexus LS600h car shown in Figure 12. The vehicle is equipped with a variety of sensors including two IBEO Lux lidars placed toward the edges of the front bumper, a TYZX stereo camera situated behind the windshield, and an Xsens MTi-G inertial sensor with GPS.



Figure 12. Lexus LS600h car equipped with two IBEO Lux lidars, a TYZX stereo camera, and a n Xsens MTi-G inertial sensor with GPS.

The stereo camera baseline is 22 cm, with a field of view of 62° . Camera resolution is 512x320 pixels with a focal length of 410 pixels. Each lidar provides four layers of up to 200 impacts with a sampling period of 20 ms. The angular range is 100° , and the angular resolution is 0.5° . The on-board computer is equipped with 8GB of RAM, an Intel Xeon 3.4 GHz processor and an NVIDIA GeForce GTX 480 for GPU. IMU data contains accelerations, velocity, GPS position and steering angle. The experiments are conducted 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). The datasets are acquired online and are used for testing of our sensor fusion and risk assessment algorithms.

5.1.5.2. Migration from Hugn to ROS middleware

Our platform described in 5.1.5.1 previously used a middleware named Hugn. Middlewares bring an abstraction layer between the sensors drivers and the processing modules. We also used this middleware to share information with modules and applications. Using a middleware facilitates and normalises the communication between modules.

Hugn has been developed by inria for the Cycab project and a team was built to add functionalities and maintain this new middleware. However, now the team has to work on other projects and it is becoming increasingly difficult to allocate resources to maintain this middleware. Given this and some other technical issues [49], we have decided to change our robotic middleware.

We find that many different middleware (AROCAM, RTMaps, ROS, ...) are being used in the robotic community [53]. Among these, Robotic Operating System (ROS) is increasingly becoming a research standard in robotics. The reason being: an important community, a lot of tools and sharing work and development. The primary goal of ROS is to develop faster robotics applications. However, before moving to ROS we also did an extensive research on the comparison between Hugn and ROS [49], that supported our this decision.

Because of this middleware change, we had to reimplement all the perception process from drivers to applications. In this regard, we have implemented the following drivers:

- the IBEO Lux lidar
- the TYZX camera
- the CAN bus
- the Xsens MTi-G (inertial sensor with GPS)

However for the Xsens MTi-G, we found an existing driver that we modified to add the GPS functionality http://www.ros.org/wiki/lse_xsens_mti.

Furthermore, we have also migrated the following modules:

- a module that fuses lidar data into an occupancy grid
- a module that generates occupancy grid from the stereo camera
- the Bayesian Occupancy filter (BOF) module
- the lane tracker

Some result images of occupancy grids and data from the lane tracker after this migration to ROS are shown below 13 . Finally, we have created a public repository at <http://gforge.inria.fr> that share our developments (both drivers and modules).

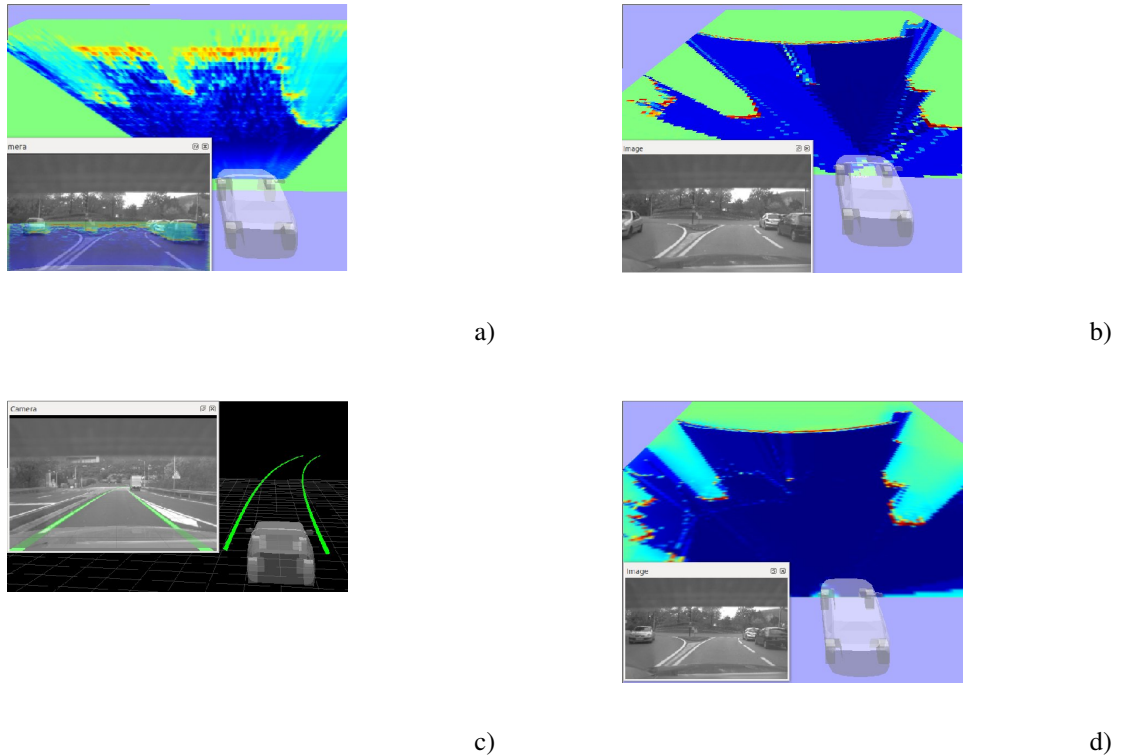


Figure 13. a) occupancy grid from the stereo camera. b) occupancy grid from the lidar. d) lanes detected by lane tracker. e) occupancy grid from the BOF.

5.1.5.3. Disparity space approach for a vision based occupancy grid

Participants: Mathias Perrollaz, Anne Spalanzani, John-David Yoder, Amaury Nègre, Christian Laugier.

To use sensors in the BOF framework, it is essential to develop an associated probabilistic sensor model that takes into consideration the uncertainty over measurements. In 2009, we proposed such a sensor model for stereo-vision [72]. The originality of the approach relied on the decision to work in the disparity space, instead of the classical Cartesian space. In 2010, we improved our sensor model, in order to mimic some features of the sensor models used for range finders. Particularly, we worked on managing visible/occluded areas of the scene [74], and on including the information from the road/obstacle segmentation of the disparity image [73]. Our approach was also designed to allow highly parallel computation of the occupancy grid. A. Nègre implemented the approach on GPU using NVIDIA CUDA to enhance the performance. The complete

processing of the stereo data can now be done in 6 ms, while more than 150 ms were necessary with the CPU implementation. The complete approach for occupancy grid computation using stereovision has been published in 2012, in [13].

5.1.6. Software and Hardware Integration for Embedded Bayesian Perception

Participants: Mathias Perrollaz, Christian Laugier, Qadeer Baig, Dizan Vasquez.

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).

This work has been started in 2012 in cooperation with CEA-LETI DACLE laboratory. During 2012, we have worked on the definition of the software/hardware architecture and we have started to re-think some components of the lower layer of the BOF software module.

The work plan has been split in two three-year-long phases, respectively leading to address a first level of integration based on mobile technologies, and a second level of integration, based on a more dedicated hardware architecture (and maybe to a SOC).

Two cooperative projects have been prepared and submitted this year for supporting this promising research: the “Permobil” project (FUI), involving industrial companies and user, and the “Perfect” project (IRT-Nano) involving the CEA-LET LIALP lab and ST-Microelectronics. Permobil is focusing on the first integration objectives (3 years) and has been recently submitted. 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. . .).

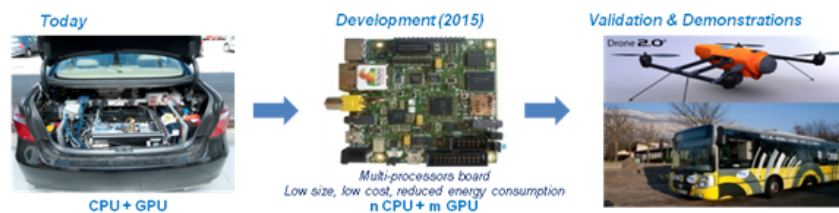


Figure 14. First objective for software/hardware of the BOF: developing and using multiple processor board from mobile technologies. The approach will be validated with real demonstrators.

5.2. Dynamic Change Prediction and Situation Awareness

5.2.1. Vision-based Lane Change Prediction

Participants: Puneet Kumar, Mathias Perrollaz, Stephanie Lefevre, Amaury Nègre, Maiwen Gault.

Predicting driver’s behaviors is a key component for future Advanced Driver Assistance Systems (ADAS). In 2012, we have proposed a novel approach for lane change prediction, using only information from a vision sensor embedded into the car. The idea is to predict in advance if our vehicle is about to change lane. Then this information can be used to properly help the driver, for instance by detecting inconsistencies with the turn lights signals.

As an input, the method uses visual data from a camera embedded into the car. A multiple-size ridge filter is used to extract low level features from the image (white markings on black road). Then road lanes are estimated and tracked over time using a particle filter. This process allows parallel computing, and thus works in real time on GPU.



Figure 15. Vision-based tracking of the road markings. From left to right: particles generated by the particle filter, low level features extracted using the ridge filter, and estimated lane.

The road markings are used to estimate the position and heading angle of our car with respect to the lane, as well as the derivatives of these variables. This information is then used as a vector of features for a classifier. The used classifier is a multi-class Support Vector Machine (SVM). The three possible classes are "no lane change" (NL), "right lane change" (CR) and "left lane change" (CL). The classifier has been trained using real data of 180 lane changes on highway, manually annotated. The output of the classification is then converted into a set of probabilities using a generalized Bradley-Terry model.

The classifier provides a very short term classification, which can contain many errors. The longer term integration of the time information is obtained by feeding the classification results into a Bayesian Filter (BF). The posterior output of the filter provides the probability distribution over possible behaviors (NL, CR, CL), hence providing the lane change prediction.

Real-world data from our vehicle is used for the purpose of training and testing. Data from different drivers on different highways were used for the robustness evaluation of the overall approach. The proposed method show promising results, because it is able to predict driver's intention to change lane 1.3 seconds (average) in advance, with maximum prediction horizon of 3.29 seconds. We are now working on a real time implementation of this approach, to demonstrate its use on real situations (e.g., for warning the driver while driving on the highway).

5.2.2. Risk estimation at road intersections for connected vehicle safety applications

Participants: Stéphanie Lefèvre, Christian Laugier.

Intersections are the most complex and dangerous areas of the road network. Statistics show that most road intersection accidents are caused by driver error and that many of them could be avoided through the use of Advanced Driver Assistance Systems. In this respect, vehicular communications are a particularly promising technology. The sharing of information between vehicles over wireless links allows vehicles to perceive their environment beyond the field-of-view of their on-board sensors. Thanks to this enlarged representation of the environment in time and space, situation assessment is improved and dangerous situations can be detected earlier.

A PhD was started on this topic in 2009, in collaboration with Renault. It tackles the problem of risk estimation at road intersections from a new perspective: a Bayesian framework is proposed for reasoning about traffic situations and collision risk at a semantic level instead of at a trajectory level. While classic approaches estimate the risk of a situation by predicting the future trajectories of the vehicles and looking for intersections between them, here dangerous situations are detected by estimating the intentions of drivers and looking for conflicts between them. This novel approach to risk assessment is very relevant in the context of road traffic, as it takes into account the fact that the road network is a highly constrained environment regulated by traffic

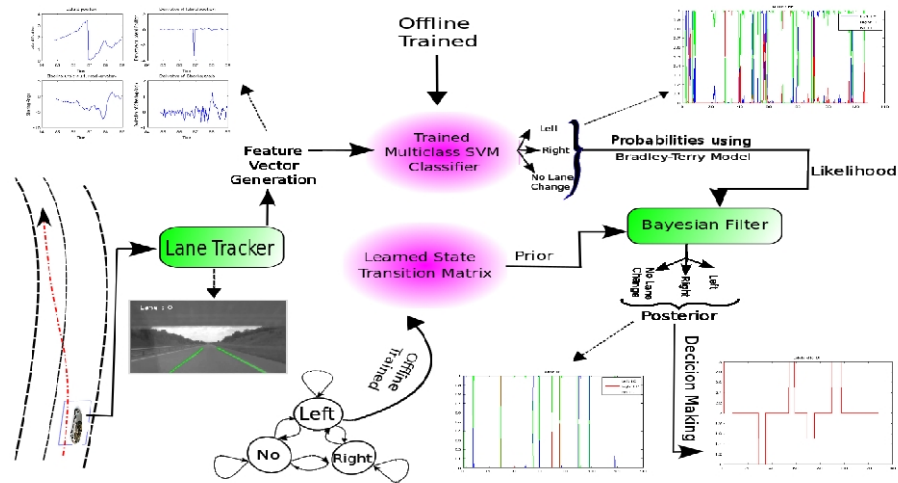


Figure 16. General architecture of the lane change prediction module.

rules. The proposed approach relies on the estimation of drivers' intentions, and the main difficulty lies in the presence of uncertainties in the estimation process: uncertainties inherent to sensor data, and ambiguities when linking vehicle behavior with driver intention. In this work the information about the state of other vehicles is obtained via vehicle-to-vehicle communication, but the proposed framework for reasoning on traffic situations and risk is general and can be applied with other types of sensors, e.g., the ones presented in 5.1.3 .

The focus of the first year (2010) was on estimating a driver's intended maneuver at an intersection (go straight, turn left, etc.) based on the current state of the vehicle (position, orientation, turn signal state) and on contextual information extracted from the digital map. The idea was to use the information on the geometry of the road network and on the connectivity between lanes to build a statistical model of the relationship between a vehicle's state and the driver's intended maneuver. The proposed solution is based on a Bayesian Network and on geometric functions which automatically extract the characteristics of the intersection from a digital map. This approach was designed and implemented during a 3-month internship in the Stanford Artificial Intelligence Laboratory, in collaboration with Sebastian Thrun's Driving Group.

During the second year (2011) we augmented the Bayesian Network with a filtering process so that new measurements could be recursively used to estimate the driver's intentions. The new version of the motion model explicitly models the influence of traffic rules on the behavior of a vehicle. While state-of-the-art approaches usually assume independence between vehicles, the proposed motion model takes into account the mutual influences between the maneuvers performed by the vehicles in the scene. These improvements were carried out by introducing two new variables in the Bayesian Network. The *"Intention to stop"* corresponds to the driver's intention to come to a halt at the intersection. The *"Expectation to stop"* corresponds to whether or not the traffic rules expect the driver to come to a halt at the intersection. The former is assumed to be dependent on the previous intention of the driver and on the current expectation. The latter is assumed to be dependent on the rules applying at the intersection and on the previous situational context, i.e., the state of the other vehicles in the scene. With this model it is possible to infer what a driver intends to do and what a driver is expected to do from the successive measurements of the pose, speed, and turn signals of the vehicles in the scene. Risk can then be computed based on the probability that intention and expectation do not match.

The focus of this year (2012) was on the evaluation of the performance of the algorithm. The proposed approach was validated in field trials using passenger vehicles equipped with vehicle-to-vehicle wireless communication modems, and in simulation. Our simulations assumed ideal perception and communication, and considered typical accident scenarios at a two-way-stop cross intersection. The tested maneuvers included crossing maneuvers, merging maneuvers, and left turn across path maneuvers (see Figure 17). A total of 240 instances of these scenarios were simulated, with both priority violations and stop violations as accident causes. The same number of instances were simulated for non-dangerous situations, by enforcing a 3 seconds safety distance between the vehicles at all times. An analysis of the collision prediction horizon led to the following conclusions:

1. There were no false alarms in non-dangerous situations, and no missed detection in the dangerous scenarios.
2. For merging and crossing maneuvers, the proposed algorithm was able to predict collisions at least 1.5 s before they occurred.
3. For left turn across path maneuvers, the proposed algorithm was able to predict collisions at least 0.6 s before they occurred.
4. Accidents caused by stop violations were detected on average 1 s earlier than the ones caused by priority violations.

Different accident avoidance strategies were tested: warning the driver of the vehicle with right-of-way, warning the driver of the other vehicle, applying autonomous braking on the vehicle with right-of-way, and applying autonomous braking on the other vehicle. It was found that the ability of each strategy to avoid an accident varies a lot with the situation. For example, the “autonomous braking on the vehicle with right-of-way” can avoid the accident in 91% of cases for stop violations, but only in 34% of cases for priority violations. “Warning the driver of the vehicle with right-of-way” can avoid the accident in 1% of cases for priority violations, while for the same scenarios “autonomous braking on the other vehicle” can avoid the accident in 99% of cases. These results were published at the conference IEEE IROS’12 [22], and as an Inria Research Report [41]. Field trials were conducted using two vehicles equipped with off-the-shelf vehicle-to-vehicle wireless communication modems. Six different drivers took part in the experiments to recreate realistic dangerous and non-dangerous situations at a T-shaped give-way intersection (see Figure 17). The risk estimation algorithm was run online in one of the vehicles, and triggered a warning for the driver when it detected a dangerous situation (see Figure 18). In the 120 tests, there were no false alarms and no missed detections. The warning was always triggered early enough that accidents were avoided by performing an emergency braking. The field trials proved that the proposed approach can operate with success in real-life situations and trigger warnings in real time. They also showed the robustness of the algorithm, since the experiments were carried out with several drivers, a positioning system with a precision of 2 meters (standard deviation) and challenging wireless communication conditions. These results were published at the conference IEEE IV’12 [23], where the paper received the Best PhD Student Paper award.

The PhD was successfully defended in October 2012 [9]. A patent application was filed with Renault in October 2012 [45]. This work will be continued within the Inria@SiliconValley program, in collaboration with the University of Berkeley, California. Ms Lefevre will conduct further research on this topic as a post-doctoral researcher at Berkeley starting January 2013.

5.2.3. *Guidance for Uncertain shooting domain*

Participant: Emmanuel Mazer.

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

Contex This project relates to the use of lock after launch missiles, both long range anti aircraft missiles such as Meteor, or air to ground strike weapons employing for example IIR or Semi Active Laser (SAL) guidance. In both cases, a target is ultimately recognized and tracked by means of a seeker which detects a characteristic signal above the noise. This could be the target reflections of a radar beam, or the spot from a designating laser.

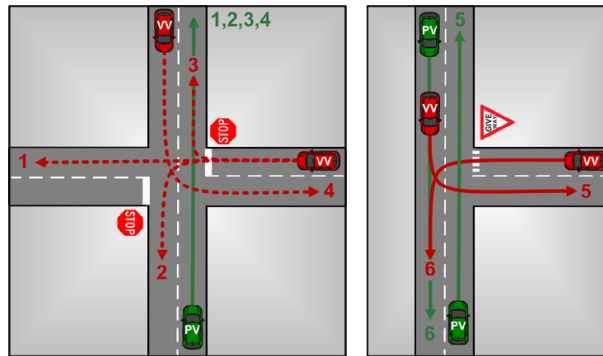


Figure 17. Scenarios tested in simulation (left) and during field trials (right).



Figure 18. Online execution of the algorithm during the field trials: warning the driver of an upcoming collision with a vehicle on the left.

However, a missile is often launched at a target range which is greater than its seeker acquisition range, although within the kinematics No Escape Zone (NEZ). It is provided with targeting geometry before launch, and maybe (via a data link) during the first part of the trajectory. However, it must fly for some period in inertial mode, and during this time the target may manoeuvre. Also, errors build up due to the imperfections in the inertial navigation system. This means that the target bearing becomes increasingly uncertain whilst the range reduces. It may be necessary to scan the seeker to acquire the target. If the scan is not matched to the possible manoeuvres, the target may escape detection. But if the scan is large, the acquisition range will be reduced, because of the reduction in search time per solid angle. As the target is acquired later, the missile's terminal manoeuvre will be more severe, and as a result the range assumed for the original kinematics NEZ may have been too optimistic. Equivalently, it is possible to be too pessimistic about the target uncertainty, hence to scan too much, and acquire the target so late that there is no longer the manoeuvre capability to reach it. Present Weapon systems optimize the probability of successful interception assuming either Gaussian uncertainties, or worst case uncertainties.

Objectives and achievements of the GUS-D system

These considerations lead to the concept of a stochastic approach for computing a probabilistic, adaptive NEZ. Probabilistic NEZ depend on the uncertain target behaviour. The uncertainties we propose to deal with are also related to the missile Inertial Navigation System (INS) precision, to sensor errors and to misalignments. Moreover, the uplink management, i.e., when to evade and breakdown the link between the launching platform and the in-flight missile plays a major role on the target localization accuracy and by the way to the size of the NEZ. Finally, there is uncertainty in the target radar cross section, which has a big effect on the seeker acquisition range. The purpose of this study is better tactical advice to the pilot about launching decision and how long maintain the uplink, and where appropriate, better matching of seeker scan strategies to target behaviours. These decisions have impacts on the probability of combat success; i.e., not only to hit the target but also on the probability to survive, as the opponent aircraft or ground threat may launch similar weapons.

The project focuses predominantly on Air to Air systems. The Meteor scan strategy has been studied deeply and is no longer critical for the engagement of fighter jets, but an objective of the study is to extend the strategy to the engagement of targets of much lower radar cross section, where the acquisition range is significantly shorter. Nevertheless all the issues apply also to Air to Ground weapon systems.

The GUS-D system is limited to one to one engagement scenario:

- one aircraft and its missile
- opponent aircraft and its missile

The main functionality of the GUS-D system is then to provide to the user a probability of successful target interception given the current engagement conditions, and the uncertainties on the target properties and behaviours.

5.3. Human Centered Navigation in the physical world

5.3.1. Goal oriented risk based navigation in dynamic uncertain environment

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

Navigation in large dynamic spaces has been addressed 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 [57]. 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 20 for an illustration), is a sampling-based partial planner guided by the risk of collision. Results concerning this work were published in [58] [59] [60]. In

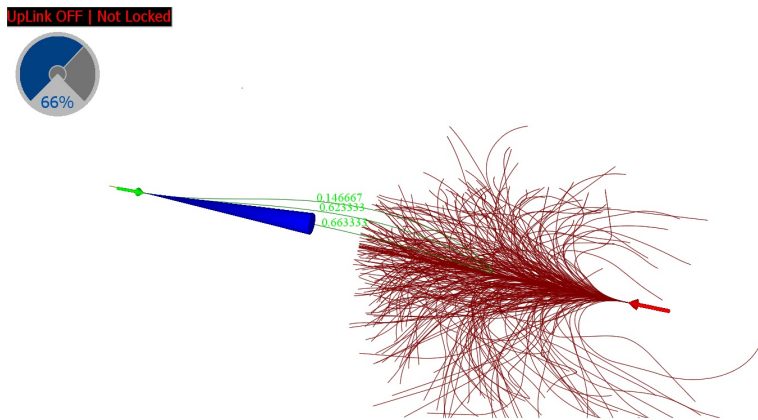


Figure 19. This Figure shows (up left corner) the “probability meter” which indicates the probability to intercept the target if the uplink is to be shutoff now. The 3D representation of the scene is displayed. The red filaments are future possible trajectories obtained with a Markov Process (MBD-UK). The blue cone modelizes the detection cone of the seeker. The green numbers indicates the probabilities to lock the target with the corresponding trajectory.

2012, We continue to work on developing probabilistic models and algorithms to analyze and learn human motion patterns from sensor data (e.g., tracker output) in order to perform inference, such as predicting the future state of people or classifying their activities. Our work has been published in the Handbook of Intelligent Vehicles [40]. We obtained some preliminary results on our robotic wheelchair combining RiskRRT with some social conventions described in section 5.3.2 . This approach and experimental results have been published at ISER 2012 [32].

This algorithms is used in the work presented in the next three sections, work conducted under the large scale initiative project PAL.

5.3.2. Socially-aware navigation

Participants: Jorge Rios-Martinez, Anne Spalanzani, Alessandro Renzaglia, Agostino Martinelli, Christian Laugier.

Our proposal to endow robots with the ability of socially-aware navigation is the Social Filter, which implements constraints inspired by social conventions in order to evaluate the risk of disturbance represented by a navigation decision.

The Social Filter receives from the perception system a list of tracked humans and a list of interesting objects in the environment. The interesting objects are designated manually according to their importance in a particular context, for example, an information screen in a bus station. After the process of such data, the Social Filter is able to output the risk of disturbance relative to people and interesting objects, on request of the planner and the decisional system. Thus, the original navigation solutions are “filtered” according to the social conventions taken into account. Notice that the concept of social filter is built as a higher layer above the original safety strategy, the planner and the decisional system are responsible to include the new constraints.

The on-board Kinect attached to our robotic platform was used to track people and to detect interactions. The Kinect sensor permits to get the position and orientation of the torso for each identified human. That information is passed to the Social Filter. Result images can be seen in Figure 21 .

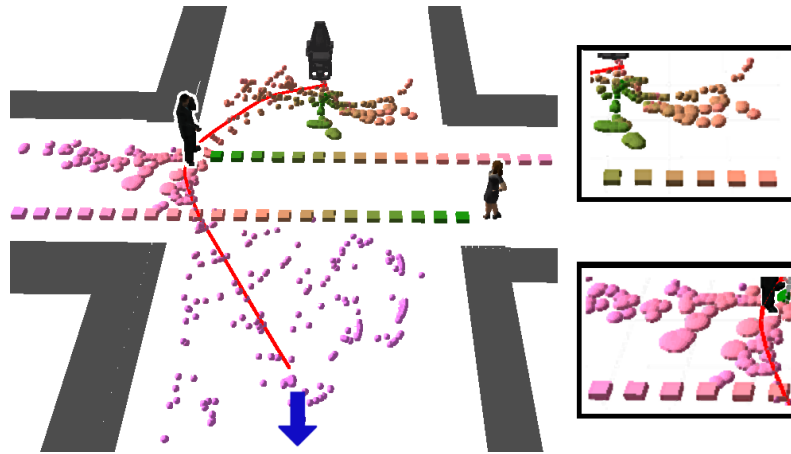
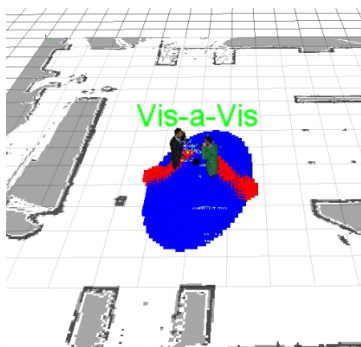


Figure 20. Predictive navigation example. RiskRRT selected a plan (red line) to the goal (blue arrow). The chosen path leads the robot to pass by the back of the first person, and then reduces the speed to let the second person to pass as well. With this strategy, the robot minimizes the risk of collision and the discomfort caused for the two pedestrians. Once second person has passed, the algorithm chooses a straighter path to the goal. Frames at the right of the figure show that estimated risk is bigger at future positions of the wheelchair (circles) which are close to predicted positions of pedestrians (squares).



(a)



(b)

Figure 21. Interaction detected with Social Filter from Kinect input for a pair of humans. Torso direction is used to estimate the main focus of interest.

In the context of socially-aware robot navigation in dynamic environments, as part of Jorge Rios-Martinez PhD thesis (to be defended in January 2013), two techniques have been proposed: one considering optimization-based navigation presented in [26] and the other a Risk-based navigation approach, previously presented in [75].

The **optimization-based navigation strategy**, done in collaboration with A. Renzaglia, is based on the Cognitive-based Adaptive Optimization (CAO) approach applied to robots [10]. We formulate the problem of socially-aware robot navigation as an optimization problem where the objective function includes, in addition to the distance to goal, information about comfort of present humans. CAO is able to efficiently handle optimization problems for which an analytical form of the function to be optimized is unknown, but the function is available for measurements at each iteration. A model of social space, contained in the Social Filter module, was integrated in order to work as a “virtual” sensor providing comfort measures. Figure 22 a) shows an image of the method implementation on ROS² framework.

Social Filter models of social conventions were combined with RiskRRT [56] by including the knowledge of human management of space (Personal Space, interaction space, activity Space). The particular considered interaction was the conversation between pedestrians which was missed in the most part of related works. The approach presented shows a way to take into account social conventions in navigation strategies providing the robot with the ability to respect the social spaces in its environment when moving safely towards a given goal. Due to the inclusion of our social models, the risk calculated for every partial path produced by RiskRRT algorithm is given by the risk of collision along the path and the risk of disturbance to human spaces.

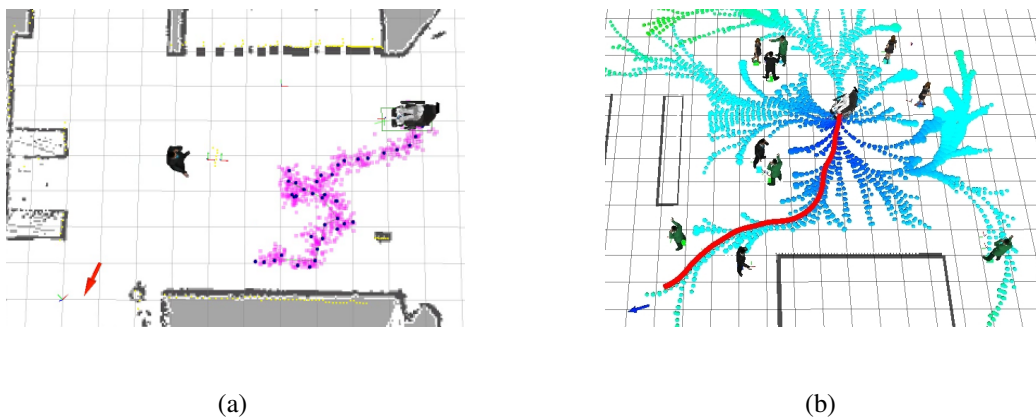


Figure 22. Results of socially-aware navigation approaches. In a) the optimization-based navigation solution avoids a region where the discomfort for the human would be higher. In b) the Risk-Based navigation technique explores the space and decides to follow a path avoiding social spaces minimizing the risk of disturbance. The goal in each case is signaled by an arrow.

One last work was presented in [25], where the socially-aware navigation based on risk was integrated with a model of human intention estimation (presented in section 5.3.4). Results exhibited emerging behavior showing a robotic wheelchair interpreting facial gesture commands, estimating the intended goal and autonomously taking the user to his/her desired goal, respecting social conventions during its navigation.

5.3.3. Navigation Taking Advantage of Moving Agents

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

²<http://www.ros.org>

Following a leader in populated environments is a form of taking advantage of the motion of the others. A human can detect cues from other humans and smartly decide in which side to pass. Humans can also easily predict the motion of the others, changing his/her path to accommodate for conflictive situations, for example. Imitating the motion of a human can also improve the social acceptance of robots and so on.

The best leader is the one whose goal is close to the robot's one. To implement that, the Growing Hidden Markov Model (GHMM) technique is used [79]. This technique provides at the same time a capability to learn and modeling typical paths, as well as learning and predicting goals associated to paths, making it ideal for the proposed approach of leader election.

Once a leader is chosen, the robot starts to track his/her path and follow it, using the RiskRRT algorithm presented in section 5.3.1. This algorithm takes into account the risk of collision with other agents, guaranteeing that the robot can avoid collisions even if its leader is lost or occluded.

Some results can be seen in the following experiments, where real human data was used together with a robot simulator.

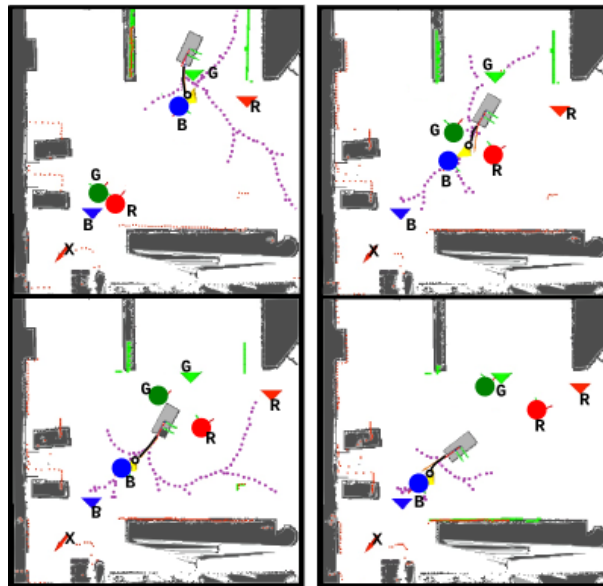


Figure 23. Robot navigation following a leader seamlessly avoid other incoming agents.

In Figure 23, the experiment demonstrates one of the advantages of following a leader to improve the robot's navigation capabilities. The direct path to the robot's goal is obstructed by two incoming humans. Normally an algorithm suited for dynamic environment would create a detour as future humans' position would conflict with the robot straight trajectory. However, as the robot is following a leader, it does not reason about the other agent's future position. Therefore, the leader knows that people will give room for he/she to pass, and the robot profits from it.

Next step will be to use this technics will navigating in a crowd, task that a common planning strategy could hardly do.

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 [77]. 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.

In 2011 a robotic wheelchair was set up as experimental platform. Some basic functions were included as the mapping of the environment using a Rao-Blackwellized Particle Filter [62], localization using an Adaptive Monte Carlo Localization approach (AMCL) [78], global planning using an A* algorithm [63] and local reactive planning using the Dynamic Window Algorithm [55]. Alongside some work was done with the kinect sensor in order to detect and track people. This behaviour was aimed to bring assistance not only to the user but also to the caregiver by allowing him to move more freely. The software implementation of the related approaches was done on the basis of the ROS middleware.

During 2012 the work was centered in the improvement of the usability of the system around three main axes:

- User intention estimation: 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 [46],[47]. 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 figure 24, the chosen approach uses a Bayesian model 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. Current research is being done to understand which information is important to take into account in order to do better estimations of the user's intention. In particular, the movements of the head are considered by the proposed inference method.

The navigation is performed using the human-aware planning algorithm developed by the team which integrates a notion of social conventions and avoidance of dynamic obstacles to prevent uncomfortable situations when the wheelchair is navigating among humans (see section 5.3.2 for details)

- 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 multimodal functionality as described in [47]. 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). The use of a touch screen control is also possible.
- Multimodal control: The wheelchair can be controlled in semi-autonomous mode employing the user's intention estimation module, described later, or in manual mode in which the user is in charge of driving by him self.

In manual mode the user controls the wheelchair's angular speed moving her head while the linear speed is controlled with vocal commands (faster, slower, break, etc).

In semi-autonomous mode the user shows the direction to his/her desired destination facing towards it. Whenever a new command is read from the face pose estimation system. The user's intention module computes the goal with the highest posterior probability. The navigation module receives the map of the environment, the list of humans present in the scene and the currently estimated goal to compute the necessary trajectory to the goal.

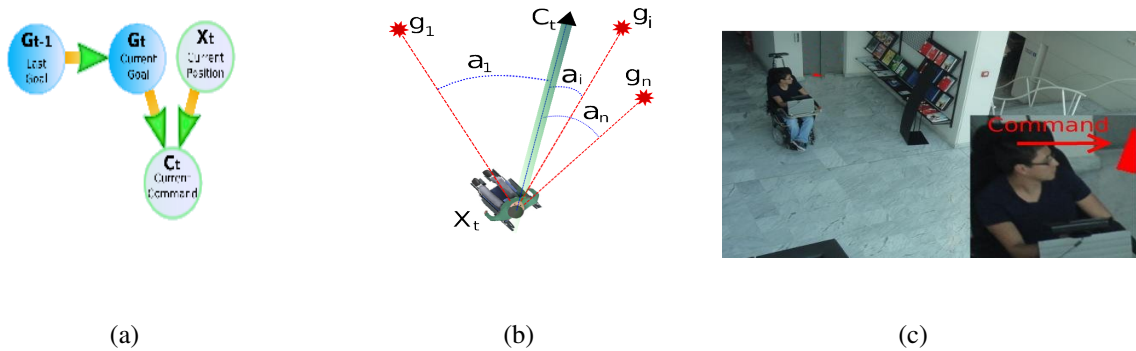


Figure 24. **left:** User's intention model. The Bayesian network used to estimate the current user's intended goal G_t . The current position X_t and the user command C_t are used as evidence. G_t is dependent on the value of the last estimation G_{t-1} . **center:** Experimental evaluation of the user's intention module. 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.

5.3.5. Multi-Robot Distributed Control under Environmental Constraints

Participants: Agostino Martinelli, Alessandro Renzaglia.

This research is the follow-up of a study begun three years ago in the framework of the European project sFly. The problem addressed is the deployment of a team of flying robots to perform surveillance coverage mission over an *unknown* terrain of complex and non-convex morphology. In such a mission, the robots attempt to maximize the part of the terrain that is visible while keeping the distance between each point in the terrain and the closest team member as small as possible. A trade-off between these two objectives should be fulfilled given the physical constraints and limitations imposed at the particular application. As the terrain's morphology is unknown and it can be quite complex and non-convex, standard algorithms are not applicable to the particular problem treated in this paper. To overcome this, a new approach based on the Cognitive-based Adaptive Optimization (CAO) algorithm is proposed and evaluated. A fundamental property of this approach is that it shares the same convergence characteristics as those of constrained gradient-descent algorithms (which require perfect knowledge of the terrain's morphology and optimize surveillance coverage subject to the constraints the team has to satisfy). Rigorous mathematical arguments and extensive simulations establish that the proposed approach provides a scalable and efficient methodology that incorporates any particular physical constraints and limitations able to navigate the robots to an arrangement that (locally) optimizes surveillance coverage.

Special focus has been devoted to adapt this general approach in order to deal with real scenarios. Specifically, this has been carried out by working in collaboration with the ETHZ (Zurich). To this regard, the approach has been adopted in the framework of the final demo of the sFly project. The demo simulates a search and rescue operation in an outdoor GPS-denied disaster scenario. No laser, no GPS, and Vicon or other external cameras are used for navigation and mapping, but just onboard cameras and IMUs. All the processing runs onboard, on a Core2Duo processing unit. The mission consists of first collecting images for creating a common global

map of the working area with 3 helicopters, then engaging positions for an optimal surveillance coverage of the area, and finally detecting the transmitter positions.

The results of this research have been published in two journals, [14], [15], and on the thesis of A. Renzaglia, [10].

5.4. Bayesian Modelling of Sensorimotor Systems and Behaviors

Results described in this section were done in collaboration with the LPPA collège de France.

5.4.1. Bayesian based decision making in multi-player video games

Participants: Gabriel Synnaeve, Pierre Bessière.

The problem addressed in this work is the autonomous replacement of a human player. It is the continuation of last year's work on the same topic as well as a follow-up of previous E-Motion Ph.D Ronan Le Hy [64]. This year, we focused on real-time strategy (RTS) games, in which the players have to build an economy, advance technology, produce and control an army to kill the opponents. From a research point of view, multi-player games are interesting because they stand for a good in-between of the real world and simulations. The world is finite and simulated (no sensors problems) but we didn't wrote the simulation and the other players are humans (or advanced robots in the case of AI competitions).

This year's research work focused on tactical prediction and decision-making as well as armies composition adaptation. For the tactical model, the idea is to have (most probably biased) lower-level heuristics from units observations, which produce information exploitable at the tactical level, and take advantage of strategic inference too. We abstract space into automatically extracted choke points and regions of StarCraft maps from a pruned Voronoi diagram (using [71]). We then assign different scores to each of these regions and learn the influence of these scores on different attack types and locations. To do that, we set up a huge data-set of professional player's games, whose game state was extracted [29]. This work was accepted for publication at Computational Intelligence in Games (IEEE CIG) 2012 in Grenada [30] and was presented at the Computer Games Workshop of the European Conference of Artificial Intelligence (ECAI) 2012 [28].

Another focus of work this year was on army composition adaptation. RTS games unit types combinations in armies can be seen as complex (soft max) rock-paper-scissors games. Our analysis boiled to down army compositions encoded as clusters (we used a Gaussian Mixtures Model) of "classic" combinations (because of economy and technology constraints during the game). This work was published at the AI in Adversarial Real-Time Games workshop of AAAI AIIDE 2012 [29].

On top of the research/evaluation implementation, we also implemented it in our StarCraft: Broodwar's bot implementation BroodwarBotQ. With this bot, we took part in AIIDE and CIG conferences AI tournaments placing respectively 4th (out of 10) and 6th (out of 10). Gabriel Synnaeve defended his thesis on October 24th 2012.

5.4.2. 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 computational 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.



Figure 25. Units movement debugging output during a StarCraft game of the BroodwarBotQ bot. Considering the unit in the upper middle of the picture, white squares represent the highest probabilities of directions, while the darker the blue, the lower the probability to go there. The unit controller searches both to minimize collisions and stay in range of the enemy targets (bottom right).

We have designed an integrative Bayesian model which allows to study auditory, motor and perceptuo-motor aspects of speech production and perception. In 2011, this model was used to work on purely theoretical simulations where we studied with diverse paradigms the decrease in the performances predicted by the different theories due to communication noise. This work led to the proof of an indistinguishability theorem : given some hypotheses on the learning process, purely motor and purely auditory models have identical answers to perception tasks. Thanks to VLAM, a vocal tract simulation tool which allows to map articulatory parameters to acoustic signals, we tested our model on vowel perception tasks. The results of both these studies are detailed in [70].

In 2012, we worked on a much more complex version of the model, which was made able to deal with plosive syllable production and perception. A first version of 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 than it is the case in practice, due to unrealistic learning methods. That's why 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.

5.4.3. Bayesian programming : book and software

Participants: Emmanuel Mazer, Pierre Bessi re.

5.4.3.1. A need for a new computing paradigm

Bayesian probability theory is a mathematical alternative to logic.

However, we want working solutions to incomplete and uncertain problems. Consequently, we require an alternative computing framework based on Bayesian probabilities.

To create such a complete computing Bayesian framework, we require a new *modeling methodology* to build probabilistic models, we require new *inference algorithms* to automate probabilistic calculus, we require new *programming languages* to implement these models on computers, and finally, we will eventually require new *hardware* to run these Bayesian programs efficiently.

Our ultimate goal is a Bayesian Computer. The purpose of this book is to describe a formalism and a computer language as first steps in this direction.

5.4.3.2. Outline of the book

Its purpose is to introduce the fundamental concepts of Bayesian Programming, to present the novelty and interest of the approach, and to initiate the reader to the Bayesian modeling. Numerous simple examples of applications are presented in different fields.

It is divided in three parts, chapters 2 *Basic-Concepts* to 6: *Bayesian-Program* which presents the principles of Bayesian Programming, chapters 7 : *Information-Fusion* to 11 : *Bayesian-Programming-Iteration* which offer a cook book for the good practice of probabilistic modeling and 12 : *Bayesian Programming Formalism* to 16 *Frequently Asked Question* which revisit the Bayesian inference and learning problems with the help of the presented formalism.

A first version of the book will be send to the reviewer selected by the editor before the end of 2012

5.4.3.3. Distributed Software

One way to read this book and learn bayesian programming is to run and modify the programs given as example. A Python package “pypl” based on Probt bindings is made available with book.

The source code of the examples as well as the Python package can be downloaded free of charge.

Many examples in the book are given with parts of real corresponding programs which could be run using the distributed package. They are given under the following format

The figure 26 has been generated using the program “chapter7/invpgm.py”. The following instruction allows to get to most probable value for the heading H given the readings.

```
PH=PHkBOB1.instantiate(sensor_reading_values)
best=PH.compile().best()
```

Bayesian programs are also used to generate the illustration of the book such as this one which illustrating the navigation based on sensor fusion.

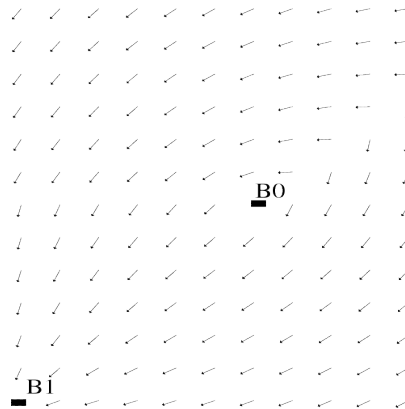


Figure 26. The vector field corresponding to $\max_h P(H = h | b_0 \wedge b_1 \wedge \pi)$

EXMO Project-Team

6. New Results

6.1. Ontology matching and alignments

We pursue our work on ontology matching and alignment support with contributions to evaluation and alignment semantics.

6.1.1. Evaluation

Evaluation of ontology matching algorithms requires to confront them with test ontologies and to compare the results. 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 two evaluation campaigns named 2011.5 and 2012. This was justified by the will to complete full evaluations using the support of the SEALS platform. Hence, the main activities carried out in 2012 were related to the automation of the evaluation. This involved providing participants with a better way to bundle their tools so that they can be evaluated both offline and within the SEALS platform. It also required to support more organisers to provide test case within the platform.

This work has been used in the OAEI 2012 evaluation campaign. OAEI 2012 offered 9 different test sets (7 of which under the SEALS platform). This issue brought the following results:

- More participants than ever (21);
- All ontology matchers running on the SEALS platform (18);
- Increased performances in terms of precision and recall;
- Matchers are now very scalable and can deal with the largest available ontologies (9 systems able to deal with the very large medical ontology SnoMed);

We have also introduced as a data set, the benchmark for multilingual ontology matching developed last year [6]. It has pushed matcher developers to address multilingual issues.

The participating systems and evaluation results were presented in the 7th Ontology Matching workshop, that was held in Boston, MA, US [22], [7]. More information on OAEI can be found at <http://oei.ontologymatching.org/>.

6.1.2. Semantics for weighted correspondences

Alignment correspondences are often assigned a weight or confidence factor by matchers. Nonetheless, few semantic accounts have been given so far for such weights. We have proposed a formal semantics for weighted correspondences between different ontologies. It is based on a classificational interpretation of correspondences: if o and o' are two ontologies used to classify a common set X , then alignments between o and o' are interpreted as encoding how elements of X classified in the concepts of o are re-classified in the concepts of o' , and weights are interpreted as measures of how precise and complete re-classifications are. This semantics is justifiable for extensional matchers. We have proven that it is a conservative extension of a semantics of absolute correspondences, and we have provided properties that relate correspondence entailment with description logic constructors [8].

This work has been made in cooperation with Alexander Borgida (Rutgers University) and Chiara Ghidini and Luciano Serafini (Fondazione Bruno Kessler).

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.

6.2.1. Keys and pseudo-keys detection for web datasets cleansing and interlinking

We have proposed a method for analysing web datasets based on key dependencies. For this purpose, we have adapted the classical notion of a key in relational databases to the case of RDF datasets [9], [16]. In order to better deal with web data of variable quality, we have introduced the definition of a pseudo-key. We have also provided an RDF vocabulary for representing keys and pseudo-keys and designed and implemented an algorithm for discovering them. Experimental results show that, even for a large dataset such as DBpedia, the runtime of the algorithm is still reasonable. We are currently working on two applications: data cleansing, i.e., detection of errors in RDF datasets and recovery, and datasets interlinking.

The algorithm is publicly available at <https://gforge.inria.fr/projects/melinda/>.

6.2.2. Data interlinking from expressive alignments

In the context of the DATALIFT project (see §7.1.1), we are developing a data interlinking module. Based on our analysis of the relationships between ontology matching and data interlinking [15], our goal is to generate data interlinking scripts from ontology alignments. For that purpose, we have integrated existing technologies within the DATALIFT platform: the Alignment API, for taking advantage of the EDOAL language and SILK, developed by Freie Universität Berlin, for processing linking scripts. So far, we have generated SILK script from ontology alignments in order to produce links.

This work is part of the PhD of Zhengjie Fan, co-supervised with François Scharffe (LIRMM).

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

Querying the semantic web is mainly done through the SPARQL language [18]. We designed one of its extensions, PSPARQL (Path SPARQL) which provides queries with paths of arbitrary length. We continue this work by connecting it to the work of the WAM team on static analysis of XPATH expressions. More specifically, we consider query containment, i.e., determining whether, for any graph, the answers to a query are contained in those of another query. This is achieved by reducing this problem to satisfiability in the μ -calculus. In this work, RDF graphs are considered as transition systems and important fragments of RDFS and SPARQL as propositional μ -calculus formulas. It is then possible to use solvers of this logic to test query containment of SPARQL queries under RDFS and OWL schema constraints [11], with paths or under particular entailment regimes [10]. We have also implemented the proposed techniques and provided a first benchmark for query containment available under <http://sparql-qc-bench.inrialpes.fr>.

This work is part of the PhD of Melisachew Wudage Chekol [4], co-supervised with Nabil Layaida (WAM).

IMAGINE Team

6. New Results

6.1. 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 2012 for each axis.

6.2. High level model for shapes

Scientist in charge: Stefanie Hahmann

Other permanent researchers: Marie-Paule Cani, Jean-Claude Léon, Damien Rohmer.

6.2.1. Implicit surface modeling

Participants: Adrien Bernhardt, Marie-Paule Cani, Maxime Quiblier, Cédric Zanni.

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

In collaboration with a researcher in formal computation, Evelyne Hubert, we improved and extended the analytical methods for computing closed form solutions for convolution surfaces [6].

Within Cédric Zanni's PhD we proposed a method based on anisotropic, surface Gabor noise, for generating procedural details on skeleton-based implicit surfaces, see Figure 4 (left). The surfaces enhanced with details can still be smoothly blended, with a natural transition between the details they carry [19].

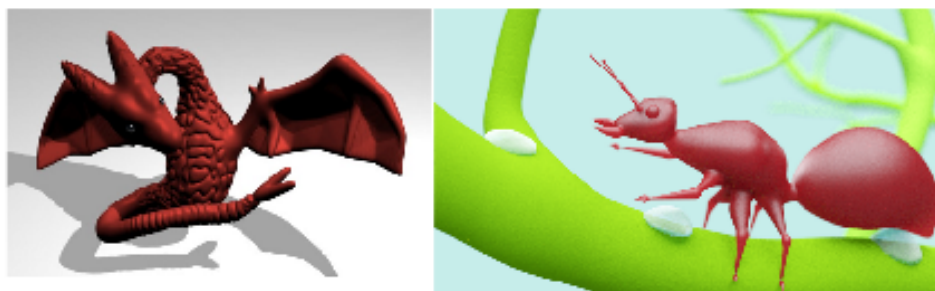


Figure 4. Left: Dragon model showing the variety of details that can be generated. Computation time was less than 2 minutes. Right: Shape obtained by the use of scale-invariant integral surfaces.

We also developed an extension to convolution surfaces, so-called scale-invariant integral surfaces, see Figure 4 (right). Thanks to blending properties that are scale invariant these surfaces have three major advantages: the radius of the surface around a skeleton can be explicitly controlled, shapes generated in blending regions are self-similar regardless of the scale of the model, and thin shape components are not smoothed-out anymore when blended into larger ones. This work has been presented at AFIG2012 [23] and submitted for international publication.

Lastly, in collaboration with Loic Barthe in Toulouse, we contributed to a new blending operator, gradient blending, which enables us to blend implicit shapes not only in function of the 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 genus of the blended shape remains the one of the union of the input one, and avoid the blur of small details. A paper has been accepted for publication in ACM ToG [4].

6.2.2. Developable surfaces

Participants: Rémi Brouet, Marie-Paule Cani, Stefanie Hahmann, Damien Rohmer.



Figure 5. Design preserving garment transfer of a multi-layer outfit from a woman to a young girl. Middle: Automatically graded patterns shown to scale. Right: The zoomed-in source and target patterns for the back panel highlight the subtle changes in shape

A developable surface is a surface, which can be unfolded (developed) into a plane without stretching or tearing. Because of this property, developable surfaces lead to a variety of applications in manufacturing with materials that are not amenable to stretching (leather for shoes or hand bags, skins of aircrafts, sails). In computer graphics developable surfaces are very popular to model, simulate or animate clothes or folded papers in virtual environments.

In collaboration with Alla Sheffer (University of British Columbia, Canada visiting Inria) we developed a fully automatic method for design-preserving transfer of garments between characters with different body shapes. The method is able to generate design-preserving versions of existing garments for target characters whose proportions and body shape significantly differ from those of the source. The work has been presented at SIGGRAPH 2012 [1].

Folded paper exhibits very characteristic shapes, due to the presence of sharp folds and to exact isometry with a given planar pattern. In the past we proposed a purely geometric solution to generate static folded paper geometry from a 2D pattern and a 3D placement of its contour curve. Current research focusses on the interactive manipulation of the folded surface without the strong requirement of starting by an initial contour curve, but using sparser positional constraints on the surface.

Damien Rohmer joined in 2012 the Hevea project: this is a project in collaboration between Vincent Borrelli (Institut Camille Jordan, Lyon), Boris Thibert (MGMI, LJK Grenoble) and Francis Lazarus (Gipsa Lab, Grenoble) focussed on the generation and visualisation of the flat torus. The flat torus is a mathematical smooth surface with the topology of a torus but having locally the metric of the plane. In other word, this

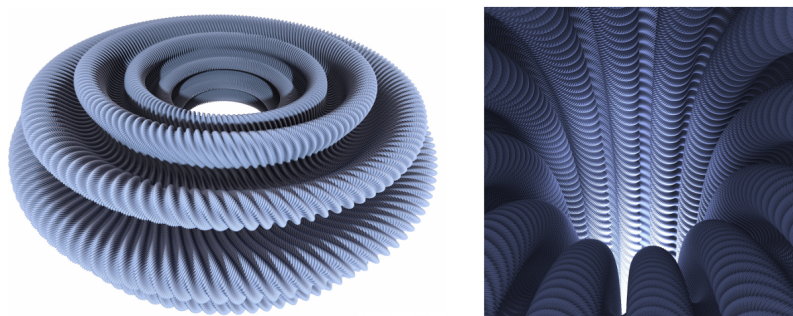


Figure 6. The first representation of a flat torus.

is a developable torus. So far, no representation of such object had ever being made. In 2012, based on a convex integration algorithm generating coherent wrinkles on the torus called *corrugations*, we generated the first representation of such object that is both C^1 while being fractal as the number of wrinkles has to tend to infinity to converge toward true developability. The rendering made by Damien Rohmer has been used for the cover image of Proceedings of the National Academy of Sciences (PNAS) (<http://www.pnas.org/content/109/19.cover-expansion>).

6.2.3. Parametric surfaces

Participant: Stefanie Hahmann.

We are developing new smooth parametric surface models defined on irregular quad meshes. They are in fact a powerful alternative to singularly parameterized tensor product surfaces since they combine the advantages of both, the arbitrary topology of quad meshes and the smoothness of the tensor product patches.

In collaboration with G.-P. Bonneau (Maverick team) several parametric triangular surface models for arbitrary topologies have been published in the past (CAGD, IEEE TVCG and ACM ToG). A new tensor product spline surface model has been developed this year. It solves the problem of defining a G^1 -continuous surface interpolating the vertices of an irregular quad mesh with low degree polynomial tensor product patches. It further aims to produce shapes of very high visual quality while reducing the number of control points. A comparison with existing methods and a journal paper are in preparation.

6.2.4. Fibrous structures

Participant: Damien Rohmer.

Due to anisotropy, fibrous structures may exhibit complex deformation properties. These properties are of main interest to understand the behavior of some human organs such as the heart.

In collaboration with Grant Gullberg, Archontis Giannakidis from Lawrence Berkeley Laboratory, and Alexander Veress from University of Washington we developed a new visualization of heart defects based on the fibrous structure organization. Combining 3D visualization with the fiber structure analysis may help to detect heart defects such as cardiac Hypertrophy. This work as been published as a book chapter [29].

6.2.5. Virtual Prototypes

Participants: Flavien Boussuge, Francois Faure, Stefanie Hahmann, Jean-Claude Léon.

In the context of virtual prototyping (process of product development involving CAD/CAE software), a DMU (digital mock up) is the container of all the components of a 3D virtual product that be used during design and simulations.

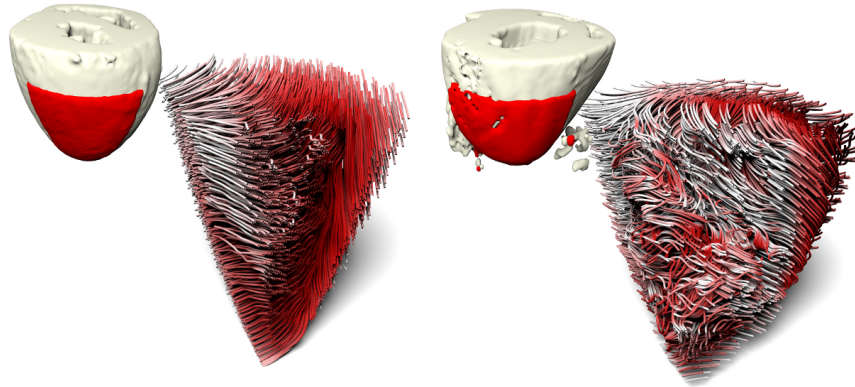


Figure 7. Visualization of fibrous structure in the heart for a normal case (left) and a defect heart (right).

Herein geometric interfaces, i.e. the imprint of a component onto each of its neighboring components, must be taken into account to generate simulation models. Indeed, a DMU does not contain these geometric interfaces. However, extensive use of CAD assemblies has led to increasingly complex DMUs with up to hundreds of thousands of components. The detection and generation of the geometric interfaces between all components with existing software is a very tedious task, which may require hours or days of user-interaction or is even not possible. As part of the ANR project ROMMA in collaboration with Georges-Pierre Bonneau and Francois Jourdes from the Maverick team, we developed a new method to rapidly detect and precisely describe the geometry of interfaces in highly complex assemblies [20].

Within the PhD of Flavien Boussuge, we take advantage of these interfaces to focus on the generation of mixed dimensional models from enriched DMUs for FE analysis of structural assemblies. The goal is to provide a methodology and operators for transforming geometries of complex assemblies so that they are directly usable for FE mesh generation. A first contribution to assembly model preparation for simulation has been presented at ECT12 [11]. Herein, a model preparation methodology has been proposed that addresses the shape transformation categories specific to assemblies. Current and future research includes the generation of construction graphs of volume models that contribute to idealization operators. These algorithms take the simulation objectives into account as part of the proposed methodology.

Another important issue connected to geometry transformation of assemblies and construction graphs of volume models relates to the global as well as partial symmetries of components and assemblies. Here, symmetry analysis is applied to B-Rep NURBS models and must be obtained within the tolerance of a geometric modeler, which differs rather significantly from approximate symmetries extracted from meshes. The symmetry analysis helps structuring the construction graphs of volume models to take into account repetitive locations of primitives. Also, symmetry properties combine with functional annotations of components to enhance their search and retrieval[16].

6.3. Models for real-time motion synthesis

Scientist in charge: François Faure

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

6.3.1. Interactive manipulation of folded paper surfaces

Participant: Damien Rohmer.

Although physically-based simulation has become very popular to model deformable surfaces such as cloth it is still not applicable to generate animations of creased paper. Due to the stiffness of this incompressible material and to the complex changes of its mechanical behavior during creasing. As a result, this standard material in every-day life almost never appears in Computer Graphics applications such as movies or video games. Animating creased paper brings two main challenges: First, such surface needs to be deformed while preserving its length in every direction according to its original pattern. Secondly, sharp features, which are not commonly handled in numerical simulators, need to be generated on the surface.

With the master work of Ulysse Vimont, we developed a prototype (as seen in fig. 8) of a deformation tool enabling to interactively manipulate a virtual sheet of paper. The approach is a procedural approach based on some geometrical apriori knowledge of behavior of paper under deformation. We plan to extend this work in the next year with a new master student Camille Shreck.

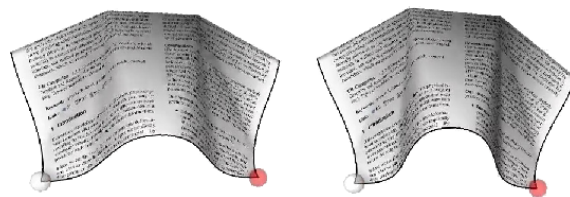


Figure 8. Example of interactive manipulation of a sheet of paper.

6.3.2. Real-time skinning deformation with contacts

Participants: Marie-Paule Cani, Damien Rohmer.

Skinning deformation based on linear blending or dual quaternion approach is a very popular technique thanks to its fast computation. However, they do not capture the complex behavior of flesh bulging and contact between body parts.

In collaboration with Loic Barthe, Rodolphe Vaillant from IRIT Toulouse, and Gael Guennebaud from LaBRI Bordeaux, we developed a skinning deformation handling both flesh bulges and collision avoidance.

An implicit surface is first fitted onto the original mesh surface. During the animation, the mesh is deformed using a standard skinning deformation while the implicit surface follows the rigid articulation of the bone. Finally, the mesh is projected back toward the deformed implicit surface enabling to both compensate for mesh collapse and self collision. This work has been presented in AFIG [10] conference and won the *best article* award. It has also been accepted for publication in the REFIG Journal.

6.3.3. Particle-based simulation of concrete structures

Participants: Marie Durand, François Faure.

In collaboration with the LIG and L3S-R labs, we have published results on gpu-accelerated simulation of concrete fracturation due to impacts [2], leading to a speedup factor of about 15 compared to a CPU implementation. This led us to notice that collision detection was the major bottleneck. Consequently, we investigated and published a new incremental sorting method to more efficiently cluster the particles along a Z-curve, by improving the Packed Memory Array data structure for fast updates [15], as illustrated in Figure 9 . We have proposed a new strategy to efficiently update the sorting, while maintaining a desired fill rate in each branch of the tree structure. Experiments show that our PMA can outperform a compact sorted array for up to 50% particle cell changes per time step.

6.3.4. Collision detection and response

Participant: François Faure.

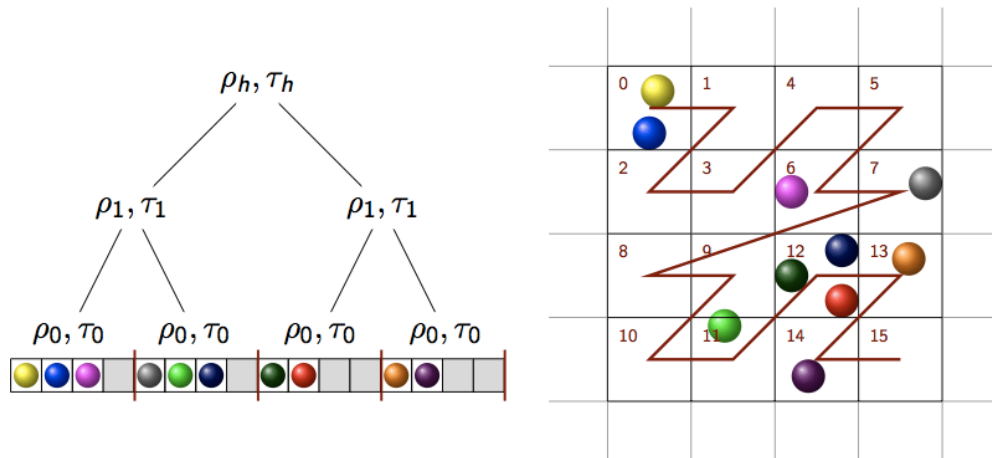


Figure 9. Z-sort using a Packed Memory Array structure. The gaps allow fast updates.

In collaboration with UBC, Vancouver, we have presented at SIGGRAPH 2012 a new method for image-based contact detection and modeling, with guaranteed precision on the intersection volume [8]. Unlike previous image-based methods, our method optimizes a nonuniform ray sampling resolution and allows precise control of the volume error. By cumulatively projecting all mesh edges into a generalized 2D texture, we construct a novel data structure, the Error Bound Polynomial Image (EBPI), which allows efficient computation of the maximum volume error as a function of ray density. Based on a precision criterion, EBPI pixels are subdivided or clustered. The rays are then cast in the projection direction according to the non-uniform resolution. The EBPI data, combined with ray-surface intersection points and normals, is also used to detect transient edges at surface intersections. This allows us to model intersection volumes at arbitrary resolution, while avoiding the geometric computation of mesh intersections. Moreover, the ray casting acceleration data structures can be reused for the generation of high quality images, as illustrated in Figure 10.

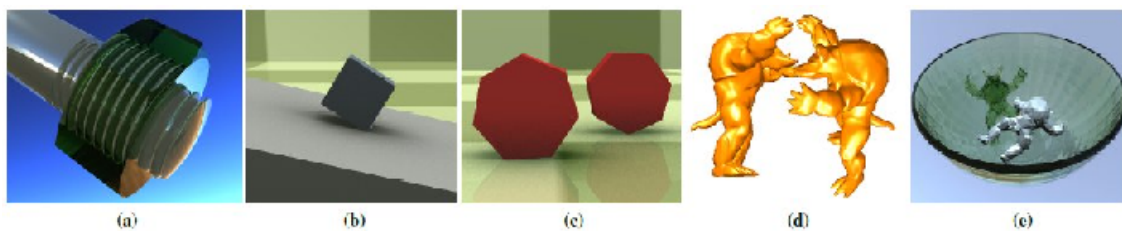


Figure 10. Examples of challenging contact scenarios handled by our method. (a) The movement of a tight fitting nut on a bolt can be simulated directly using the geometric models. (b) Very small geometric features on a flat surface can dramatically change the behavior of objects sliding on it. (c) “Ruina wheels.” Two almost identical wheels have obviously different rolling behavior due to subtle features (one is slightly convex and another is slightly concave); our method can simulate this contact behavior realistically. (d) A simulation with 4.4 million triangles. (e) A snapshot of an interactive simulation with ray-traced rendering.

6.3.5. Action representation, segmentation and recognition

Participant: Remi Ronfard.

Following Daniel Weinland's PhD thesis, we published a survey of modern methods for representing, segmenting and recognizing full-body actions in video [32]. A taxonomy of methods is elaborated in that paper, where actions can be represented with local, structured or global features both in time and in space. The potential for future work in grammar-based action recognition is emphasized, with possible applications in corpus-based procedural modeling of actions.

6.3.6. Simulation software architecture

Participants: Ali-Hamadi Dicko, Guillaume Bousquet, François Faure.

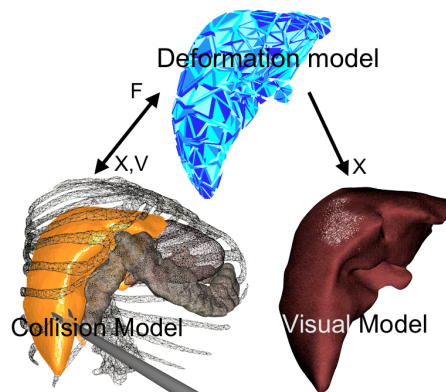


Figure 11. A simulated Liver. Three representations are used: one master model for the internal deformable mechanics, one for the collisions, and one for the visualization. Mappings (black arrows) are used to propagate positions (X) and velocities (V) from master to slaves, while forces (F) are propagated in the opposite direction

We continue the development of SOFA, the open source simulation library, which is becoming an international reference, and we have published a chapter on it in a Springer book [28]. SOFA facilitates collaborations between specialists from various domains, by decomposing complex simulators into components designed independently and organized in a scenegraph data structure. Each component encapsulates one of the aspects of a simulation, such as the degrees of freedom, the forces and constraints, the differential equations, the main loop algorithms, the linear solvers, the collision detection algorithms or the interaction devices. The simulated objects can be represented using several models, each of them optimized for a different task such as the computation of internal forces, collision detection, haptics or visual display, as illustrated in Figure 11. These models are synchronized during the simulation using a mapping mechanism. CPU and GPU implementations can be transparently combined to exploit the computational power of modern hardware architectures. Thanks to this flexible yet efficient architecture, SOFA can be used as a test-bed to compare models and algorithms, or as a basis for the development of complex, high-performance simulators

6.3.7. Real time fluid animation on GPU

Participant: Martin Guay.

In collaboration with Manuel Vennier (Maverick, Inria), we developed a simple and fast method to animate fluids on the GPU. Inspired from the classical SPH method (Smooth Particles Hydrodynamics), we express a weakly compressible formulation for the fluid animation. Contrary to standard approaches, we fully developed the formulation on a grid, leading to an efficient GPU implementation. The method replaces the implicit formulation of pressure by an explicit one based on density invariance. We therefore propose a method to simulate 3D Eulerian gaseous fluids in a single pass on the GPU. The results published in [22] are less accurate than a standard fluid simulation approach, but lead to real-time fluid-looking models (see fig. 12) which are practicable for video games or other interactive applications.



Figure 12. Example of fluid results obtained by our approach in [22].

6.4. Knowledge-based models for narrative design

Scientist in charge: Rémi Ronfard

Other permanent researchers: François Faure, Jean-Claude Léon, Olivier Palombi

6.4.1. Computational model of film editing

Participants: Remi Ronfard, Quentin Galvane.

Collaboration with the Mimetic team (Marc Christie) is continuing on this topic as part of the CINECITA (ANR jeune chercheur) and CHROME (ANR) projects.

We presented a survey of automatic video editing and new results from our ongoing collaboration at the first workshop on intelligent cinematography and editing (WICED) which took place during the Foundation of Digital Games (FDG) international conference [18], [14].

6.4.2. Stochastic Plex Grammars

Participant: Remi Ronfard.

During Quentin Doussot's master thesis, we experimented with stochastic plex grammars, which proved to be efficient for generating 3D scenes in the style of Keith Haring [17]. The model is able to generate static scenes by assembling colorful body parts into Keith Haring figures. The model is also able to simulate Markov chains of such figures by randomly changing attributes and composition of the scene.

6.4.3. Reframing theatre performances

Participants: Remi Ronfard, Vineet Gandhi.

In 2012, we made full-hd video recordings of rehearsals and performances at Celestins - Theatre de Lyon:

- A l'Ouest, directed by Nathalie Fillion. Coproduction - Théâtre du Rond-Point, Célestins, Théâtre de Lyon, Cie Théâtre du Baldaquin, AskUs , Le Gallia Théâtre-Saintes. Coproduction - Théâtre du Rond-Point, Célestins, Théâtre de Lyon, Cie Théâtre du Baldaquin, AskUs , Le Gallia Théâtre, Saintes.
- Lorenzaccio, directed by Claudia Stavisky, Théâtre de Saint Petersburg.
- Mort d'un commis, directed by Claudia Stavisky, Célestins, Théâtre de Lyon.

As part of his PhD thesis, Vineet Gandhi developed novel algorithms for actor detection and naming. This has been tested on movies as well as theatre performances. Current work is focusing on automatically reframing those recordings into cinematically-valid shots focusing on one or more actors.

A related thread of work was started for semantic annotation of the recordings using the syntax and semantics of blocking notations, a symbolic notation used in North-American theatres [25].

6.4.4. *Virtual theatre rehearsals*

Participant: Remi Ronfard.

We are starting to investigate the possibility of rehearsing theatre plays with real and virtual actors, using extensions of interactive scores initially proposed for computer music. A position paper was presented to researchers in theatre studies during a seminar on the notation of theatre [24].

6.4.5. *Extracting functional information from assembly models*

Participants: Jean-Claude Léon, Ahmad Shahwan, Olivier Palombi.

Assembly models of products, as available from CAD software reduce to a set of independent geometric models of its components and a logical structure of the assembly described as a tree containing components' names. Such a model lacks of geometric connections between its components and the work performed at 6.5 contributes already to structure the geometric model of each component with its geometric interfaces. However, the assembly tree structure and components' names still have no connection with the geometric model of components and their names don't convey robust information because their are user chosen. Here, the purpose is to set tight connections between components' geometric models and their functions. Using dualities between geometric interfaces and interaction forces, it is possible to initialize qualitative mechanical values at each geometric interface, producing different possible configurations.

Then, the proposed approach builds upon relationships between function, behavior and shape to derive functional information from the geometry of component interfaces. Among these concepts, the concept of behavior is more difficult to set up and connect to the geometry of interfaces and functions. Indeed, states and design rules are introduced to express the behavior of components through a qualitative reasoning process [7]. This reasoning process, in turn, takes advantage of domain knowledge rules and facts, checking the validity of certain hypotheses that must hold true all along a specific state of the product's lifecycle, such as operational, stand-by or relaxed states. Eliminating configurations at geometric interfaces that contradict one or more of those hypotheses in their corresponding reference state reduces ambiguity, subsequently producing functional information in a bottom-up manner.

This bottom-up process starts with the generation of a Conventional Interfaces Graph (CIG) with components as nodes, and conventional interfaces (CI: the geometric interfaces) as arcs. A CI is initially defined by a geometric interaction that can be a contact or an interference between two components. CIs are then populated with Functional Interpretations (FI) according to their geometric properties, producing potentially many combinations. A first step of the reasoning process, the validation against reference states, reduces the number of FIs per CI. Then, a matching process takes place using inferences of an ontology reasoner to produce a functional designation of each component. The ontology is based on several taxonomies: conventional interfaces, functional interfaces and functional designations that are connected through the qualitative reasoning process. As a result, the geometric model of each component assigned with a functional designation becomes intrinsically structured with functional interfaces (see Figure 13). Structured models can

be used to perform high level shape transformations like virtual prototypes. MyCorporisFabrica is a software framework we started to connect to. This activity is part of the ANR ROMMA project. It is a first contribution to simulation scenarios.

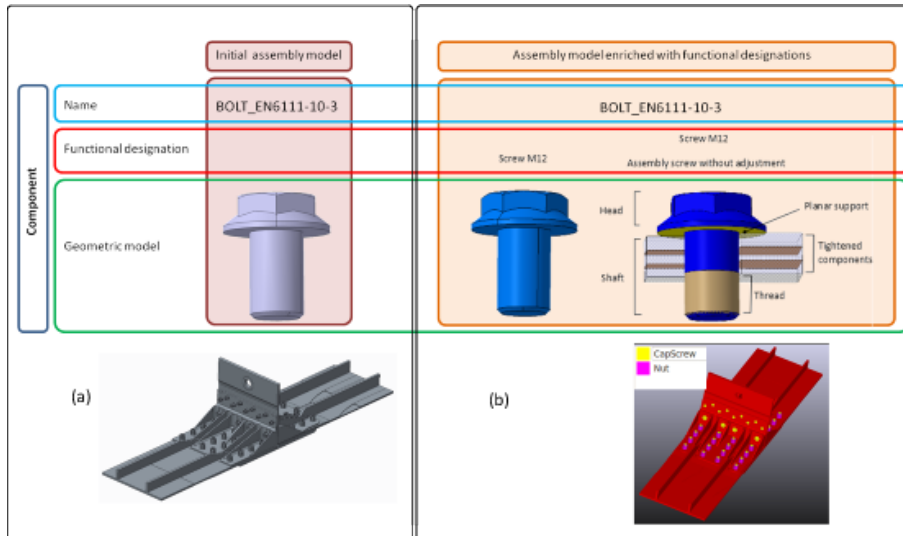


Figure 13. An example of assembly before (a) and after (b) the extraction of functional information. The upper part shows the influence of the extraction process on the structure of components' geometric models. The lower part illustrates the extraction process applied to a mechanical assembly.

6.4.6. Anatomical models

Participants: Ali-Hamadi Dicko, Olivier Palombi, François Faure.

We continue the development and the exploitation of MyCF, our ontology-centered anatomical knowledge base, in collaboration with the Grenoble University Hospital, and the DEMAR team in Montpellier (Benjamin Gilles).

We have presented a novel pipeline for the construction of biomechanical simulations by combining generic anatomical knowledge with specific data [27], [21], as illustrated in Figure 14. Based on functional descriptors supplied by the user, the list of the involved anatomical entities (currently bones and muscles) is generated using formal knowledge stored in ontologies, as well as a physical model based on reference geometry and mechanical parameters. This simulation-ready model can then be registered to subject-specific geometry to perform customized simulations. The user can provide additional specific geometry, such as a simulation mesh, to assemble with the reference geometry. Subject-specific information can also be used to individualize each functional model. The model can then be visualized and animated. This pipeline dramatically eases the creation of biomechanical models.

6.4.7. Managing morphological and functional information of the human body

Participants: Olivier Palombi, Ali-Hamadi Dicko, François Faure, Jean-Claude Léon, Ahmad Shahwan.

My Corporis Fabrica (MyCF) is an anatomical knowledge database. During 2012, we have linked functional entities defined in MyCF to the involved anatomical structures. The scope has been limited to the musculo-skeletal system. Based on this brain new formal description of the functional anatomy of limbs, we present a novel pipeline for the construction of biomechanical simulations by combining generic anatomical knowledge with specific data. This pipeline dramatically eases the creation of biomechanical models [27].

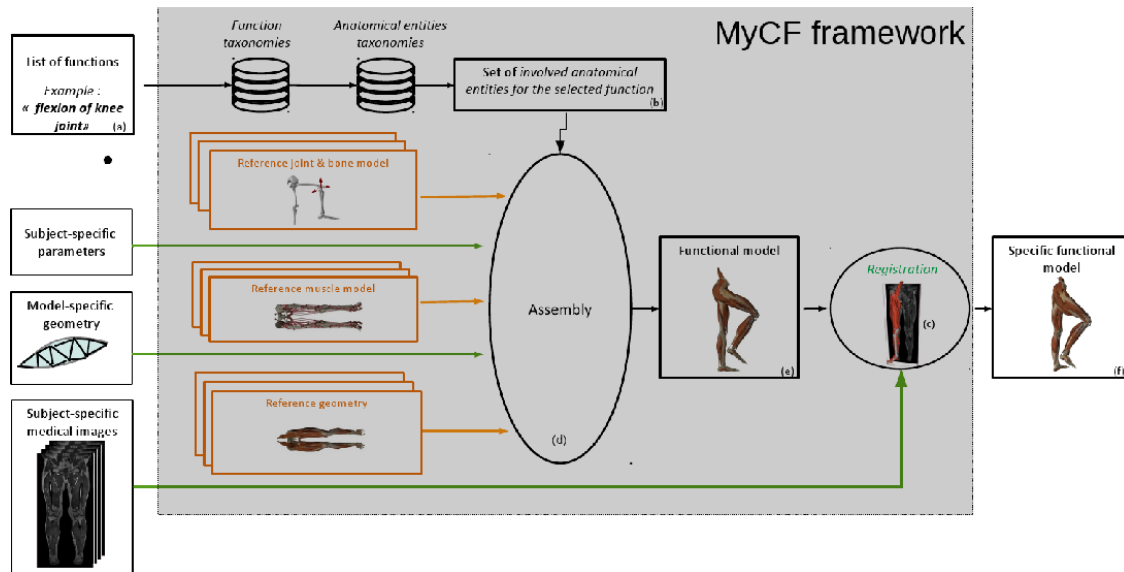


Figure 14. An overview of our modeling framework. On the left, the user input is a list of functions to simulate, optionally complemented with specific data. On the right, the output is a mechanical model ready for simulation. The modeling pipeline uses symbolic knowledge to select anatomical entities to assemble. The final model can be composed of a mix of reference and specific parameters and geometry.

MyCF-Browser which is the GUI of MyCF has been completely reviewed and rewritten. The MyCf's style software architecture is REST (Representational State Transfer) that has emerged as a predominant Web service design model. The anatomical knowledge is now available through a WEB service. The next step is to write a full web MyCF-Browser. MyCF browser is now available on line: <http://www.mycorporisfabrica.org/>. The MyCf's generic programming framework can be used for other domains. The link with semantic and 3D models matches research activities of IMAGINE towards interactive digital creation media. Anatomy can be seen as a study case.

6.5. Creating and interacting with virtual prototypes

Scientist in charge: Jean-Claude Léon

Other permanent researchers: Marie-Paule Cani, Rémi Ronfard, Olivier Palombi

6.5.1. Space deformations

Participant: Stefanie Hahmann.

Free Form Deformation (FFD) is a well-established technique for deforming arbitrary object shapes in space. Although more recent deformation techniques have been introduced, amongst them skeleton-based deformation and cage based deformation, the simple and versatile nature of FFD is a strong advantage, and justifies its presence in nowadays leading commercial geometric modeling and animation software systems. Several authors have addressed the problem of volume preserving FFD. These previous approaches however make either use of expensive non-linear optimization techniques, or resort to first order approximation suitable only for small-scale deformations. Our approach was to take advantage from the multi-linear nature of the volume constraint in order to derive a simple, exact and explicit solution to the problem of volume preserving FFD. Two variants of the algorithm have been developed, without and with direct shape manipulation.

Moreover, we showed that the linearity of our solution enables to implement it efficiently on GPU. This work has been done in collaboration with Gershon Elber from TECHNION, Hans Hagen from TU Kaiserslautern, Georges-Pierre Bonneau and Sébastien Barbier from Maverick Inria. It has been published in the journal *The Visual Computer* [5].



Figure 15. Comparison between standard FFD deformation (middle) and our method preserving the volume (right) from an initial rest shape (left).

Within Lucian Stanculescu PhD, we developed a mesh structure that dynamically adapts to the deformation defined by the user. Thanks to the quasi-uniform property of the mesh, it can be locally extended by any arbitrary deformation, and the mesh can also handle changes of topologies to be used as a virtual sculpting tool. This year we extend this work to handle local features such as sharp edges. In defining features (points or curves) over the surface we can interactively define meaningful regions limiting the influence of the deformation tools, or to ease artistic decoration mapping such as textures or extra geometric layers. We aim to generate a new tool enabling to sculpt objects which blend between organic to CAD-style appearance.

6.5.2. Procedural modeling of terrains and cities

Participants: Adrien Bernhard, Marie-Paule Cani, Arnaud Emilien.

Within the PhD of Adrien Bernhard we introduced a real-time terrain modeling tool using a fast GPU-based terrain solver with a lightweight CPU-based data structure.

We then work on adding roads and settlements on this terrain within the PhD of Arnaud Emilien. We focused on the modeling of small, European villages that took benefit of terrain features to settle in safe, sunny or simply convenient places. We introduced a three step procedural method [3] for generating scattered settlements on arbitrary terrains, enabling villages and hamlets, with the associated roads, forests and fields to be built on arbitrary landscapes.



Figure 16. Fortified village at the top of a cliff, using a war-time growth scenario followed by farming style settlement.

6.5.3. Hand Navigator

Participant: Jean-Claude Léon.

The different deformation models we developed in the past few years open the problem of providing intuitive interaction tools for specifying the desired deformations in real-time. Therefore, work has focused on developing new devices to investigate interactions incorporating a rather large number of parameters. For the past three years, we focused on developing a peripheral device similar to a mouse, called the HandNavigator, enabling to control simultaneously ten or more degrees of freedom of a virtual hand. This device developed in collaboration with Jean-Rémy Chardonnet (Inst. Image, Arts et Métiers ParisTech) consists in a 3D mouse for the position and orientation of the hand in 3D space, enhanced with many sensors for moving and monitoring the virtual fingers. Thanks to a pre-industrialization project funded by the incubator GRAVIT, the first prototype, patented by Inria, has been extended with the incorporation of new sensors and new shapes to improve the device efficiency and evolve toward a passive haptic device (see Figure 17). An extension of the patent and a partnership with HAPTION company are new steps toward the industrialization of this device. The partnership with HAPTION focuses on grasping actions to use the Hand Navigator as a complement to their haptic feedback device. Publications took place after setting up the patent extension [12], [13]. The ongoing BQR INTUACTIVE funded by Grenoble-INP will lead to further scientific topics regarding interactions during grasping as well as with deformable bodies and a partnership is ongoing with GIPSA-Lab to study the muscular activity during interactions. A specific experiment has been set up to study the user's muscles activity.

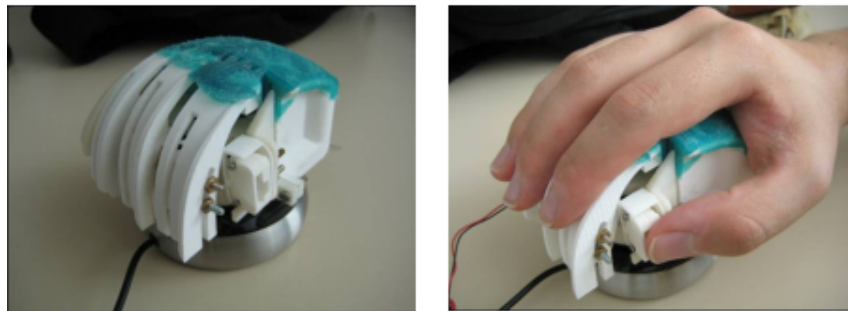


Figure 17. Current version of the HandNavigator prototype with three sensors per finger and a vibration damping structure.

LEAR Project-Team

6. New Results

6.1. Visual recognition in images

6.1.1. Correlation-Based Burstiness for Logo Retrieval

Participants: Matthijs Douze, Jerome Revaud, Cordelia Schmid.

Detecting logos in photos is challenging. A reason is that logos locally resemble patterns frequently seen in random images. In [21] we propose to learn a statistical model for the distribution of incorrect detections output by an image matching algorithm. It results in a novel scoring criterion in which the weight of correlated keypoint matches is reduced, penalizing irrelevant logo detections. In experiments on two very different logo retrieval benchmarks, our approach largely improves over the standard matching criterion as well as other state-of-the-art approaches.



Figure 1. Illustration of a logo detected by our method.

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

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

In [19] we propose a benchmark of several objective functions for large-scale image classification: we compare the one-vs-rest, multiclass, ranking and weighted average ranking SVMs. Using stochastic gradient descent optimization, we can scale the learning to millions of images and thousands of classes. Our experimental evaluation shows that ranking based algorithms do not outperform a one-vs-rest strategy and that the gap between the different algorithms reduces in case of high-dimensional data. We also show that for one-vs-rest, learning through cross-validation the optimal degree of imbalance between the positive and the negative samples can have a significant impact. Furthermore, early stopping can be used as an effective regularization strategy when training with stochastic gradient algorithms. Following these “good practices”, we were able to improve the state-of-the-art on a large subset of 10K classes and 9M of images of ImageNet from 16.7% accuracy to 19.1%. Some qualitative results can be seen in Figure 2 .

6.1.3. Discriminative Spatial Saliency for Image Classification

Participants: Frédéric Jurie [Université de Caen], Cordelia Schmid, Gaurav Sharma.

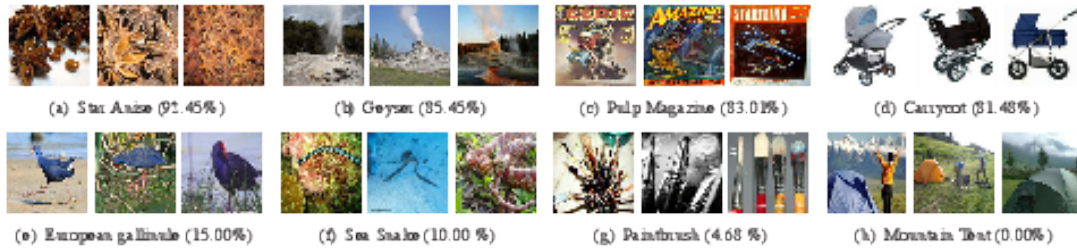


Figure 2. ImageNet10K results (top-1 accuracy in %) obtained with w -OVR and 130K-dim Fisher vectors. (a-d) Sample classes among the best performing ones. (e-h) Sample classes among the worst performing ones.

In many visual classification tasks the spatial distribution of discriminative information is (i) non uniform e.g. “person reading” can be distinguished from “taking a photo” based on the area around the arms i.e. ignoring the legs, and (ii) has intra class variations e.g. different readers may hold the books differently. Motivated by these observations, we propose in [22] to learn the discriminative spatial saliency of images while simultaneously learning a max-margin classifier for a given visual classification task. Using the saliency maps to weight the corresponding visual features improves the discriminative power of the image representation. We treat the saliency maps as latent variables and allow them to adapt to the image content to maximize the classification score, while regularizing the change in the saliency maps. See Figure 3 for an illustration. Our experimental results on three challenging datasets, for (i) human action classification, (ii) fine grained classification, and (iii) scene classification, demonstrate the effectiveness and wide applicability of the method.

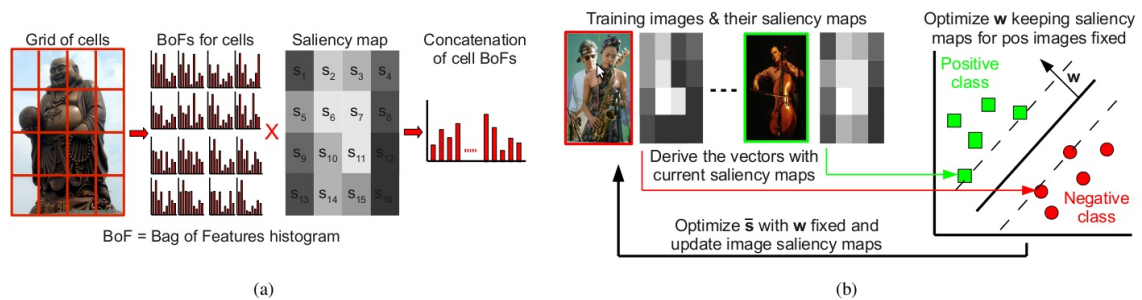


Figure 3. (a) The images are represented by concatenation of cell bag-of-features weighted by the image saliency maps. (b) We propose to use a block coordinate descent algorithm for learning our model. As in a latent SVM, we optimize in one step the weight vector w keeping the saliency maps of the positive images fixed, and in the other step we optimize the saliency keeping w fixed.

6.1.4. Tree-structured CRF Models for Interactive Image Labeling

Participants: Gabriela Csurka [XRCE], Thomas Mensink, Jakob Verbeek.

In [8] we propose structured prediction models for image labeling that explicitly take into account dependencies among image labels. In our tree structured models, image labels are nodes, and edges encode dependency relations. To allow for more complex dependencies, we combine labels in a single node, and use mixtures of

trees. Our models are more expressive than independent predictors, and lead to more accurate label predictions. The gain becomes more significant in an interactive scenario where a user provides the value of some of the image labels at test time. Such an interactive scenario offers an interesting trade-off between label accuracy and manual labeling effort. The structured models are used to decide which labels should be set by the user, and transfer the user input to more accurate predictions on other image labels. We also apply our models to attribute-based image classification, where attribute predictions of a test image are mapped to class probabilities by means of a given attribute-class mapping. Experimental results on three publicly available benchmark data sets show that in all scenarios our structured models lead to more accurate predictions, and leverage user input much more effectively than state-of-the-art independent models.

6.1.5. Metric Learning for Large Scale Image Classification: Generalizing to new classes at near-zero cost

Participants: Gabriela Csurka [XRCE], Thomas Mensink, Florent Perronnin [XRCE], Jakob Verbeek.

In [18], [27] we consider the task of large scale image classification in open ended datasets. Many real-life datasets are open-ended and dynamic: new images are continuously added to existing classes, new classes appear over time and the semantics of existing classes might evolve too. In order to be able to handle new images and new classes at near-zero cost we consider two distance based classifiers, the k-nearest neighbor (k-NN) and nearest class mean (NCM) classifiers. For the NCM classifier we introduce a new metric learning approach, which has advantageous properties over the classical Fisher Discriminant Analysis. We also introduce an extension of the NCM classifier to allow for richer class representations, using multiple centroids per class. Experiments on the ImageNet 2010 challenge dataset, which contains over one million training images of thousand classes, show that, surprisingly, the NCM classifier compares favorably to the more flexible k-NN classifier. Moreover, the NCM performance is comparable to that of linear SVMs which obtain current state-of-the-art performance. Experimentally we study the generalization performance to classes that were not used to learn the metrics. Using a metric learned on 1,000 classes, we show results for the ImageNet-10K dataset which contains 10,000 classes, and obtain performance that is competitive with the current state-of-the-art, while being orders of magnitude faster. Furthermore, we show how a zero-shot class prior based on the ImageNet hierarchy can improve performance when few training images are available. See Figure 4 for an illustration.

6.2. Learning and statistical models

6.2.1. Image categorization using Fisher kernels of non-iid image models

Participants: Ramazan Cinbis, Cordelia Schmid, Jakob Verbeek.

Bag of visual words treat images as an orderless sets of local regions and represent them by visual word frequency histograms. Implicitly, regions are assumed to be identically and independently distributed (iid), which is a very poor assumption from a modeling perspective; see Figure 5 for an illustration. In [13], we introduce non-iid models by treating the parameters of bag-of-word models as latent variables which are integrated out, rendering all local regions dependent. Using the Fisher kernel we encode an image by the gradient of the data log-likelihood with respect to hyper-parameters that control priors on the model parameters. In fact, our models naturally generate transformations similar to taking square-roots, providing an explanation of why such non-linear transformations have proven successful. Using variational inference we extend the basic model to include Gaussian mixtures over local descriptors, and latent topic models to capture the co-occurrence structure of visual words, both improving performance. Our models yields state-of-the-art image categorization performance using linear classifiers, without using non-linear kernels, or (approximate) explicit embeddings thereof, e.g. by taking the square-root of the features.

6.2.2. Conditional gradient algorithms for machine learning

Participants: Zaid Harchaoui, Anatoli Juditsky [UJF], Arkadi Nemirovski [Georgia Tech].


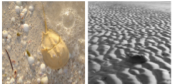







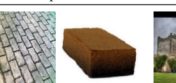










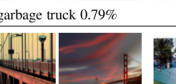


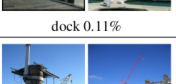






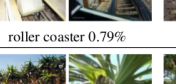


 <p>Cliff dwelling L2 11.0% - Mah. 99.9%</p>	 <p>horseshoe crab 0.99%</p>	 <p>African elephant 0.99%</p>	 <p>mongoose 0.94%</p>	 <p>Indian elephant 0.88%</p>	 <p>dingo 0.87%</p>	L2
	 <p>cliff 0.07%</p>	 <p>dam 0.00%</p>	 <p>stone wall 0.00%</p>	 <p>brick 0.00%</p>	 <p>castle 0.00%</p>	Mah.
 <p>Gondola L2 4.4% - Mah. 99.7%</p>	 <p>shopping cart 1.07%</p>	 <p>unicycle 0.84%</p>	 <p>covered wagon 0.83%</p>	 <p>garbage truck 0.79%</p>	 <p>forklift 0.78%</p>	L2
	 <p>dock 0.11%</p>	 <p>canoe 0.03%</p>	 <p>fishing rod 0.01%</p>	 <p>bridge 0.01%</p>	 <p>boathouse 0.01%</p>	Mah.
 <p>Palm L2 6.4% - Mah. 98.1%</p>	 <p>crane 0.87%</p>	 <p>stupa 0.83%</p>	 <p>roller coaster 0.79%</p>	 <p>bell cote 0.78%</p>	 <p>flagpole 0.75%</p>	L2
	 <p>cabbage tree 0.81%</p>	 <p>pine 0.30%</p>	 <p>pandanus 0.14%</p>	 <p>iron tree 0.07%</p>	 <p>logwood 0.06%</p>	Mah.

Figure 4. Examples of three classes, and the five most similar classes for each according to the standard ℓ_2 metric and our learned Mahalanobis metric.

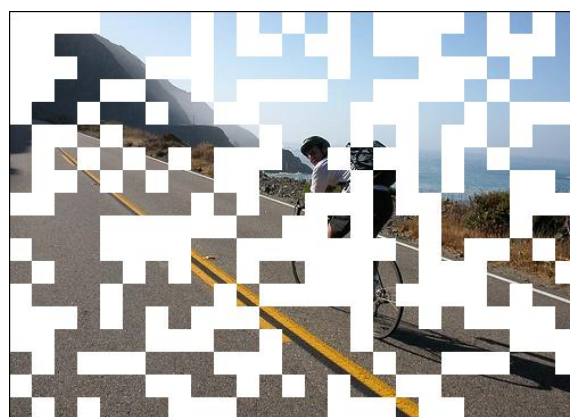


Figure 5. Illustration of why local image patches are not independent: we can easily guess the image content in the masked areas.

In [17] we consider convex optimization problems arising in machine learning in high-dimensional settings. For several important learning problems, such as e.g. noisy matrix completion, state-of-the-art optimization approaches such as composite minimization algorithms are difficult to apply and do not scale up to large datasets. We study three conditional gradient-type algorithms, suitable for large-scale problems, and derive their finite-time convergence guarantees. Promising experimental results are presented on two large-scale real-world datasets.

6.2.3. *Large-scale classification with trace-norm regularization*

Participants: Matthijs Douze, Miro Dudik [Microsoft Research], Zaid Harchaoui, Jérôme Malick [BiPoP Team Inria Grenoble], Mattis Paulin [ETHZ].

In [16] we introduce a new scalable learning algorithm for large-scale multi-class image classification, based on the multinomial logistic loss and the trace-norm regularization penalty. Reframing the challenging non-smooth optimization problem into a surrogate infinite-dimensional optimization problem with a regular ℓ_1 -regularization penalty, we propose a simple and provably efficient accelerated coordinate descent algorithm. Furthermore, we show how to perform efficient matrix computations in the compressed domain for quantized dense visual features, scaling up to 100,000s examples, 1,000s-dimensional features, and 100s of categories. Promising experimental results on the “Fungus”, “Ungulate”, and “Vehicles” subsets of ImageNet are presented, where we show that our approach performs significantly better than state-of-the-art approaches for Fisher vectors with 16 Gaussians.

6.2.4. *Tree-walk kernels for computer vision*

Participants: Francis Bach [Inria SIERRA team], Zaid Harchaoui.

In [25] we propose a family of positive-definite kernels between images, allowing to compute image similarity measures respectively in terms of color and of shape. The kernels consists in matching subtree-patterns called “tree-walks” of graphs extracted from the images, e.g. the segmentation graphs for color similarity and graphs of the discretized shapes or the point clouds in general for shape similarity. In both cases, we are able to design computationally efficient kernels which can be computed in polynomial-time in the size of the graphs, by leveraging specific properties of the graphs at hand such as planarity for segmentation graphs or factorizability of the associated graphical model for point clouds. Our kernels can be used by any kernel-based learning method, and hence we present experimental results for supervised and semi-supervised classification as well as clustering of natural images and supervised classification of handwritten digits and Chinese characters from few training examples.

6.2.5. *Lifted coordinate descent for learning with trace-norm regularization*

Participants: Miro Dudik [Microsoft Research], Zaid Harchaoui, Jérôme Malick [BiPoP Team Inria Grenoble].

In [14] we consider the minimization of a smooth loss with trace-norm regularization, which is a natural objective in multi-class and multi-task learning. Even though the problem is convex, existing approaches rely on optimizing a non-convex variational bound, which is not guaranteed to converge, or repeatedly perform singular-value decomposition, which prevents scaling beyond moderate matrix sizes. We lift the non-smooth convex problem into an infinitely dimensional smooth problem and apply coordinate descent to solve it. We prove that our approach converges to the optimum, and is competitive or outperforms the state of the art.

6.3. Recognition in video

6.3.1. *Large-scale multi-media event detection in video*

Participants: Matthijs Douze, Zaid Harchaoui, Dan Oneata, Danila Potapov, Jerome Revaud, Cordelia Schmid, Jochen Schwenninger [Fraunhofer Institute, Bonn], Jakob Verbeek, Heng Wang.

This year we participated in the TrecVid Multimedia Event Detection (MED) task. The goal is to detect event categories (such as “birthday party”, or “changing a vehicle tire”) in a large collection of around 100,000 videos with a total duration of around 4,000 hours. To this end we implemented an efficient system based on our recently developed MBH video descriptor (see Section 5.4), SIFT descriptors and, MFCC audio descriptors (contributed by Fraunhofer Institute). All these low-level descriptors are encoded using the Fisher vector representation (see Section 5.3). In addition we implemented an optical character recognition (OCR) system to extract textual features from the video. The system is described in a forthcoming paper [31], and ranked first and second in two evaluations among the 17 systems submitted by different international teams participating to the task. See Figure 6 for an illustration.



Figure 6. Illustration of videos retrieved for two event categories. From left to right, we show for each a frame from (i) the top ranked video, (ii,iii) the first negative video, and the positive just before, and (iv) the last positive video.

6.3.2. Learning Object Class Detectors from Weakly Annotated Video

Participants: Javier Civera, Vittorio Ferrari, Christian Leistner, Alessandro Prest, Cordelia Schmid.

Object detectors are typically trained on a large set of still images annotated by bounding-boxes. In [20] we introduce an approach for learning object detectors from real-world web videos known only to contain objects of a target class. We propose a fully automatic pipeline that localizes objects in a set of videos of the class and learns a detector for it. The approach extracts candidate spatio-temporal tubes based on motion segmentation and then selects one tube per video jointly over all videos. See Figure 7 for an illustration. To compare to the state of the art, we test our detector on still images, i.e., Pascal VOC 2007. We observe that frames extracted from web videos can differ significantly in terms of quality to still images taken by a good camera. Thus, we formulate the learning from videos as a domain adaptation task. We show that training from a combination of weakly annotated videos and fully annotated still images using domain adaptation improves the performance of a detector trained from still images alone.

6.3.3. Recognizing activities with cluster-trees of tracklets

Participants: Adrien Gaidon, Zaid Harchaoui, Cordelia Schmid.

In [15] we address the problem of recognizing complex activities, such as pole vaulting, which are characterized by the composition of a large and variable number of different spatio-temporal parts. We represent a video as a hierarchy of mid-level motion components. This hierarchy is a data-driven decomposition specific to each video. We introduce a divisive clustering algorithm that can efficiently extract a hierarchy over a large

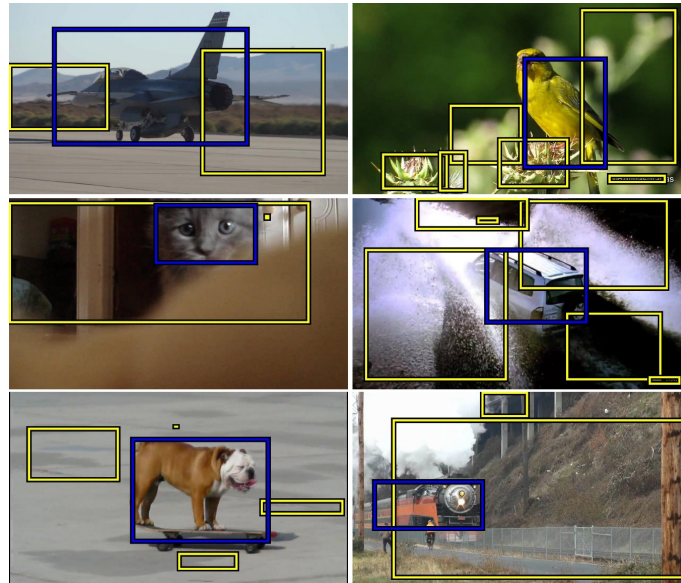


Figure 7. Yellow boxes represent tubes extracted by our method on the YouTube-Objects dataset. Blue boxes indicate the automatically selected tubes.

number of local trajectories. We use this structure to represent a video as an unordered binary tree. This tree is modeled by nested histograms of local motion features, see Figure 8 . We provide an efficient positive definite kernel that computes the structural and visual similarity of two tree decompositions by relying on models of their edges. Contrary to most approaches based on action decompositions, we propose to use the full hierarchical action structure instead of selecting a small fixed number of parts. We present experimental results on two recent challenging benchmarks that focus on complex activities and show that our kernel on per-video hierarchies allows to efficiently discriminate between complex activities sharing common action parts. Our approach improves over the state of the art, including unstructured activity models, baselines using other motion decomposition algorithms, graph matching, and latent models explicitly selecting a fixed number of parts.

6.3.4. Action Detection with Actom Sequence Models

Participants: Adrien Gaidon, Zaid Harchaoui, Cordelia Schmid.

We address the problem of detecting actions, such as drinking or opening a door, in hours of challenging video data. In [26] 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 the temporal structure of actions as a sequence of histograms of actom-anchored visual features, see Figure 9 for an illustration. Our representation, which can be seen as a temporally structured extension of the bag-of-features, is flexible, sparse, and discriminative. Training requires the annotation of actoms for action examples. At test time, actoms are detected 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 temporal action detection: "Coffee and Cigarettes" and the "DLSB" dataset. We also adapt our approach to a classification by detection 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 detection, as well as baselines that detect actions with a sliding window method combined with bag-of-features.

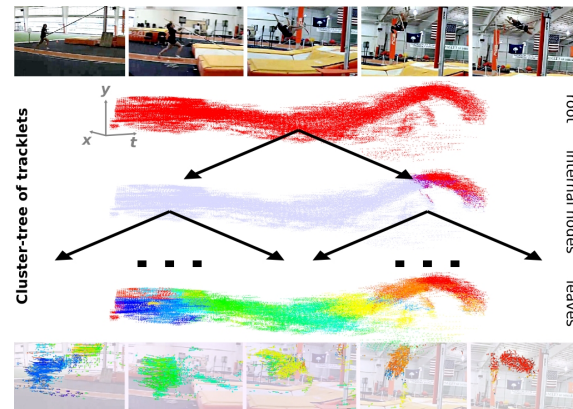


Figure 8. Illustration of tracklets found in a video and their hierarchical decomposition.

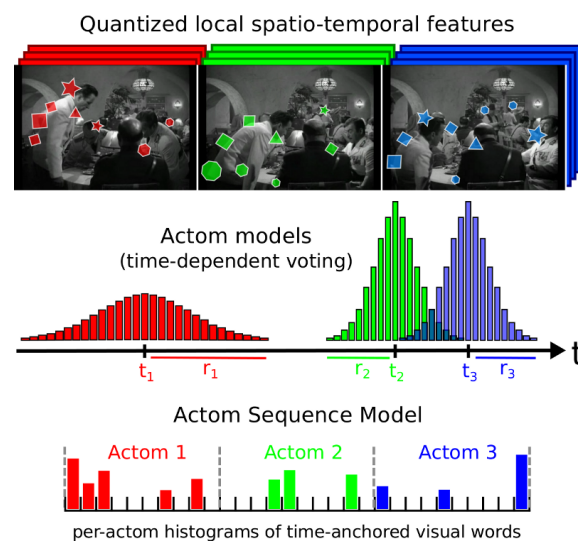


Figure 9. Illustration of the "Actom" video representation, see text for details.

6.3.5. Action recognition by dense trajectories

Participants: Alexander Kläser, Cheng-Lin Liu [Chinese Academy of Sciences], Cordelia Schmid, Heng Wang [Chinese Academy of Sciences].

In [28] we introduce 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 10), 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.

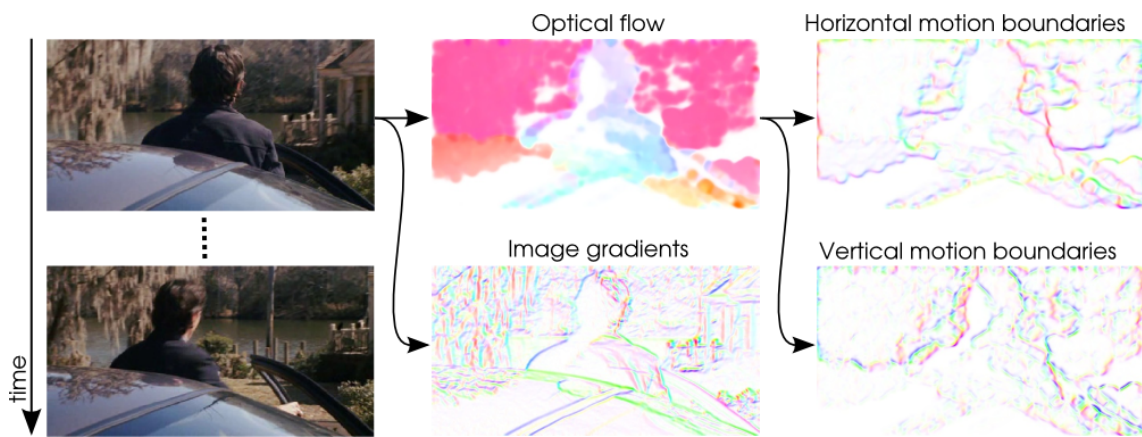


Figure 10. Illustration of the information captured by HOG, HOF, and MBH descriptors. Gradient/flow orientation is indicated by color (hue) and magnitude by saturation. The optical flow (top, middle) shows constant motion in the background, which is due to the camera movements. The motion boundaries (right) encode the relative motion between the person and the background.

MAVERICK Team

6. New Results

6.1. Computer visualization

6.1.1. Immersive Virtual Environment for Visuo-Vestibular Therapy: Preliminary Results

Participants: Jean-Dominique Gascuel, Henri Payno, Sébastien Schmerber, Olivier Martin.

The sense of equilibrium aggregates several interacting cues. On patients with vestibular loss, vision plays a major role. In this study, the goal is to propose a new immersive therapy based on 3D opto-kinetic stimulation. We propose to demonstrate that 3D monoscopic optical flows are an efficient tool to stimulate adaptive postural adjustment. We developed an immersive therapeutic platform that enables to tune the balance task difficulty by managing optic flow speed and gaze anchoring (Figure 5). **METHODOLOGY:** the immersive sessions proposed to vestibular areflexic patients are composed of a repetition of dynamic optic flows, with varying speed and presence or not of a gaze anchor. The balance adjustments are recorded by a force plate, and quantified by the length of the center of pressure trajectory. **RESULTS:** Preliminary analysis shows that (i) Patients report a strong immersion feeling in the motion flow, triggering more intense motor response to "fight against fall" than in standard opto-kinetic protocols; (ii) An ANOVA factorial design shows a significant effect of flow speed, session number and gaze anchor impact. **CONCLUSION:** This study shows that 3D immersive stimulation removes essential limits of traditional opto-kinetic stimulators (limited 2D motions and remaining fixed background cues). Moreover, the immersive optic flow stimulation is an efficient tool to induce balance adaptive reactions in vestibular patients. Hence, such a platform appears to be a powerful therapeutic tool for training and relearning of balance control processes.

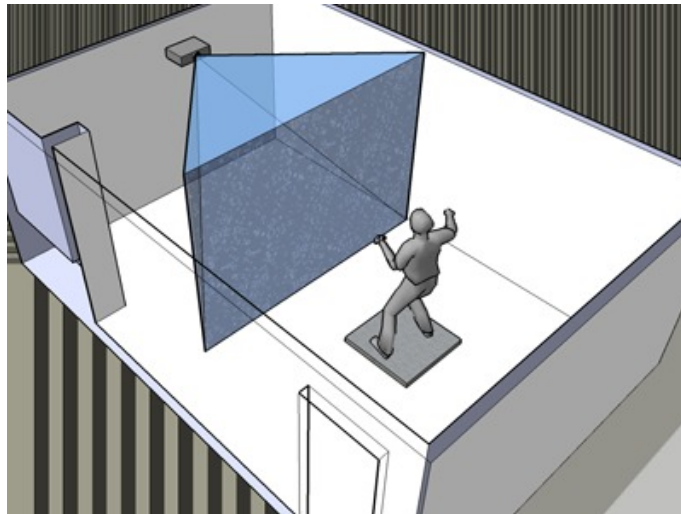


Figure 5. The immersive platform, installed in an available room of the hospital. The large retro projected screen is at 60 cm of the patient, covering most of its visual field. The patient is standing on a force plate, recording CoP.

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

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

we present a user study on the use of Depth of Field for depth perception in Direct Volume Rendering (Figure 6). 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.

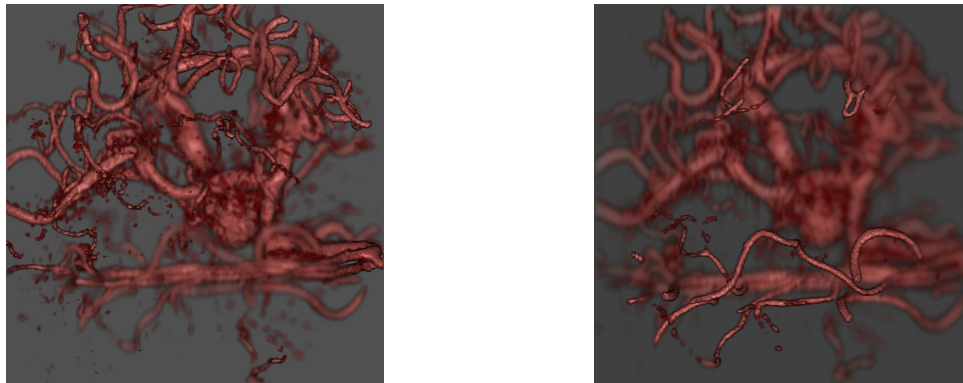


Figure 6. Aneurism. Depth of Field reinforces the perception of which feature is in front of the other.

6.1.3. Volume Preserving FFD for Programmable Graphics Hardware

Participants: Stefanie Hahmann, Georges-Pierre Bonneau, Sébastien Barbier, Gershon Elber, Hans Hagen.

Free Form Deformation (FFD) is a well established technique for deforming arbitrary object shapes in space. Although more recent deformation techniques have been introduced, amongst them skeleton-based deformation and cage based deformation, the simple and versatile nature of FFD is a strong advantage, and justifies its presence in nowadays leading commercial geometric modeling and animation software systems. Since its introduction in the late 80's, many improvements have been proposed to the FFD paradigm, including control lattices of arbitrary topology, direct shape manipulation and GPU implementation. Several authors have addressed the problem of volume preserving FFD. These previous approaches either make use of expensive non-linear optimization techniques, or resort to first order approximation suitable only for small-scale deformations. In this paper we take advantage from the multi-linear nature of the volume constraint in order to derive a simple, exact and explicit solution to the problem of volume preserving FFD. Two variants of the algorithm are given, without and with direct shape manipulation. Moreover, the linearity of our solution enables to implement it efficiently on GPU (Figure 7).

6.1.4. Sharp feature preserving MLS surface reconstruction based on local feature line approximations

Participants: Christopher Weber, Stefanie Hahmann, Hans Hagen, Georges-Pierre Bonneau.

Sharp features in manufactured and designed objects require particular attention when reconstructing surfaces from unorganized scan point sets using moving least squares (MLS) fitting. It's an inherent property of MLS fitting that sharp features are smoothed out. Instead of searching for appropriate new fitting functions our approach computes a modified local point neighborhood so that a standard MLS fitting can be applied

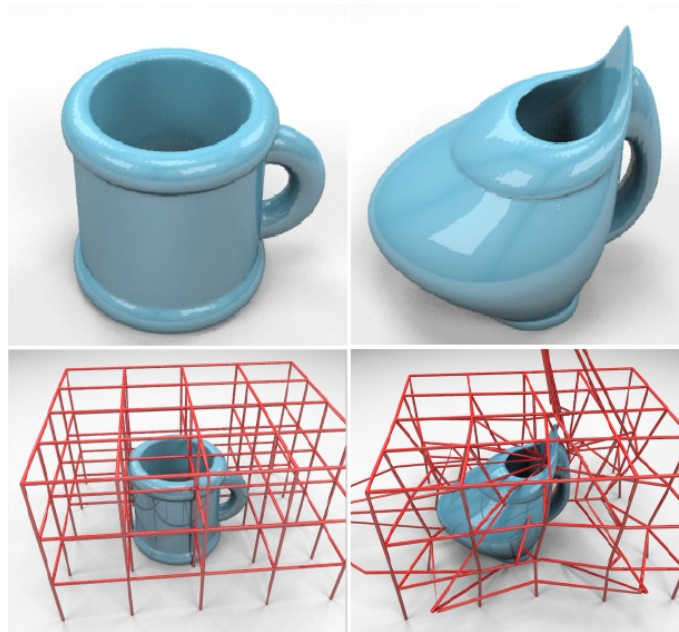


Figure 7. Sculptured cup with volume preservation.

enhanced by sharp features reconstruction. We present a two-stage algorithm. In a pre-processing step sharp feature points are marked first. This algorithm is robust to noise since it is based on Gauss map clustering. In the main phase, the selected feature points are used to locally approximate the feature curve and to segment and enhance the local point neighborhood. The MLS projection thus leads to a piecewise smooth surface preserving all sharp features. The method is simple to implement and able to preserve line-type features as well as corner-type features during reconstruction (Figure 8).

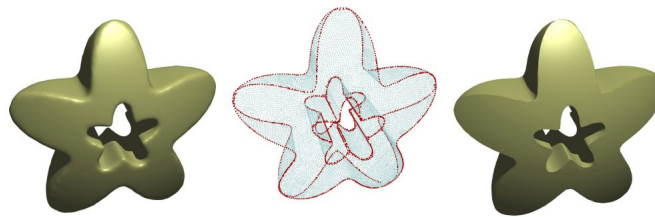


Figure 8. Left: standard MLS surface. Middle: feature point detection in point cloud. Right: sharp feature preserving MLS.

6.2. Expressive rendering

6.2.1. Active Strokes: Coherent Line Stylization for Animated 3D Models

Participants: Pierre Bénard, Lu Jingwan, Forrester Cole, Adam Finkelstein, Joëlle Thollot.

We present a method for creating coherently animated line drawings that include strong abstraction and stylization effects (Figure 9). These effects are achieved with active strokes: 2D contours that approximate and track the lines of an animated 3D scene. Active strokes perform two functions: they connect and smooth unorganized line samples, and they carry coherent parameterization to support stylized rendering. Line samples are approximated and tracked using active contours ("snakes") that automatically update their arrangement and topology to match the animation. Parameterization is maintained by brush paths that follow the snakes but are independent, permitting substantial shape abstraction without compromising fidelity in tracking. This approach renders complex models in a wide range of styles at interactive rates, making it suitable for applications like games and interactive illustrations.

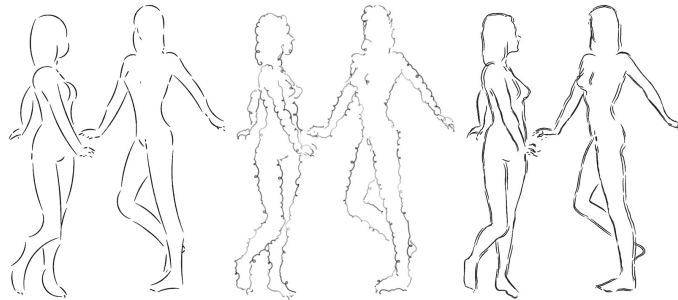


Figure 9. Stylization examples. Woman in two poses and three styles: arcs, loopy offsets, and overdrawn.

6.2.2. Temporally Coherent Video Stylization

Participants: Pierre Bénard, Joëlle Thollot, John Collomosse.

The transformation of video clips into stylized animations remains an active research topic in Computer Graphics. A key challenge is to reproduce the look of traditional artistic styles whilst minimizing distracting flickering and sliding artifacts; i.e. with temporal coherence. This chapter surveys the spectrum of available video stylization techniques, focusing on algorithms encouraging the temporally coherent placement of rendering marks, and discusses the trade-offs necessary to achieve coherence. We begin with flow-based adaptations of stroke based rendering (SBR) and texture advection capable of painting video. We then chart the development of the field, and its fusion with Computer Vision, to deliver coherent mid-level scene representations. These representations enable the rotoscoping of rendering marks on to temporally coherent video regions, enhancing the diversity and temporal coherence of stylization. In discussing coherence, we formalize the problem of temporal coherence in terms of three defined criteria, and compare and contrast video stylization using these.

6.3. Illumination simulation

6.3.1. Accurate fitting of measured reflectances using a Shifted Gamma micro-facet distribution

Participants: Mahdi M. Bagher, Cyril Soler, Nicolas Holzschuch.

Material models are essential to the production of photo-realistic images. Measured BRDFs provide accurate representation with complex visual appearance, but have larger storage cost. Analytical BRDFs such as Cook-Torrance provide a compact representation but fail to represent the effects we observe with measured appearance. Accurately fitting an analytical BRDF to measured data remains a challenging problem. In this paper we introduce the SGD micro-facet distribution for Cook-Torrance BRDF. This distribution accurately

models the behavior of most materials. As a consequence, we accurately represent all measured BRDFs using a single lobe. Our fitting procedure is stable and robust, and does not require manual tweaking of the parameters (Figure 10).

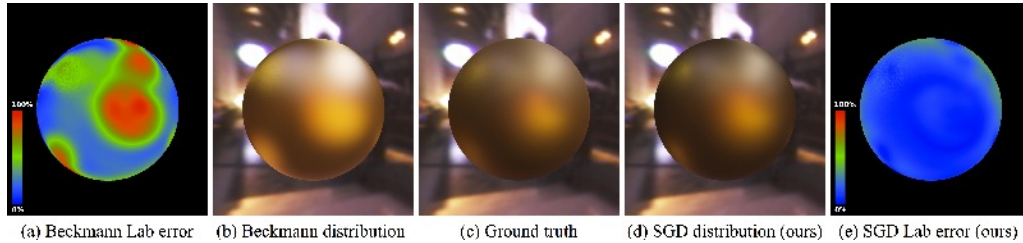


Figure 10. fitting of measured reflectances: comparison between ground truth and our approach.

6.3.2. Interactive rendering of acquired materials on dynamic geometry using bandwidth prediction

Participants: Mahdi M. Bagher, Cyril Soler, Kartic Subr, Laurent Belcour, Nicolas Holzschuch.

Shading complex materials such as acquired reflectances in multi-light environments is computationally expensive. Estimating the shading integral involves sampling the incident illumination independently at several pixels. The number of samples required for this integration varies across the image, depending on an intricate combination of several factors. Adaptively distributing computational budget across the pixels for shading is therefore a challenging problem. In this paper we depict complex materials such as acquired reflectances, interactively, without any precomputation based on geometry. We first estimate the approximate spatial and angular variation in the local light field arriving at each pixel. This *local bandwidth* 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 bandwidth information to adaptively sample for reconstruction and integration. For example, fewer pixels per area are shaded for pixels projecting onto diffuse objects, and fewer samples are used for integrating illumination incident on specular objects (Figure 11).

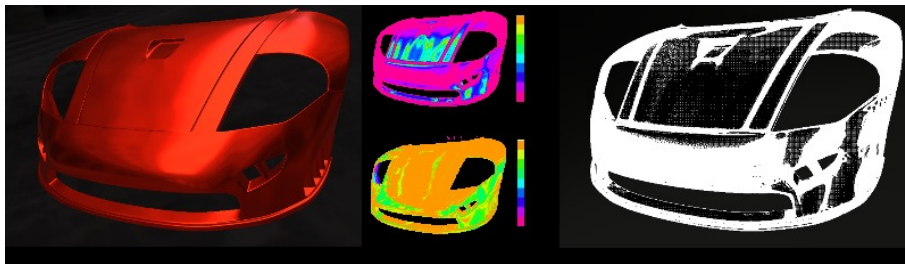


Figure 11. Interactive rendering of aquired materials. Center: predicted bandwidth and variance. Right: sample points where we compute illumination. Left: rendered result.

6.3.3. Real-Time Rendering of Rough Refraction

Participants: Charles De Rousiers, Adrien Bousseau, Kartic Subr, Nicolas Holzschuch, Ravi Ramamoorthi.

We present an algorithm to render objects made of transparent materials with rough surfaces in real-time, under all-frequency distant illumination (Figure 12). Rough surfaces cause wide scattering as light enters and exits objects, which significantly complicates the rendering of such materials. We present two contributions to approximate the successive scattering events at interfaces, due to rough refraction: First, an approximation of the Bidirectional Transmittance Distribution Function (BTDF), using spherical Gaussians, suitable for real-time estimation of environment lighting using pre-convolution; second, a combination of cone tracing and macro-geometry filtering to efficiently integrate the scattered rays at the exiting interface of the object. We demonstrate the quality of our approximation by comparison against stochastic ray-tracing. Furthermore we propose two extensions to our method for supporting spatially varying roughness on object surfaces and local lighting for thin objects.

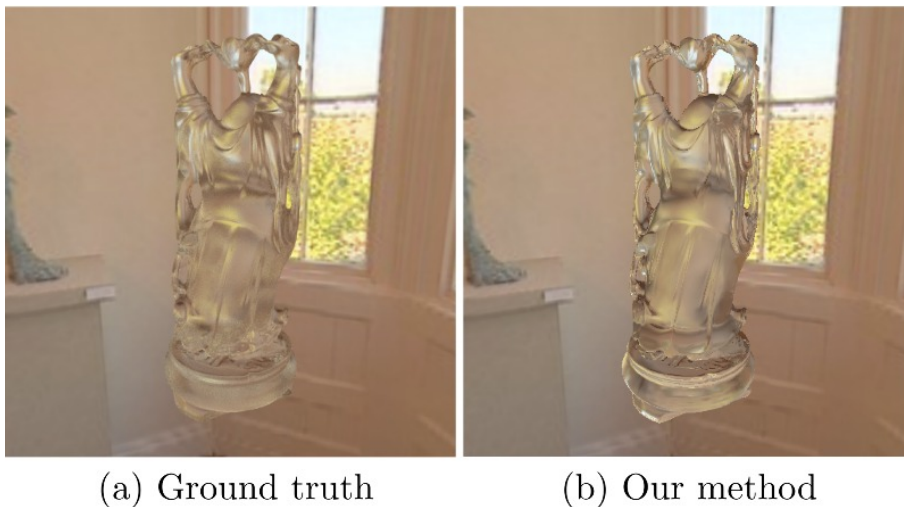


Figure 12. Real-Time Rendering of Rough Refraction

6.3.4. Multiple-scattering and double-scattering effects in translucent materials

Participants: Jean-Dominique Gascuel, Nicolas Holzschuch.

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 difficulty 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. In this paper, we show 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 for accelerating multiple scattering effects. We also provide a model for fast computation of double-scattering events, using a precomputed density function we store in a compact way (Figure 13).

6.3.5. Frequency analysis of participating media

Participants: Laurent Belcour, Cyril Soler, Kavita Bala.

Computing global illumination in participating media is frustratingly expensive: while the computation itself is long and complicated, the result involve very smooth regions of illumination. This motivates an a priori approach to find out how fast the resulting image will vary in space (i.e. it's spatial frequency) to adapt computation effort to reach the maximal efficiency. For this we are extending the theory of Fourier Analysis of

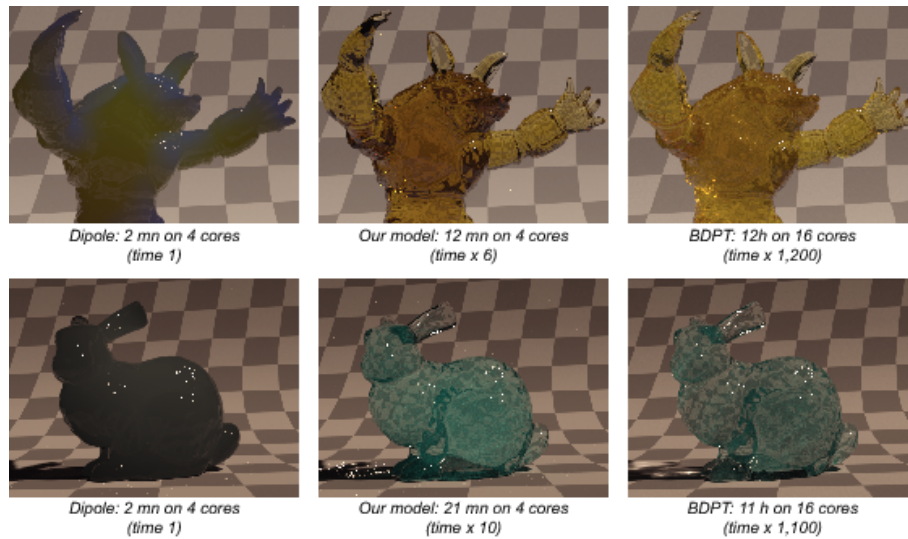


Figure 13. Multiple-scattering and double-scattering effects in translucent materials

light transport to participating media. Our work builds on the covariance analysis of light transport developed by Laurent Belcour in his PhD Thesis. It offers the possibility to drastically accelerate the algorithms involved in the computation of the illumination in scenes with participating media (Figure 14).

6.4. Complex scenes

6.4.1. A Survey of Non-linear Pre-filtering Methods for Efficient and Accurate Surface Shading

Participants: Eric Bruneton, Fabrice Neyret.

Rendering a complex surface accurately and without aliasing requires the evaluation of an integral for each pixel, namely a weighted average of the outgoing radiance over the pixel footprint on the surface. The outgoing radiance is itself given by a local illumination equation as a function of the incident radiance and of the surface properties. Computing all this numerically during rendering can be extremely costly. For efficiency, especially for real-time rendering, it is necessary to use precomputations. When the fine scale surface geometry, reflectance and illumination properties are specified with maps on a coarse mesh (such as color maps, normal maps, horizon maps or shadow maps), a frequently used simple idea is to pre-filter each map linearly and separately. The averaged outgoing radiance, i.e., the average of the values given by the local illumination equation is then estimated by applying this equation to the averaged surface parameters. But this is really not accurate because this equation is non-linear, due to self-occlusions, self-shadowing, non-linear reflectance functions, etc. Some methods use more complex pre-filtering algorithms to cope with these non-linear effects. This paper is a survey of these methods. We start with a general presentation of the problem of pre-filtering complex surfaces. We then present and classify the existing methods according to the approximations they make to tackle this difficult problem. Finally, an analysis of these methods allows us to highlight some generic tools to pre-filter maps used in non-linear functions, and to identify open issues to address the general problem.

6.4.2. Real-time Realistic Rendering and Lighting of Forests

Participants: Eric Bruneton, Fabrice Neyret.

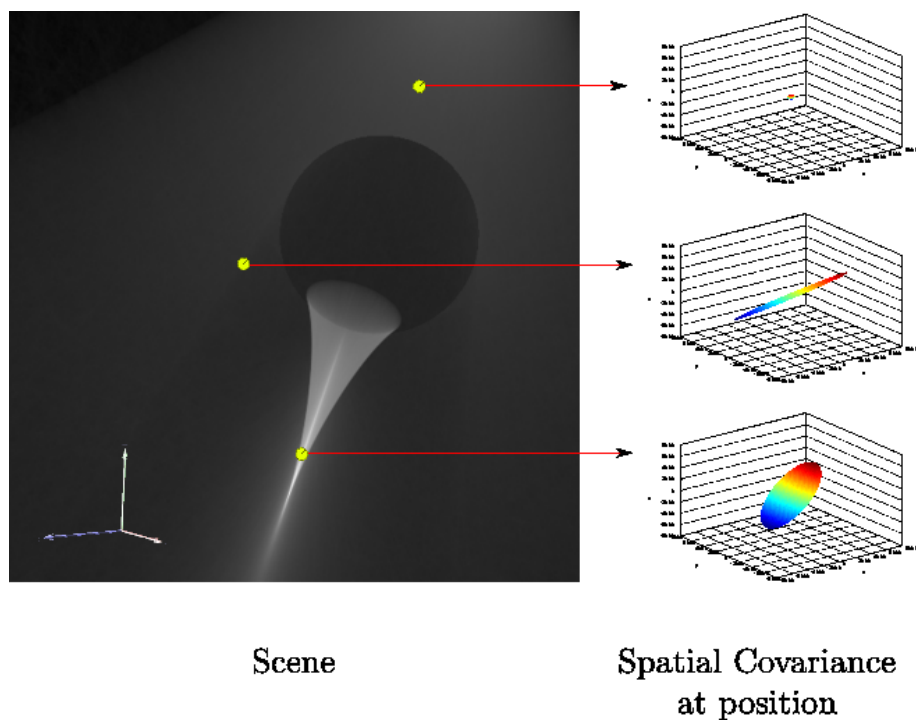


Figure 14. A glass sphere casting a volumetric caustic in participating media with multiple scattering. We can predict the covariance of the spectrum of the illumination locally everywhere in the volume so as to adapt the computation effort.

Realistic real-time rendering and lighting of forests is an important aspect for simulators and video games. This is a difficult problem, due to the massive amount of geometry: aerial forest views display millions of trees on a wide range of distances, from the camera to the horizon. Light interactions, whose effects are visible at all scales, are also a problem: sun and sky dome contributions, shadows between trees, inside trees, on the ground, and view-light masking correlations. In this paper we present a method to render very large forest scenes in realtime, with realistic lighting at all scales, and without popping nor aliasing (Figure 15). Our method is based on two new forest representations, z-fields and shader-maps, with a seamless transition between them. Our first model builds on light fields and height fields to represent and render the nearest trees individually, accounting for all lighting effects. Our second model is a location, view and light dependent shader mapped on the terrain, accounting for the cumulated subpixel effects. Qualitative comparisons with photos show that our method produces realistic results.



Figure 15. Some real-time results obtained with our method, showing large forest scenes with a wide range of view distances, various tree densities and lighting conditions.

6.4.3. Representing Appearance and Pre-filtering Subpixel Data in Sparse Voxel Octrees

Participants: Eric Heitz, Fabrice Neyret.

Sparse Voxel Octrees (SVOs) represent efficiently complex geometry on current GPUs. Despite the fact that LoDs come naturally with octrees, interpolating and filtering SVOs are still issues in current approaches. In this paper, we propose a representation for the appearance of a detailed surface with associated attributes stored within a voxel octree. We store macro- and micro-descriptors of the surface shape and associated attributes in each voxel. We represent the surface macroscopically with a signed distance field and we encode subvoxel microdetails with Gaussian descriptors of the surface and attributes within the voxel. Our voxels form a continuous field interpolated through space and scales, through which we cast conic rays. Within the ray marching steps, we compute the occlusion distribution produced by the macro-surface inside a pixel footprint, we use the microdescriptors to reconstruct light- and view-dependent shading, and we combine fragments in an A-buffer way. Our representation efficiently accounts for various subpixel effects. It can be continuously interpolated and filtered, it is scalable, and it allows for efficient depth-of-field. We illustrate the quality of these various effects by displaying surfaces at different scales, and we show that the timings per pixel are scale-independent (Figure 16).



Figure 16. Our method allows for correct filtering of color variations, like anti-aliasing demonstrated here.

MORPHEO Team

6. New Results

6.1. A discrete 3D+t Laplacian framework for mesh animation processing

In this work we extend the discrete 3D Laplacian framework to mesh animations, represented as temporally coherent sequences of meshes (Figure 3). In order to let the user control the motion influence with respect to the geometry, we introduce a parameter for the time dimension. Our discrete 3D+t Laplace operator holds the same properties as the discrete 3D Laplacian, as soon as this parameter is non negative. We demonstrate the usefulness of this framework by extending Laplacian-based mesh editing and fairing techniques to mesh animations [15].



Figure 3. 3D+t Laplacian

6.2. Surface Flow

Recovering dense motion information is a fundamental intermediate step in the image processing chain upon which higher level applications can be built, such as tracking or segmentation. For that purpose, pixel observations in the image provide useful motion cues through temporal variations of the intensity function. We have studied the estimation of dense, instantaneous 3D motion fields over non-rigidly moving surface observed by multi-camera systems. The motivation arises from multi-camera applications that require motion information for arbitrary subjects, in order to perform tasks such as surface tracking or segmentation. To this aim, we have proposed a novel framework that allows to efficiently compute dense 3D displacement fields using low level visual cues and geometric constraints. The main contribution is a unified framework that combines flow constraints for small displacements with temporal feature constraints for large displacements

and fuses them over the surface using local rigidity constraints. The resulting linear optimization problem allows for variational solutions and fast implementations. Experiments conducted on synthetic and real data demonstrated the respective interests of flow and feature constraints as well as their efficiency to provide robust surface motion cues when combined.

As an extension of this work, we also studied the situation where a depth camera and one or more color cameras are available, a common situation with recent composite sensors such as the Kinect. In this case, geometric information from depth maps can be combined with intensity variations in color images in order to estimate smooth and dense 3D motion fields. We propose a unified framework for this purpose, that can handle both arbitrary large motions and sub-pixel displacements. The novelty with respect to existing scene flow approaches is that it takes advantage of the geometric information provided by the depth camera to define a surface domain over which photometric constraints can be consistently integrated in 3D. Experiments on real and synthetic data provide both qualitative and quantitative results that demonstrated the interest of the approach[12].

6.3. Progressive Shape Models

In this work we address the problem of recovering both the topology and the geometry of a deformable shape using temporal mesh sequences (Figure 4). The interest arises in multi-camera applications when unknown natural dynamic scenes are captured. While several approaches allow recovery of shape models from static scenes, few consider dynamic scenes with evolving topology and without prior knowledge. In this nonetheless generic situation, a single time observation is not necessarily enough to infer the correct topology of the observed shape and evidences must be accumulated over time in order to learn this topology and to enable temporally consistent modelling. This appears to be a new problem for which no formal solution exists. We have proposed a principled approach based on the assumption that the observed objects have a fixed topology. Under this assumption, the topology can be progressively learned during the capture of a dynamic scene evolutions. The approach has been successfully experimented on several standard 4D datasets and we believe that it paves the way to more general multi-view scene capture and analysis[8].

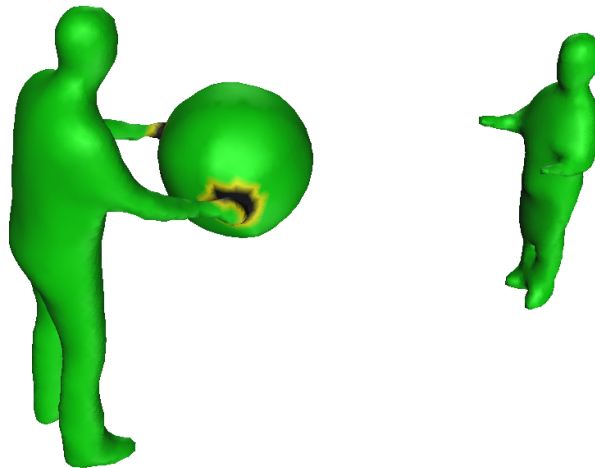


Figure 4. Progressive Shape Models : the balloon can be separated from humans

6.4. Principal Geodesic Dynamics

This work presents a new integration of a data-driven approach using dimension reduction and a physically-based simulation for real-time character animation (Figure 5). We exploit Lie group statistical analysis techniques (Principal Geodesic Analysis, PGA) to approximate the pose manifold of a motion capture sequence by a reduced set of pose geodesics. We integrate this kinematic parametrization into a physically-based animation approach of virtual characters, by using the PGA-reduced parametrization directly as generalized coordinates of a Lagrangian formulation of mechanics. In order to achieve real-time without sacrificing stability, we derive an explicit time integrator by approximating existing variational integrators. Finally, we test our approach in task-space motion control. By formulating both physical simulation and inverse kinematics time stepping schemes as two quadratic programs, we propose a features-based control algorithm that interpolates between the two metrics. This allows for an intuitive trade-off between realistic physical simulation and controllable kinematic manipulation[9].

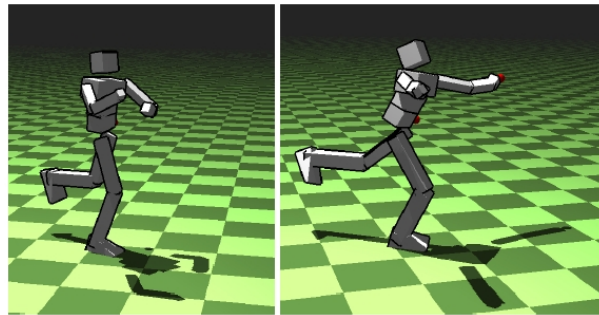


Figure 5. Principal Geodesic Dynamics : test of the balance controller

6.5. A Minimal Solution for Camera Calibration Using Independent Pairwise Correspondences

We have proposed a minimal algorithm for fully calibrating a camera from 11 independent pairwise point correspondences with two other calibrated cameras. Unlike previous approaches, our method neither requires triple correspondences, nor prior knowledge about the viewed scene. This algorithm can be used to insert or re-calibrate a new camera into an existing network, without having to interrupt operation. Its main strength comes from the fact that it is often difficult to find triple correspondences in a camera network. This makes the algorithm, for the specified use cases, probably the most suited calibration solution that does not require a calibration target, and hence can be performed without human interaction [10].

6.6. N-Tuple Color Segmentation for Multi-View Silhouette Extraction

We have presented a new method to extract multiple segmentations of an object viewed by multiple cameras, given only the camera calibration. This method relies on the n-tuple color model to express inter-view consistency when inferring in each view the foreground and background color models permitting the final segmentation. A color n-tuple is a set of pixel colors associated to the n projections of a 3D point. The first goal is set as finding the MAP estimate of background/foreground color models based on an arbitrary sample set of such n-tuples, such that samples are consistently classified, in a soft way, as "empty" if they project in the background of at least one view, or "occupied" if they project to foreground pixels in all views. An Expectation Maximization framework is then used to alternate between color models and soft classifications. In a final step, all views are segmented based on their attached color models. The approach is significantly

simpler and faster than previous multi-view segmentation methods, while providing results of equivalent or better quality. [6].

6.7. Cage-based Motion Recovery using Manifold Learning

We have proposed a flexible model-based approach for the recovery of parameterized motion from a sequence of 3D meshes without temporal coherence (Figure 6). Unlike previous model-based approaches using skeletons, we embed the deformation of a reference mesh template within a low polygonal representation of the mesh, namely the cage, using Green Coordinates. The advantage is a less constrained model that more robustly adapts to noisy observations while still providing structured motion information, as required by several applications. The cage is parameterized with a set of 3D features dedicated to the description of human morphology. This allows to formalize a novel representation of 3D meshed and articulated characters, the Oriented Quads Rigging (OQR). To regularize the tracking, the OQR space is subsequently constrained to plausible poses using manifold learning. Results are shown for sequences of meshes, with and without temporal coherence, obtained from multiple view videos preprocessed by visual hull. Motion recovery applications are illustrated with a motion transfer encoding and the extraction of trajectories of anatomical joints. Validation is performed on the HumanEva II database[7].

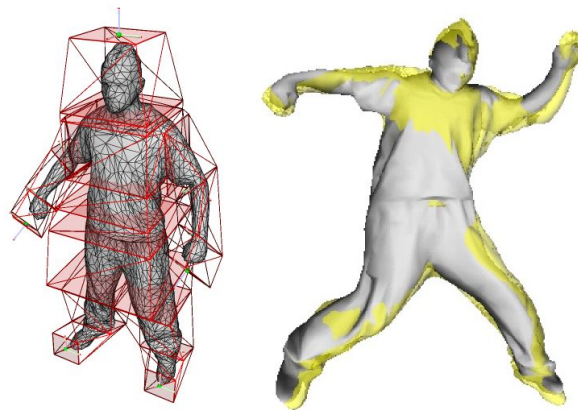


Figure 6. Cage-based Motion Recovery using Manifold Learning

6.8. 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 7). 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[3].

6.9. Keypoints and Local Descriptors of Scalar Functions on 2D Manifolds

This work addresses the problem of describing surfaces using local features and descriptors. While methods for the detection of interest points in images and their description based on local image features are very well

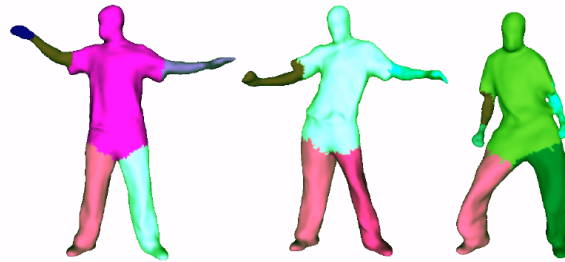


Figure 7. Segmentation of temporal mesh sequences into rigidly moving components

understood, their extension to discrete manifolds has not been well investigated. We provide a methodological framework for analyzing real-valued functions defined over a 2D manifold, embedded in the 3D Euclidean space, e.g., photometric information, local curvature, etc. Our work is motivated by recent advancements in multiple-camera reconstruction and image-based rendering of 3D objects: there is a growing need for describing object surfaces, matching two surfaces, or tracking them over time. Considering polygonal meshes, we propose a new methodological framework for the scale-space representations of scalar functions defined over such meshes. We propose a local feature detector (MeshDOG) and region descriptor (MeshHOG). Unlike the standard image features, the proposed surface features capture both the local geometry of the underlying manifold and the scale-space differential properties of the real-valued function itself. We provide a thorough experimental evaluation. The repeatability of the feature detector and the robustness of feature descriptor are tested, by applying a large number of deformations to the manifold or to the scalar function[4].

PERCEPTION Team

6. New Results

6.1. 3D shape analysis and registration

We address the problem of 3D shape registration and we propose a novel technique based on spectral graph theory and probabilistic matching. Recent advancement in shape acquisition technology has led to the capture of large amounts of 3D data. Existing real-time multi-camera 3D acquisition methods provide a frame-wise reliable visual-hull or mesh representations for real 3D animation sequences. The task of 3D shape analysis involves tracking, recognition, registration, etc. Analyzing 3D data in a single framework is still a challenging task considering the large variability of the data gathered with different acquisition devices. 3D shape registration is one such challenging shape analysis task. The main contribution of this chapter is to extend the spectral graph matching methods to very large graphs by combining spectral graph matching with Laplacian embedding. Since the embedded representation of a graph is obtained by dimensionality reduction we claim that the existing spectral-based methods are not easily applicable. We discuss solutions for the exact and inexact graph isomorphism problems and recall the main spectral properties of the combinatorial graph Laplacian; We provide a novel analysis of the commute-time embedding that allows us to interpret the latter in terms of the PCA of a graph, and to select the appropriate dimension of the associated embedded metric space; We derive a unit hyper-sphere normalization for the commute-time embedding that allows us to register two shapes with different samplings; We propose a novel method to find the eigenvalue-eigenvector ordering and the eigenvector sign using the eigensignature (histogram) which is invariant to the isometric shape deformations and fits well in the spectral graph matching framework, and we present a probabilistic shape matching formulation using an expectation maximization point registration algorithm which alternates between aligning the eigenbases and finding a vertex-to-vertex assignment. See [22], [34], [19] for more details.

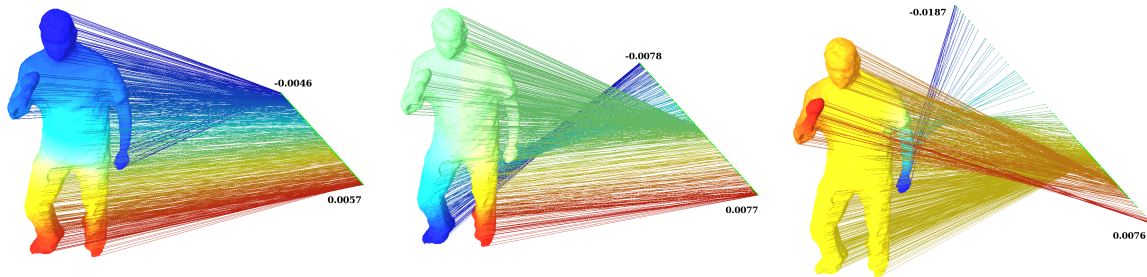


Figure 5. This is an illustration of the concept of the PCA of a shape embedding. The shapes's vertices are projected onto the second, third and fourth eigenvectors of the Laplacian matrix. These eigenvectors can be viewed as the principal directions of the shape (see [34] for more details).

6.2. 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. See [27], [33] for more details.

6.3. 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 analyse 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 [25], [24], [26]

6.4. Sound localization and recognition with a humanoid robot

We addressed the problem of localizing recognizing everyday sound events in indoor environments with a consumer robot. For localization, we use the four microphones that are embedded into the robot’s head. We developed a novel method that uses four non-coplanar microphones and that guarantees that for each set of pairwise TDOA (time difference of arrival) there is a unique 3D source location. For recognition, sounds are represented in the spectrotemporal domain using the stabilized auditory image (SAI) representation. The SAI is well suited for representing pulse-resonance sounds and has the interesting property of mapping a time-varying signal into a fixed-dimension feature vector space. This allows us to map the sound recognition problem into a supervised classification problem and to adopt a variety of classifications schemes. We developed a complete system that takes as input a continuous signal, splits it into significant isolated sounds and noise, and classifies the isolated sounds using a catalogue of learned sound-event classes. The method is validated with a large set of audio data recorded with a humanoid robot in a typical home environment. Extended experiments showed that the proposed method achieves state-of-the-art recognition scores with a twelve-class problem, while requiring extremely limited memory space and moderate computing power. A first real-time embedded implementation in a consumer robot show its ability to work in real conditions. See [23], [28] for more details.

6.5. Audiovisual fusion based on a mixture model

The problem of multimodal clustering arises whenever the data are gathered with several physically different sensors. Observations from different modalities are not necessarily aligned in the sense there is no obvious way to associate or to compare them in some common space. A solution may consist in considering multiple clustering tasks independently for each modality. The main difficulty with such an approach is to guarantee that the unimodal clusterings are mutually consistent. In this paper we show that multimodal

clustering can be addressed within a novel framework, namely conjugate mixture models. These models exploit the explicit transformations that are often available between an unobserved parameter space (objects) and each one of the observation spaces (sensors). We formulate the problem as a likelihood maximization task and we derive the associated expectation-maximization algorithm. The algorithm and its variants are tested and evaluated within the task of 3D localization of several speakers using both auditory and visual data. See [36], [30], [29] for more details.

PRIMA Project-Team

6. New Results

6.1. Social behaviors recognition

Participants: Wafa Benkaouar, Claudine Combe, Dominique Vaufreydaz [correspondant].



Figure 5. On the left image, one can see the telemeter range in red, the foot detection (blue spot) and the angle view from the Kinect (in green). the middle and right image represent RGB camera en depth view from the Kinect.

Recognition of social behaviors is an unconscious innate cognitive process vital to human communication. This skill enables anticipation and increases interactive exchanges quality between humans. Among social behaviors, engagement is the expression of intention for interaction. During engagement phase, many non-verbal signals are used to communicate this intention to the partner, e.g. posture, gaze, spatial information, gestures, vocal cues. Within the context of frail or elderly people at home, companion robots must also be able to detect the engagement of humans in order to adapt their responses during interaction with humans to increase their acceptability.

Classical approaches in the domain are dealing with spatial information. Our hypothesis was that relative spatial information of people and robot are not discriminative in a home-like environment [15]. Our approach integrates multimodal features gathered using a robot companion equipped with a Kinect from Microsoft (see figure 5). Confronted to a robot centered dataset for multimodal social signal processing recorded in a home-like environment, the evaluation highlights its robustness and validates use of such technique in real environment (50% of error reduction). Our experimentations also confirm results from cognitive science domain [61].

6.2. Live monitoring and correction of 3DTV broadcasts

Participants: Pierre Arquier, Frédéric Devernay [correspondant], Sylvain Duchêne, Sergi Pujades-Rocamora, Matthieu Volat.

6.2.1. 3D broadcast monitoring and correction:

One of the achievements of the 3DLive FUI project was the transfer of real-time 3D video monitoring and correction algorithms to the Binocle company, and their integration into the TaggerLive product, which was used during several 3DTV broadcasts between 2010 and 2012 for live monitoring and correction of stereoscopic video. The algorithms that were developed within the PRIMA team and transferred into the TaggerLive are:

Multiscale view-invariant feature detection and matching on the GPU.

Computation of a temporally smooth and robust correction (or rectification) to remove the vertical disparity in the stereoscopic video while keeping the image aspect.

Real-time monitoring of the “depth budget”, or the histogram of the horizontal disparity;

Live alerts when stereoscopic production rules are broken, such as when the disparities are too large, or when there is a stereoscopic window violation.

Real-time implementation of a state-of-the-art dense stereo matching method on the GPU.

6.2.2. 3D content adaptation:

3D shape perception in a stereoscopic movie depends on several depth cues, including stereopsis. For a given content, the depth perceived from stereopsis highly depends on the camera setup as well as on the display size and distance. This can lead to disturbing depth distortions such as the cardboard effect or the puppet theater effect. As more and more stereoscopic 3D content is produced in 3D (feature movies, documentaries, sports broadcasts), a key point is to get the same 3D experience on any display. For this purpose, perceived depth distortions can be resolved by performing view synthesis. We have proposed [19] a real time implementation of a stereoscopic player based on the open-source software Bino, which is able to adapt a stereoscopic movie to any display, based on user-provided camera and display parameters.

6.2.3. Focus mismatch detection:

Live-action stereoscopic content production requires a stereo rig with two cameras precisely matched and aligned. While most deviations from this perfect setup can be corrected either live or in post-production, a difference in the focus distance or focus range between the two cameras will lead to unrecoverable degradations of the stereoscopic footage. We have developed a method [18] to detect focus mismatch between views of a stereoscopic pair in four steps. First, we compute a dense disparity map. Then, we use a measure to compare focus in both images. After this, we use robust statistics to find which images zones have a different focus. Finally, to give useful feedback, we show the results on the original images and give hints on how to solve the focus mismatch.

6.3. Simultaneous localization and mapping (SLAM)

Participants: James Crowley, Marion Decrouez, Frédéric Devernay.

Localisation, place recognition, object recognition. Live processing of a video sequence taken from a single camera enables to model an a priori unknown 3D scene. Metrical SLAM (Simultaneous Localization and Mapping) algorithms track the camera pose while reconstructing a sparse map of the visual features of the 3D environment. Such approaches provide the geometrical foundation for many augmented reality applications in which informations and virtual objects are superimposed on live images captured by a camera. Improving such systems will enable in the future precise industrial applications such as guided-maintenance or guided-assembly in wide installations.

A problem with current methods is the assumption that the environment is static. Indoor environments such as supermarket aisles and factory floors may contain numerous objects that are likely to be moved, disrupting a localization and mapping system. We explore methods for automatic detection and modeling of such objects. We define the scene as a static structure that may contain moving objects and objects are defined as a set of visual features that share a common motion compared to the static structure [39]. Using several explorations of a camera in the same scene, we detect and model moved objects while reconstructing the environment. Experiments highlight the performance of the method in a real case of localization in an unknown indoor environment.

6.4. Post-production tools for 3-D Cinema

Participants: Laurent Boiron, Frédéric Devernay [correspondant], Sylvain Duchêne, Sergi Pujades-Rocamora.

Over the past 6 years, we have been developing 3D movie processing techniques which have been used for the production and post-production of 3D movies (mainly feature-length movies, documentaries and commercials). These include image alignment, view interpolation, depth map computation, etc. These algorithms were developed as C++ libraries, and can be executed using standalone tools. Since the movie post-production workflow relies mainly on standard tools for compositing, color grading, etc., and these tools can be extended by plugin mechanisms, we integrated our post-production algorithms into such a tool, namely Nuke by The Foundry.

We also developed a new method for stereoscopic video cut and paste. Video cut-and-paste consists in semi-interactively segmenting a video object from a video stream, and pasting the segmented video object in another video. The object segmentation is done using a small number of strokes made on a few frames of the video, and can be corrected interactively. Existing methods only worked on monoscopic videos, and extending it to stereoscopic videos required solving important challenges:

The video object must not only remain consistent over time, but also between the left and right views.

The video object may be partially occluded in one or both views.

The camera setup may be different between the first and the second video, causing depth distortion or different depth effects.

We solved the first two challenges by adding left-right stereo consistency based on dense stereo matching, as well as temporal consistency based on optical flow, in an optimization framework based on graph cuts. The user interface was also taken into consideration in the algorithm: any correction of the results (i.e. new strokes on an image) will only propagate forward in time.

6.5. Scene flow computation from RGBZ data

Participants: Frédéric Devernay [correspondant], Julian Quiroga.

The scene flow describes the motion of each 3D point between two times steps. With the arrival of new depth sensors, as the Microsoft Kinect, it is now possible to compute scene flow with a single camera, with promising repercussion in a wide range of computer vision scenarios. We proposed [22] a novel method to compute scene flow by tracking in a Lucas-Kanade framework. Scene flow is estimated using a pair of aligned intensity and depth images, but rather than computing a dense scene flow as in most previous methods, we get a set of 3D motion fields by tracking surface patches. Assuming a 3D local rigidity of the scene, we propose a rigid translation flow model that allows to solve directly for the scene flow by constraining the 3D motion field both in intensity and depth data. In our experimentation we achieve very encouraging results. Since this approach solves simultaneously for the 2D tracking and for the scene flow, it can be used for action recognition in existing 2D tracking based methods or to define scene flow descriptors.

6.6. Knit your Ideas Into Smart Spaces

Participants: Joelle Coutaz, Alexandre Demeure [correspondant], Emeric Fontaine.

We developed KISS (Knit your Ideas Into Smart Spaces), an end-user development system for the home. KISS enables users to program their home with sentences expressed in a pseudo-natural language. Programs can be tested either with the virtual home or in the real home. We led an evaluation that shows that users are able to program a real-life scenario. This work is described in the phd manuscript of Emeric Fontaine [46]. An experimental evaluation shows that KISS can be used to programm a real life scenario.

Participants encountered some difficulties related to the restricted vocabulary used for the experiment. Some difficulties also occurred relative to the understanding of "progressive verbs". To overcome these problems, we envision a system for co-constructing vocabulary with the system, which may lead to the definition of multiple language for communicating with the system.

6.7. Attention-Based Navigation

Participants: Thomas Fisher, Thierry Fraichard [correspondant], Patrick Reignier.

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.

WAM Project-Team

6. New Results

6.1. Multimedia Models and Formats

In the context of the CLAIRE project (see section 7.1.1), a new model for educational documents has been defined. The objectives of this model are:

- to seamlessly handle conventional and richmedia content in the context of a unique pedagogical web platform.
- to be able to store and recover any multimedia document including its spatial and time structure, consistent with HTML5 and Timesheets specifications.
- to have a data model which is format agnostic to cope with existing and future rendering systems.
- to cope with the authoring needs of all users.

We have more specifically worked on the multimedia modelling part for defining spatial and temporal fragment types. These types are used to express the synchronization between different elements within the document.

We are now using this model in the definition and implementation of a web-based authoring user interface.

6.2. XML Processing

In the area of XML processing, we have obtained new results in several directions:

- We have introduced the first system capable of statically verifying properties of a given cascading style sheet (CSS) over the whole set of documents to which this style sheet applies [5]. Properties include coverage of styling information and absence of erroneous rendering.
- In a joint work with the EXMO team, we have introduced a novel approach for deciding the SPARQL query containment problem in the presence of schemas, that paves the way for future extensions [4] [3] [8] [1].
- We have revisited the problem of XML Query-Update Independence Analysis, and showed the relevance of an approach that has been neglected in the literature so far [6]. In particular, we have compared an SMT-modulo with a tree logic approach to Independence Analysis.
- We have made progress on the characterization of the impacts of schema changes on XQuery programs [7].
- We have formally proved a result about the factorization power of the Lean: a construction that we use to speed up the **XML Reasoning Solver**. We have characterized which kind of duplicate subformulas this construction eliminates, and how [10].
- We have proposed a novel technique and a tool for the static type-checking of XQuery programs, using backward type inference [11].
- We have defined a type system for integrating session types for objects in object-oriented languages such as Java, with full structural subtyping, without altering the language semantics [9]. Session types are protocol specifications which describe which sequences of method calls are allowed or disallowed on a given object.

We briefly review these results below.

6.2.1. Automated Analysis of Cascading Style Sheets (CSS)

Developing and maintaining cascading style sheets (CSS) is an important issue to web developers as they suffer from the lack of rigorous methods. Most existing means rely on validators that check syntactic rules, and on runtime debuggers that check the behavior of a CSS style sheet on a particular document instance. However, the aim of most style sheets is to be applied to an entire set of documents, usually defined by some schema. To this end, a CSS style sheet is usually written w.r.t. a given schema. While usual debugging tools help reducing the number of bugs, they do not ultimately allow to prove properties over the whole set of documents to which the style sheet is intended to be applied. We have developed a novel approach to fill this lack [5]. The main ideas are borrowed from the fields of logic and compile-time verification and applied to the analysis of CSS style sheets. We have implemented an original tool (see section 5.1.1) based on recent advances in tree logics. The tool is capable of statically detecting a wide range of errors (such as empty CSS selectors and semantically equivalent selectors), as well as proving properties related to sets of documents (such as coverage of styling information), in the presence or absence of schema information. This new tool can be used in addition to existing runtime debuggers to ensure a higher level of quality of CSS style sheets.

6.2.2. Deciding Satisfiability and Containment for Semantic Web Queries

The problem of SPARQL query containment is defined as determining if the result of one query is included in the result of another for any RDF graph. Query containment is important in many areas, including information integration, query optimization, and reasoning about Entity-Relationship diagrams [1].

We encode this problem into an expressive logic called μ -calculus: where RDF graphs become transition systems, queries and schema axioms become formulas [4] [3]. Thus, the containment problem is reduced to formula satisfiability test. Beyond the logic's expressive power, satisfiability solvers are available for it. Hence, this study allows to exploit these advantages.

In addition, in order to experimentally assess implementation limitations, we have designed a benchmark suite offering different experimental settings depending on the type of queries, projection and reasoning (RDFS) [8]. We have applied this benchmark to three available systems using different techniques highlighting the strengths and weaknesses of such systems.

6.2.3. XML Query-Update Independence Analysis Revisited

XML transformations can be resource-costly in particular when applied to very large XML documents and document sets. Those transformations usually involve lots of XPath queries and may not need to be entirely re-executed following an update of the input document. In this context, a given query is said to be independent of a given update if, for any XML document, the results of the query are not affected by the update. We have revisited Benedikt and Cheney's framework for query-update independence analysis and we have shown that performance can be drastically enhanced, contradicting their initial claims [6]. The essence of our approach and results resides in the use of an appropriate logic, to which queries and updates are both succinctly translated. Compared to previous approaches, ours is more expressive from a theoretical point of view, equally accurate, and more efficient in practice. We have illustrated this through practical experiments and comparative figures.

6.2.4. Toward Automated Schema-directed Code Revision

Updating XQuery programs in accordance with a change of the input XML schema is known to be a time-consuming and error-prone task. We have designed an automatic method aimed at helping developers realign the XQuery program with the new schema [7]. First, we have devised a taxonomy of possible problems induced by a schema change. This allows to differentiate problems according to their severity levels, e.g. errors that require code revision, and semantic changes that should be brought to the developer's attention. Second, we have provided the necessary algorithms to detect such problems using our solver (see section 5.1) to check satisfiability of XPath expressions.

6.2.5. Logical Combinators for Rich Type Systems

We have developed a functional approach to design rich type systems based on an elegant logical representation of types [10]. The representation is not only clean but it also avoids exponential increases in combined complexity due to subformula duplication. This opens the way to solving a wide range of problems such as subtyping in exponential-time even though their direct translation into the underlying logic results in an exponential blowup of the formula size, yielding an incorrectly presumed two-exponential time complexity.

6.2.6. Backward type inference for XQuery

We have designed a novel technique and a tool for static type-checking of XQuery programs [11]. 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 compared to 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.

6.2.7. Session types

Session types allow communication protocols to be specified type-theoretically so that protocol implementations can be verified by static type checking. In [9], we extend previous work on session types for distributed object-oriented languages in three ways. (1) We attach a session type to a class definition, to specify the possible sequences of method calls. (2) We allow a session type (protocol) implementation to be *modularized*, i.e. partitioned into separately-callable methods. (3) We treat session-typed communication channels as objects, integrating their session types with the session types of classes. The result is an elegant unification of communication channels and their session types, distributed object-oriented programming, and a form of typestate supporting non-uniform objects, i.e. objects that dynamically change the set of available methods. We define syntax, operational semantics, a sound type system, and a sound and complete type checking algorithm for a small distributed class-based object-oriented language with structural subtyping. Static typing guarantees that both sequences of messages on channels, and sequences of method calls on objects, conform to type-theoretic specifications, thus ensuring type-safety. The language includes expected features of session types, such as delegation, and expected features of object-oriented programming, such as encapsulation of local state. The main ideas have been implemented as a prototype, extending Java 1.4.

6.3. Multimedia Authoring

In cooperation with EPFL (Lausanne) we pursue our research on template-driven editing for XML multimedia contents (see section 3.3.2). Experiments with very different types of contents have been done with the AXEL library. AXEL is developed by EPFL, based on our joint work on template languages. It is an innovative multi-purpose client-side authoring framework intended for web users with limited skills.

We have addressed the issue of authoring XML multimedia content on the web, focusing on methods that apply to such contents as structured documents, factual data, and multimedia objects [2]. We have shown that a template-based approach enhances the ability for multiple applications to use the produced content.

6.4. Augmented Environments

Most results in the area of augmented environments were presented through various software products and prototypes, including:

- IXE, Interactive eXtensible Engine (see section 5.5 for details). In particular, IXE allowed us to show that a precision of one step is attainable, guidance being done through a mix of spatialized vocal instructions and 3D audio.
- GIF Demonstrator: This application was used to showcase our technologies at the Grenoble Innovation Fair (GIF). Augmented reality was used to find the various booths and products, while our indoor navigation system was guiding visitors to any booth.
- Interactive Audio Panorama: A fun interactive experience with virtual audio. It immerses the user in a complete 360° audio panorama and allows her/him to discover a futuristic house. It demonstrates the authoring possibilities offered by the MAUDL interactive audio language.
- PDRTrack: An indoor localization utility demonstrating the various correction parameters of our IMU-based localization system. The user can record data sets and simulate using various parameters to find out the effect of different map matching settings and their result on localization accuracy. The user can also simply walk in real-time with tracking enabled on a given OpenStreetMap network.
- Sugimotocho Stn: A model of this railway station has been built with the help of the GISLab (Osaka City University). An electronic kick-scooter was used to measure distances and a navigation network was designed to help people to move around in the station.

These products and prototypes were presented in various fora in 2012, in particular at:

- [Grenoble Innovation Fair](#)
- [4I Forum](#)
- [6th European eAccessibility Forum](#)
- [State Of The Map 2012](#)