Activity Report 2011

Section New Results
<table>
<thead>
<tr>
<th>Project-Team</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARENAIRE Project-Team</td>
<td>4</td>
</tr>
<tr>
<td>COMPSYS Project-Team</td>
<td>9</td>
</tr>
<tr>
<td>LICIT Exploratory Action</td>
<td>15</td>
</tr>
<tr>
<td>POP ART Project-Team</td>
<td>18</td>
</tr>
<tr>
<td>VASY Project-Team</td>
<td>30</td>
</tr>
<tr>
<td>BIPOP Project-Team</td>
<td>40</td>
</tr>
<tr>
<td>MISTIS Project-Team</td>
<td>46</td>
</tr>
<tr>
<td>NANO-D Team</td>
<td>54</td>
</tr>
<tr>
<td>NECS Project-Team</td>
<td>63</td>
</tr>
<tr>
<td>OPALE Project-Team</td>
<td>74</td>
</tr>
<tr>
<td>BAMBOO Team</td>
<td>83</td>
</tr>
<tr>
<td>Beagle Team</td>
<td>90</td>
</tr>
<tr>
<td>DRACULA Project-Team</td>
<td>92</td>
</tr>
<tr>
<td>IBIS Project-Team</td>
<td>93</td>
</tr>
<tr>
<td>MOISE Project-Team</td>
<td>98</td>
</tr>
<tr>
<td>NUMED Project-Team (section vide)</td>
<td>114</td>
</tr>
<tr>
<td>STEEP Exploratory Action</td>
<td>115</td>
</tr>
<tr>
<td>AMAZONES Team</td>
<td>117</td>
</tr>
<tr>
<td>DNET Team</td>
<td>121</td>
</tr>
<tr>
<td>GRAAL Project-Team</td>
<td>123</td>
</tr>
<tr>
<td>MESCAL Project-Team</td>
<td>137</td>
</tr>
<tr>
<td>MOAIS Project-Team</td>
<td>141</td>
</tr>
<tr>
<td>PLANETE Project-Team</td>
<td>143</td>
</tr>
<tr>
<td>RESO Project-Team</td>
<td>158</td>
</tr>
<tr>
<td>SARDES Project-Team</td>
<td>163</td>
</tr>
<tr>
<td>SWING Team</td>
<td>170</td>
</tr>
<tr>
<td>ARTIS Project-Team</td>
<td>178</td>
</tr>
<tr>
<td>E-MOTION Project-Team</td>
<td>188</td>
</tr>
<tr>
<td>EVASION Project-Team</td>
<td>209</td>
</tr>
<tr>
<td>EXMO Project-Team</td>
<td>217</td>
</tr>
<tr>
<td>LEAR Project-Team</td>
<td>219</td>
</tr>
<tr>
<td>Morpheeo Team</td>
<td>225</td>
</tr>
<tr>
<td>PERCEPTION Project-Team</td>
<td>231</td>
</tr>
<tr>
<td>PRIMA Project-Team</td>
<td>233</td>
</tr>
<tr>
<td>WAM Project-Team</td>
<td>234</td>
</tr>
</tbody>
</table>
6. New Results

6.1. Hardware Arithmetic and Architecture

Participants: Florent de Dinechin, Hong Diep Nguyen, Bogdan Pasca, Honoré Takeugming, Álvaro Vázquez Álvarez, Nicolas Brunie, Sylvain Collange.

6.1.1. FPGA-specific arithmetic

Reconfigurable computing has the opportunity of using exotic operators that would not make sense in a general-purpose microprocessor [43], for instance the constant dividers studied in 6.1.2. Such operators must be also be matched to the precision and performance needed by applications. F. de Dinechin and B. Pasca described the FloPoCo framework that assists the construction of correct pipelines and the automatic testing of such operators [28]. For this context, B. Pasca, with H. D. Nguyen, now at U.C. Berkeley, and T. Preusser, from T. U. Darmstadt, described improved architectures for short-latency adders on modern FPGAs [39]. With Ch. Alias and A. Plesco (Compsys project-team), he studied the integration in of deeply pipelined arithmetic datapath in high-level synthesis tools [51].

6.1.2. 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 [26]. 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 [44].

6.1.3. Elementary Functions

A. Vázquez worked with J. Bruguera, from U. Santiago de Compostella, on hardware architectures for evaluating $q$-th roots [66]. Their solution composes digit-recurrence operators for reciprocal, logarithm, multiplication and exponential.

6.1.4. Extensions of the fused-multiply-and-add operator

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 [56].

6.1.5. Emerging throughput-oriented architecture

On massively multi-threaded processors like GPUs, neighbor threads are likely to operate on similar data. S. Collange showed with A. Kouyoumdjian how it is possible to take advantage of this inter-thread value correlation at the hardware level with a hardware cache-compression technique on GPUs [59]. With D. Sampaio, R. Martins, and F. Magno Quintão Pereira (U. Minas Gerais), he then addressed this question also at the compiler level using a compiler stage to identify statically data patterns in GPGPU programs [65].

Current GPU architectures require specific instruction sets with control-flow reconvergence annotations, and only support a limited number of control-flow constructs. S. Collange and N. Brunie, with G. Diamos (NVIDIA) generalized dynamic vectorization to arbitrary control flow on standard instruction sets with no compiler involvement [46], [57], [54]. In addition, this technique allows divergent branches to be executed in parallel, as a way to increase the throughput of parallel architectures [55].
6.2. Efficient Floating-Point Arithmetic and Applications


6.2.1. Correctly Rounded Sums

P. Kornerup (Odense Univ., Denmark), V. Lefèvre, N. Louvet and J.-M. Muller have given a study of some basic blocks needed in the design of floating-point summation algorithms. In particular, in radix-2 floating-point arithmetic, they 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 have investigated the possible use of another algorithm, Dekker’s Fast2Sum algorithm, in radix-10 arithmetic. Under reasonable conditions, they have also proven that no algorithms performing only round-to-nearest additions/subtractions exist to compute the round-to-nearest sum of at least three floating-point numbers. Starting from an algorithm due to Boldo and Melquiond, they have also presented new results about the computation of the correctly-rounded sum of three floating-point numbers [21].

6.2.2. Error of an FMA

The fused multiply-add (FMA) instruction, specified by the IEEE 754-2008 Standard for Floating-Point Arithmetic, eases some calculations, and is already available on some current processors such as the Power PC or the Itanium. S. Boldo (EPI Proval) and J.-M. Muller first extended an earlier work on the computation of the exact error of an FMA (by giving more general conditions and providing a formal proof). Then, they presented a new algorithm that computes an approximation to the error of an FMA, and provide error bounds and a formal proof for that algorithm [16].

6.2.3. Accurate computation of \(ad - bc\) with an FMA

C.-P. Jeannerod, N. Louvet, and J.-M. Muller have provided in [60] a detailed rounding error analysis of Kahan’s FMA-based algorithm for the computation of expressions of the form \(ad - bc\). They showed that Kahan’s algorithm is always highly accurate, and under mild assumptions on the radix and the precision gave an optimal bound on the absolute error and an asymptotically optimal bound on the relative error. They also studied how the relative error varies as a function of the relative order of magnitude of the two products \(ad\) and \(bc\). Finally, they investigated whether the error bounds can be improved in special cases like sums of squares and discriminants.

6.2.4. Performing Arithmetic Operations on Round-to-Nearest Operations

During any composite computation, there is a constant need for rounding intermediate results before they can participate in further processing. Recently, a class of number representations denoted RN-Codings were introduced, allowing an unbiased rounding-to-nearest to take place by a simple truncation, with the property that problems with double-roundings are avoided. P. Kornerup (Odense Univ., Denmark), J.-M. Muller and A. Panhaleux first investigate a particular encoding of the binary representation. This encoding is generalized to any radix and digit set; however, radix complement representations for even values of the radix turn out to be particularly feasible. The encoding is essentially an ordinary radix complement representation with an appended round-bit, but still allowing rounding-to-nearest by truncation, and thus avoiding problems with double-roundings. Conversions from radix complement to these round-to-nearest representations can be performed in constant time, whereas conversion the other way, in general, takes at least logarithmic time. Not only is rounding-to-nearest a constant time operation, but so is also sign inversion, both of which are at best log-time operations on ordinary two’s complement representations. Addition and multiplication on such fixed-point representations are first analyzed and defined in such a way that rounding information can be carried along in a meaningful way, at minimal cost. The analysis is carried through for a compact (canonical) encoding using two’s complement representation, supplied with a round-bit. Based on the fixed-point encoding, it is shown possible to define floating-point representations, and a sketch of the implementation of an FPU is presented [22].
6.2.5. Augmented Precision Square Roots, 2-D Norms, and Discussion on Correctly Rounding \( \sqrt{x^2 + y^2} \)

Define an “augmented precision” algorithm as an algorithm that returns, in precision-\( p \) floating-point arithmetic, its result as the unevaluated sum of two floating-point numbers, with a relative error of the order of \( 2^{-2p} \). Assuming an FMA instruction is available, N. Brisebarre, M. Joldeș, P. Kornerup (Odense University, Denmark), E. Martin-Dorel and J.-M. Muller perform a tight error analysis of an augmented precision algorithm for the square root, and introduce two slightly different augmented precision algorithms for the 2D-norm \( \sqrt{x^2 + y^2} \). Then they give tight lower bounds on the minimum distance (in ulps) between \( \sqrt{x^2 + y^2} \) and a midpoint when \( \sqrt{x^2 + y^2} \) is not itself a midpoint. This allows them to determine cases when their algorithms make it possible to return correctly-rounded 2D-norms [30].

6.2.6. Midpoints and exact points of some algebraic functions in floating-point arithmetic

When implementing a function \( f \) in floating-point arithmetic, if we wish correct rounding and good performance, it is important to know if there are input floating-point values \( x \) such that \( f(x) \) is either the middle of two consecutive floating-point numbers (assuming rounded-to-nearest arithmetic), or a floating-point number (assuming rounded toward \( \pm \infty \) or toward 0 arithmetic). In the first case \( f(x) \) is a midpoint, and in the second case it is an exact point. In [20] C.-P. Jeannerod, N. Louvet, J.-M. Muller, and A. Panhaux have studied whether such midpoints and exact points exist for some usual algebraic functions and various floating-point formats. When midpoints or exact points exist, they have been characterized or, when possible, listed exhaustively. The results and the techniques presented in this paper can be used in particular to deal with both the binary and the decimal formats defined in the IEEE 754-2008 standard for floating-point arithmetic.

6.3. Correct Rounding of Elementary Functions

Participants: Florent de Dinechin, Vincent Lefèvre, Jean-Michel Muller, Bogdan Pasca, Serge Torres.

6.3.1. FPGA Acceleration of the Search For Hardest-to-Round Cases

The IEEE 754-2008 standard for floating-point arithmetic recommends (yet does not dictate) that some elementary functions should be correctly rounded. That is, given a rounding function \( \circ \) (e.g., round to nearest even, or round to \( \pm \infty \)), when evaluating function \( f \) at the floating-point number \( x \), the system should always return \( \circ(f(x)) \).

Building a fast correctly rounded library for some target floating-point (FP) format requires preliminarily solving a problem called the table maker’s dilemma. This requires very large computations which may use environments and formats totally different from the target environment and format. F. de Dinechin, V. Lefèvre, J.-M. Muller, B. Pasca and A. Plesco suggest performing these computations on an FPGA. Their paper [45] won the best paper award at the ASAP2011 conference.

6.3.2. Hierarchical Polynomial Approximation of a Function by Polynomials

Algorithms used to search for the hardest-to-round cases of a function requires the approximation of the function by small-degree polynomials on small intervals. This can be done efficiently by a hierarchical polynomial approximation. Work is being done to improve this method by replacing interval arithmetic (as partly used in the current tools) by static error bounds. This will allow us to better control the precision needed to compute the coefficient of the polynomials. The implementation will also be simpler.

6.4. Validation and Automation

6.4.1. Efficient Implementation of Algorithms for Interval Linear Algebra

H.-D. Nguyen and N. Revol proposed an algorithm to solve linear systems with interval coefficients. The same approach can be used to verify the solution of a linear system with floating-point coefficients, i.e. to compute an interval enclosing the error between the exact solution and an approximate solution. The goal is twofold: on the one hand the accuracy of the solution is desired up to the last bit of the floating-point solution, on the other hand the efficiency of the implementation is obtained through the use of optimized BLAS3 routines [48]. The PhD thesis of H.-D. Nguyen [13] contains in particular the algorithm [24] and its properties. Its complexity has been established [47] and its potential use for symbolic-numeric computations has been discussed [50].

6.4.2. Standardization of Interval Arithmetic

We contributed to the creation, and now chair, the IEEE 1788 working group on the standardization of interval arithmetic http://grouper.ieee.org/groups/1788/. The main discussion topics of this working group [49], for the year 2011, were exception handling (via decorations). An emerging topic is the repeatability and reproducibility of interval computations, on the same platform or across different platforms.

6.4.3. Formal Proofs of the Arithmetic on Polynomial Models

Using as starting point [9], the calculus with polynomial models, such as Taylor models, based on floating-point coefficients and floating-point operations, has been formalized and checked in Coq [18]. This calculus is at the core of Ariadne, an environment for the study of hybrid systems: the idea is to prove the environment itself, instead of using model-checking on the systems.

6.4.4. Formal Proof Generation for Elementary functions

The proof of the correct rounding property for an elementary function requires tight bounds on the error involved in the function code. F. de Dinechin, with Ch. Lauter (LIP6) and G. Melquiond (INRIA Proval) have described the use of the Gappa proof assistant to compute such tight bounds rigorously [27].

6.4.5. Code Generation for Polynomial Evaluation

A given arithmetic expression may be evaluated on a computer in several ways, depending on the parenthesisation and the ordering of terms in use. Among all the possible evaluations, one may want to choose one that is as fast and accurate as possible. In [12] Ch. Mouilleron introduced a set of algorithms in order to generate all these possible evaluations, to count them, and to find an optimal or nearly optimal one according to a given criteria. Thanks to this work, several sequences related to numbers of evaluations have been discovered and added to Sloane’s on-line encyclopedia of integer sequences (OEIS). Moreover, this allowed to show experimentally that an algorithm by Paterson and Stockmeyer for the evaluation of a polynomial p at a matrix point is optimal for small degrees of p. Finally, this work has led to the revamping of the software tool CGPE presented in [38] (see also § 5.5).

6.5. Arithmetic and Algorithms

Participants: Guillaume Hanrot, Claude-Pierre Jeannerod, Adeline Langlois, Ivan Morel, Christophe Mouilleron, Andrew Novocin, Xavier Pujol, Damien Stehlé, Gilles Villard.

6.5.1. Faster Lattice Reduction

Andrew Novocin, Damien Stehlé and Gilles Villard [40] designed an algorithm, $\tilde{L}^1$, with the following specifications: It takes as input an arbitrary basis $B$ in $\mathbb{Z}^{d \times d}$ of a lattice $L$; It computes a basis of $\tilde{L}$ which is reduced for a mild modification of the Lenstra-Lenstra-Lovász reduction; It terminates in time $O(d^3 \beta + d^{e+1} \beta)$ where $\beta = \log \|B\|$ (and $\omega$ is a valid exponent for matrix multiplication). This is the first LLL-reducing algorithm with a time complexity that is quasi-linear in the bit-length beta of the entries and polynomial in the dimension $d$. A critical ingredient for achieving this result was the study of the effect of small perturbations on the LLL-reducedness of a lattice basis [17].
6.5.2. Computing Short Lattice Vectors

Among all known lattice reduction algorithms, BKZ provides the best trade-off between run-time and smallness of the computed lattice basis. Guillaume Hanrot, Xavier Pujol and Damien Stehlé [32] showed that BKZ can be terminated long before its completion, while still providing bases of excellent quality. More precisely, if it is terminated within a polynomial number of calls to a lower-dimensional Shortest Vector Problem solver, then the bounds on the output quality are as close as desired to the bounds that can be obtained by letting BKZ run until completion.

Guillaume Hanrot, Xavier Pujol and Damien Stehlé also surveyed the known algorithms for solving the Shortest Vector Problem [31].

6.5.3. Lattice-Based Cryptography

NTRUEncrypt is the fastest known lattice-based encryption scheme. Its moderate key-sizes, excellent asymptotic performance and conjectured resistance to quantum computers could make it a desirable alternative to factorisation and discrete-log based encryption schemes. Damien Stehlé and Ron Steinfeld [41] showed how to modify NTRUEncrypt to make it provably resistance to Chosen Plaintext Attacks, under the assumed quantum hardness of standard worst-case lattice problems restricted to a family of lattices related to some cyclotomic fields.

6.5.4. Lattices and Communication Theory

Cong Ling, Shuiyin Liu, Laura Luzzi and Damien Stehlé studied and optimized lattice algorithms that are relevant for MIMO communications [23], [37]. These algorithms tackle the Bounded Distance Decoding Problem: Given a point within a small prescribed distance to a given lattice, find the lattice vector closest to it.

6.5.5. Other Applications of Lattice Reduction Algorithms

In [35] Jürgen Klüners, Mark van Hoeij, and Andrew Novocin showed how to use the LLL lattice reduction algorithm for computing a compact representation of the set of all subfields of any given number field. William Hart (Warwick Mathematics Institute, UK), Mark van Hoeij (Florida State University, USA) and Andrew Novocin exploited the very latest progress in lattice reduction to propose a fine-tuned cutting-edge implementation of a polynomial factorization algorithm.

6.5.6. Polynomial Arithmetic

With William Hart and Mark van Hoeij, A. Novocin proposed in [33] a state of the art algorithm for factoring polynomials in \( \mathbb{Z}[x] \). The algorithm is fast in practice, saving in a large class of common examples, without sacrificing performance on worst-case polynomials. The presented algorithm is structured along the lines of algorithms with the best theoretical complexity. In [34] William Hart and A. Novocin proposed an efficient algorithm for computing the composition of two univariate polynomials. Their work builds upon the Brent-Kung algorithm.

6.5.7. Exact Linear Algebra

Transforming a matrix over a field to echelon form, or decomposing the matrix as a product of structured matrices that reveal the rank profile, is a fundamental building block of computational exact linear algebra. For such tasks the best algorithms available so far were either rank sensitive (i.e., of complexity expressed in terms of the exponent of matrix multiplication and the rank of the input matrix) or in place (i.e., using essentially no more memory that what is needed for matrix multiplication). In [61] C.-P. Jeannerod, Clément Pernet (U. Joseph Fourier, Grenoble), and Arne Storjohann (U. Waterloo, Canada) have proposed algorithms that are both rank sensitive and in place. These algorithms are based on a new matrix factorization, namely \( A = CUP \) with \( C \) a column echelon form revealing the row rank profile of \( A \), \( U \) a unit upper triangular matrix, and \( P \) a permutation matrix.
6. New Results

6.1. Introduction

This section presents the results obtained by Compsys in 2011. For clarity, some earlier results are also recalled, when they were continued or extended during the year 2011.

6.2. Studying Optimal Spilling in the Light of SSA

Participants: Florian Brandner, Quentin Colombet, Alain Darte.

Recent developments in register allocation, mostly linked to static single assignment (SSA) form, have shown that it is possible to decouple the 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. We focused on the first phase, for which many open questions remained: in particular, we studied the notion of optimal spilling (what can be expressed?) and the impact of SSA form (does it help?).

To identify the important features for optimal spilling on load-store architectures, we developed a new integer linear programming formulation, more accurate and expressive than previous approaches. Among other features, we can express SSA $\phi$-functions, memory-to-memory copies, and the fact that a value can be stored simultaneously in a register and in memory. We implemented this formulation in LAO and analyzed in details the static and dynamic results obtained for the SPEC INT 2000 and EEMBC 1.1 benchmarks. We can draw, among others, the following conclusions: a) rematerialization is extremely important, b) SSA complicates the formulation of optimal spilling, especially because of memory coalescing when the code is not in CSSA, c) micro-architectural features are significant and thus have to be accounted for, d) significant savings can be obtained in terms of static spill costs, cache miss rates, and dynamic instruction counts.

This work has been presented at the CASES’11 conference [8].

6.3. Copy Elimination on Data Dependence Graphs

Participants: Florian Brandner, Quentin Colombet.

Register allocation recently regained much interest due to new decoupled strategies that split the problem into separate phases: spilling, register assignment, and copy elimination.

A common assumption of existing copy elimination approaches is that the original ordering of the instructions in the program is not changed. We worked on an extension of a local recoloring technique that we developed earlier, called Parallel Copy Motion [30]. 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.

This work will be presented at the conference SAC’12 [6].

6.4. Graph-Coloring and Treescan Register Allocation Using Repairing

Participants: Quentin Colombet, Benoît Boissinot, Philip Brisk [University of California, Riverside], Sebastian Hack [Saarland University], Fabrice Rastello.
Graph coloring and linear scan are two appealing techniques for register allocation as the underlying formalism are extremely clean and simple. Our previous work advocated the use of a decoupled approach that first lowers the register pressure by spilling variables, then performs live-range splitting, coalescing, and coloring in a separate phase. This enables the design of simpler, cleaner, and more efficient register allocators.

In this context, we introduced a new and more general approach to deal with register constraints. This approach, called repairing, does not require any preliminary live-range splitting and does not introduce additional spill code. It ignores register constraints during coloring/coalescing and repairs the violated constraints afterwards. We applied this method to develop both a graph-based and a scan-based decoupled approach: one based on the iterated register coalescer (IRC) and the other on a scan algorithm (the treescan) that uses static single assignment (SSA) properties.

Our experimental evaluation shows that, for the graph-based approach, we reduced the number of vertices (edges) in the interference graph by 26% (33%) without compromising the quality of the generated code. The treescan algorithm improved the compile time of the allocation process by 6.97× over IRC while providing comparable results for the quality of the generated code.

This work was part of a collaboration with the Saarland University and the University of California, Riverside. It has been presented at the conference CASES'11 [7].

6.5. Decoupled Graph-Coloring Register Allocation with Hierarchical Aliasing

Participants: Andre Tavares [UFMG, Brazil], Quentin Colombet, Mariza Bigonha [UFMG, Brazil], Christophe Guillen [STMicroelectronics], Fernando Pereira [UFMG, Brazil], Fabrice Rastello.

Decoupling spilling from register assignment, as mentioned in previous sections, has the main advantage of simplifying the implementation of register allocators. However, the decoupled model faces many problems when dealing with register aliasing, a phenomenon typical in architectures usually seen in embedded systems, such as ARM.

We introduced the semi-elementary form, a program representation that brings decoupled register allocation to architectures with register aliasing. The semi-elementary form is much smaller than program representations used by previous decoupled solutions, which leads to register allocators that perform better in terms of time and space. Furthermore, this representation reduces the number of copies that traditional allocators insert into assembly programs. We have empirically validated our results by showing that how our representation improves two well-known graph-coloring-based allocators, namely the iterated register coalescer (IRC) and Bouchez et al.’s brute force (BF) method, both augmented with Smith et al. extensions to handle aliasing. Running our techniques on SPEC CPU 2000, we have reduced the number of nodes in the interference graphs by a factor of 4 to 5, hence speeding-up the allocation time by a factor of 3 to 5. In addition, the semi-elementary form reduces by 8% the number of copies that IRC leaves uncoalesced.

This work is part of a collaboration with the Federal University of Minas Gerais. It has been presented at SCOPES’11 [13].

6.6. A Non-Iterative Data-Flow Algorithm for Computing Liveness Sets in Strict SSA Programs

Participants: Benoit Boissinot, Florian Brandner, Alain Darte, Benoît Dupont de Dinechin [Kalray], Fabrice Rastello.

We revisited the problem of computing liveness sets, i.e., the sets of variables live-in and live-out of basic blocks, for programs in strict static single assignment (SSA). In strict SSA, aka SSA with dominance property, the definition of a variable always dominates all its uses. We exploited this property and the concept of loop-nesting forest to design a fast two-phase data-flow algorithm: a first pass traverses the control-flow graph (CFG), propagating liveness information backwards, a second pass traverses the loop-nesting forest, updating liveness sets within loops. The algorithm is proved correct even for irreducible CFGs.
We analyzed its algorithmic complexity and evaluated its efficiency on SPEC INT 2000. Compared to traditional iterative data-flow approaches, which perform updates until a fixed point is reached, our algorithm is 2 times faster on average. Other approaches are possible that propagate from uses to definitions, one variable at a time, instead of unioning sets as in data-flow analysis. Our algorithm is 1.43 times faster than the fastest alternative on average, when sets are represented as bitsets and for optimized programs, i.e., when there are more variables and larger live-sets and live-ranges.

This work has been presented at the conference APLAS’11 [5].

6.7. SSI Revisited: A Program Representation for Sparse Dataflow Analyses

Participants: Andre Tavares [UFMG, Brazil], Mariza Bigonha [UFMG, Brazil], Roberto Bigonha [UFMG, Brazil], Benoit Boissinot, Fernando Pereira [UFMG, Brazil], Fabrice Rastello.

Dataflow analyses usually associate information about variables to program regions. Informally, if these regions are too small, e.g., a point between two consecutive statements, we call the analysis dense. On the other hand, if these regions include many such points, then we call it sparse. We developed a systematic method to build program representations that support forward and/or backward sparse analyses. To pave the way that leads to this framework, we first clarified the literature on intermediate program representations. We revisited the static single information (SSI) form introduced in the 90s and showed how to simplify the construction of program representations for unidirectional dataflow analyses. We showed how to cope with live-ranges that have multiple uses/definitions without losing the equivalence property with the initial dataflow analysis problem. This allows us to simplify, for unidirectional problems, the SSI construction algorithm.

We also showed that our approach, up to a parameter choice, subsumes other program representations such as the SSA, SSI, and e-SSA forms. We can produce intermediate representations isomorphic to the sparse evaluation graphs (SEGs) of Choi et al. This data structure enables sparse solutions to the class of dataflow problems called partitioned dataflow analysis (PDA). However, contrary to SEGs, we can handle - sparsely - problems that are not PDA. We have implemented this framework in the LLVM compiler and have empirically compared different program representations in terms of size and construction time.

This work is part of a collaboration with the Federal University of Minas Gerais 8.2 and is under reviewing process for the journal Science of Computer Programming.

6.8. Incremental Spilling

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

This work addresses the minimization of the spill code overhead in the contexts of both coupled and decoupled register allocation. We devised a heuristic approach called stacking; it incrementally allocates clusters of variables, as opposed to the conventional incremental spilling approach. We describe two polynomial methods, a stacking-optimal allocator and a greedy stacking-independent-set allocator. The first method is very close to the optimal allocation; the second method outperforms state-of-the-art heuristics for just-in-time compilation.

This work has been submitted for publication.

6.9. Program Analysis and Communication Optimizations for HLS

Participants: Christophe Alias, Alain Darte, Alexandru Plesco.

High-level synthesis (HLS) tools are now getting more mature for generating hardware accelerators with an optimized internal structure, thanks to efficient scheduling techniques, resource sharing, and finite-state machines generation. However, interfacing them with the outside world, i.e., integrating the automatically-generated hardware accelerators within the complete design, with optimized communications, so that they achieve the best throughput, remains a very hard task, reserved to expert designers. The goal of our research on HLS is to study and to develop source-to-source strategies to improve the design of these interfaces, trying to consider the HLS tool as a back-end for more advanced front-end transformations.
Using the C2H HLS tool from Altera, which can synthesize hardware accelerators communicating to an external DDR-SDRAM memory, we showed that it is possible to automatically restructure the application code, to generate adequate communication processes in C, and to compile them all with C2H, so that the resulting application is highly-optimized, with full usage of the memory bandwidth.

These transformations and optimizations, which combine techniques such as double buffering, array contraction, loop tiling, software pipelining, among others, were incorporated in an automatic source-to-source transformation tool, called CHUBA (see Section 5.7), based on the polyhedral model representation. Our study shows that HLS tools can indeed be used as back-end optimizers for front-end optimizations, as it is the case for standard compilation with high-level transformations developed on top of assembly-code optimizers. We believe this is the way to go for making HLS tools viable. The complete automation of the process will be presented at PPoPP’12 [3] and Impact’12 [15].

We also showed how to extend this method to programs with irregular control and array accesses. The main difficulty arises when some data may be redefined in the accelerator but this is not sure. We showed that techniques based on parametric polyhedral optimizations can be used to generate the sets of data to be loaded (resp. stored) just before (resp. after) each tile. An interesting feature is that the previous method appears nicely as a particular case when no approximation is needed. This work is fully described in a research report [23], but is not yet published in a conference or journal.

6.10. Compilation of Hardware Accelerators with Pipelined Arithmetic

Participants: Christophe Alias, Bogdan Pasca [PhD student, Arénaire Inria Team], Alexandru Plesco.

By nature, source-level optimizations cannot perform fine optimizations of the datapath and the control, sometimes mandatory to obtain performances. In high-level synthesis, the circuit generated must be efficient and must produce quality results. This last point is the specialty of the Arénaire Inria-team, which develops the tool FLopCO, an open-source FPGA-specific generator of pipelined floating-point arithmetic operators, from a functional description as \((x, y, z) \mapsto e^x + y \cdot \cos(z)\). The user can specify the precision (mantissa, exponent) and the maximum frequency, then FLopCO generates the corresponding operator.

We have developed an algorithm to automatically generate, from a C program, an hardware accelerator that efficiently uses these pipelined operators. The main issue is to reschedule the initial program execution in order to keep the operator’s pipeline as busy as possible, while minimizing memory access. Then, the new execution schedule is used to generate the VHDL code of finite state machines (FSM) controlling the data-flow through the arithmetic operator. This work has been published at the conference ARC’11 [4].

We also showed how our method can be used as a tool to generate control FSMs of multiple parallel computing cores accelerating the same application [25]. This work has been submitted to the journal Microprocessors and Microsystems.

This is still a work in progress and many issues need to be addressed, for example how to extend the program model to general nested loops with more general dependences. These extensions will require to handle properly the communications between the operators and temporary buffers. We believe that the array contraction technique developed in Compsys can be helpful in this context too.

6.11. FPGA Optimized Table Maker’s Dilemma Architecture

Participants: Alexandru Plesco, Florent De Dinechin [Assistant Prof., ARENAIRE Inria Team], Jean-Michel Muller [Research Director, ARENAIRE Inria Team], Bogdan Pasca [PhD student, ARENAIRE Inria Team].

In this work, with some members of the Arénaire team, we developed an algorithm that enables to perform the table maker’s dilemma on a very regular architecture such as an FPGA. The algorithm is crucially different from the algorithm implemented on a standard PC and exploits efficiently the FPGA optimized high-performance arithmetic operators.
The core component (TaMaDi core) of our design is the polynomial evaluator based on the tabulated differences method that uses Remez algorithm. We instantiated multiple TaMaDi cores that work on $2^n$ disjoint intervals obtained from the splitting of the input interval. The TaMaDi cores are connected using a pipelined communication architecture based on the credit communication methodology. We focused on the communication architectures internal to a FPGA that can scale with limited impact on the frequency of the generated data processing architecture. We defined a new pipelined credit based communication interface that leads to a full-rate pipeline data transmission with the cost of small transfer initialization latency. This design has been proven to be effective on parallel design of TaMaDi FPGA specific algorithms. This strategy can be reused in any parallel FPGA design.

This work has been published at ASAP’11 [14] where it received a best paper award.

6.12. Termination of Big Programs

Participants: Christophe Alias, Laure Gonnord, Guillaume Andrieu [Undergraduate Internship, Polytech’Lille].

In a previous work with Alain Darte and Paul Feautrier, we showed how to prove the termination of a certain class of irregular programs with while loops [29]. Our technique amounts to compute a sequential schedule – called ranking function – for the program, reinvesting most of the techniques from [37] to schedule static loops. Our termination method is based on the resolution of a linear programming instance, and does not scale.

In most of big programs, all the information is not relevant for proving termination. Only a few slices need a termination proof. Moreover, inside a loop nest, termination can be proved incrementally. Inner loops are proved to terminate, and then are replaced by a summary. Our contribution is thus a reduction to prove a single loop termination, the price being the approximation made while computing the summary of a given (nest of) loop(s). This method can also be used to schedule irregular programs, which is clearly a need for program optimization.

A prototype has been designed and is currently under testing. A paper is in preparation. This method has been developed during the undergraduate internship of Guillaume Andrieu, from the Engineering School Polytech’Lille.

6.13. Simplification of Boolean Affine Formulas

Participant: Paul Feautrier.

Boolean Affine Formulas, in which affine inequalities are combined by boolean connectives, are ubiquitous in computer science: static analysis, code and hardware generation, symbolic model checking and many other techniques use them as a compact representation of large or infinite sets. Common algorithms tend to generate large and highly redundant formulas, hence the necessity of a simplifier for keeping the overall complexity under control. Simplification is a difficult problem, at least as hard as SMT solving, with a worst case complexity exponential in the number of affine inequalities. Paul Feautrier has proposed a new method, based on path cutting in Ordered Binary Decision Diagrams, which is able to take advantage of any regularity in the subject formula to speed up simplification. The method has been implemented and tested on benchmarks from several application domains.

The method has been presented at the 4th “Rencontres de la communauté française de compilation”. A detailed description has been submitted for publication; see also [28].

6.14. Retiming for Faust

Participants: Alain Darte, Alexandre Isoard [Master 1 student, ENS-Lyon], Yann Orlarey [Grame].

Faust (Functional Audio Stream) is a formal specification stream-like language designed for real-time signal processing and synthesis. Faust programs are compiled into equivalent C++ programs and optimized for parallel execution. One of the core optimizations in Faust is a preliminary phase called normalization, which amounts to change the delays between operators into an equivalent normalized form used for, among others, redundancy elimination.
We showed that this normalization is actually a special form of retiming, a well-known technique in circuit design and loop transformations. However, an important subtlety needs to be considered: the problem of initialization of signals, which is usually not even mentioned in the retiming literature. We proved that the problem comes from “time-dependent” operators in Faust and that all these operators can be decomposed into combinations of regular operators (in the sense that standard retiming applies) and a single elementary time-dependent operator, called $init$, which transfers any signal with no modification for time $t \geq 0$ and outputs a constant for time $t < 0$.

This collaboration between Compsys and Grame did not go beyond the Master internship of A. Isoard yet and did not lead to an actual implementation within Faust.
5. New Results

5.1. Liability issues in software engineering

Software contracts usually include strong liability limitations or even exemptions of the providers for damages caused by their products. This situation does not favour the development of high quality software because software editors do not have sufficient economic incentives to apply stringent development and verification methods. Indeed, experience shows that products tend to be of higher quality and more secure when the actors in position to influence their development are also the actors bearing the liability for their defects. The usual argument to justify this lack of liability is the fact that software products are too complex and versatile objects whose expected features (and potential defects) cannot be characterised precisely, and which thus cannot be treated as traditional (tangible) goods. Taking up this challenge is one of our objectives \cite{12}: we study liability issues both from the legal and the technical points of view with the aim to put forward a formal framework to (1) define liability in a precise and unambiguous way and (2) establish such liability in case of incident.

Obviously, specifying all liabilities in a formal framework is neither possible nor desirable. Usually, the parties wish to express as precisely as possible certain aspects which are of prime importance for them and prefer to state other aspects less precisely (either because it is impossible to foresee at contracting time all the events that may occur or because they do not want to be bound by too precise commitments). Taking this requirement into account, we provide a set of tools and methods to be used on a need basis in the contract drafting process (as opposed to a monolithic, “all or nothing” approach). Our model is based on execution traces which are abstractions of the log files of the system. In a nutshell, liability is specified as a function taking as parameters a claim and an execution trace and returning a set of “responsible” actors. This set of actors (ideally a singleton) depends on the claim and the errors occurring in the trace. Both errors and claims are expressed as trace properties. The liability function can be made as precise or detailed as necessary by choosing the claims and errors relevant for a given situation \cite{5}.

In order to provide a more generic way to define liabilities, we have also introduced a concept of “logical causality” \cite{11}. Causality has been studied for a long time in computer science, but with quite different perspectives and goals. In the distributed systems community, causality is seen essentially as a temporal property. We have defined several variants of logical causality allowing us to express the fact that an event \( e_2 \) (e.g. a failure) would not have occurred if another event \( e_1 \) had not occurred (“necessary causality”) or the fact that \( e_2 \) could not have been avoided as soon as \( e_1 \) had occurred (“sufficient causality”). We have applied these technical definitions of causality to real case studies and related them to the legal views of causality.

As far as legal issues are concerned, we have studied the legal validity of the technical solutions proposed in the project both in terms of legal evidence and allocation of liabilities \cite{6}. Contract templates have been defined in collaboration with lawyers to allow the parties to effectively integrate our results in a legal contract \cite{6}.

5.2. Privacy

Despite apparently strong legal protections, many citizens feel that information technologies have invaded so much of their lives that they no longer have suitable guarantees about their privacy. As a matter of fact, many aspects of new information technologies render privacy protection difficult to put into practice. A lot of data communications already take place nowadays on the Internet without the users’ notice and the situation is going to get worse with the advent of “ambient intelligence” or “pervasive computing” \cite{19}. One of the most challenging privacy issues in this context is to reconcile this continuous flow of data with privacy protection. One possible option to improve the situation when data has to be disclosed (or when it is practically impossible to object to its disclosure) is to enhance the obligations of the controllers and enforce more stringent rules on the use of personal data. We have followed this approach, considering both
• technical means to define and enforce obligations and
• possible evolutions of data protection regulations to avoid discriminations based on the use of personal data.

Technical means: specification and a posteriori verification of obligations
A major challenge for the formalization of privacy policies is the integration of deontic and temporal operators. Deontic operators are required because privacy policies are typically expressed in terms of obligations and interdictions. Temporal operators are necessary because obligations and interdictions usually come with deadlines: for example, the controller must inform the data subject before forwarding his data to a third party or must delete the data within a given period of time. On the theoretical side, the limitations of Standard Deontic Logic (SDL) have constantly been pointed out, almost since its introduction. However, no other unified mathematical formalization of this logic has been proposed so far. Instead, many specialized logics have been put forward, each aimed at addressing one particular issue. To address this challenge, we have proposed a language called FLAVOR (Formal Language for A posteriori Verification Of legal Rules) for the expression of privacy policies and, more generally, obligations to be fulfilled by organizations. Indeed, organizations have to comply with a growing number of legal rules stemming from law, regulations, corporate policies or contractual agreements. Generally speaking, the actions to be monitored can be checked either a priori or a posteriori. A priori checks are stronger in the sense that they make it possible to ensure that no breach will occur. However, they are too constraining, if not inapplicable, in many situations. Even when they could be implemented, a priori checks are not desirable in situations in which it could be legitimate to bypass the rules. For instance, it is necessary to provide emergency procedures to access personal health records when human lives are at stake, even if the medical practitioner on duty does not have sufficient permissions. The essential features provided by FLAVOR are the possibility to express “contrary to duty” obligations (substitute obligations to be fulfilled in case of breach of the primary obligation), obligations with deadlines and contextual obligations. We have defined a strength ordering between obligations and illustrated the language with typical privacy policy rules [10]. We have also considered the delegation of obligations between actors in [8] and studied the impact of delegation on different types of responsibilities (causal, functional, legal).

Legal means: privacy and non discrimination
In order to address the new threats to individual rights that are made possible by the progress of information technologies, we have proposed to distinguish two very different types of data collection [9]:

1. The collection of data as part of formal procedures with clearly identified parties or in the course of clearly identified events, recognized as such by the subjects (e.g. when submitting a file, filling a questionnaire, using a smart card or providing one’s fingerprint to get access to a building).
2. The apparently insignificant and almost continuous collection of data that will become more and more common in the digital society (digital audit trails, audio and video recordings, etc.). This collection may be more or less perceived or suspected by the subject or remain completely invisible and unsuspected. Another worrying phenomenon is the automatic generation of new knowledge using data mining and knowledge inference techniques. In this kind of situation, the subject may ignore not only the process but also the generated knowledge itself, even if this knowledge is about him and could be used to take actions affecting him (e.g. not offering him a job or an insurance contract or adjusting the price of a service up to the level he would be prepared to pay).

The regulations on personal data protection were originally designed to address the first type of situation. Efforts are made to adapt them to the complex issues raised by the second type of data collection but they tend to be increasingly ineffective in these situations. The main cause of this ineffectiveness is their underlying philosophy of a priori and procedural controls. Starting from this observation, we have argued that a possible option is to strengthen a posteriori controls on the use of personal data and to ensure that the victims of data misuses can get compensations which are significant enough to represent a deterrence for data controllers. We have also argued that the consequences of such misuses of personal data often take the form of unfair discriminations and this trend is likely to increase with the generalization of the use of profiles. For this reason, we advocate the establishment of stronger connections between anti-discrimination and data protection laws,
in particular to ensure that any data processing resulting in unfair differences of treatments between individuals is prohibited and is subject to effective compensations and sanctions [9].
6. New Results

6.1. Dependable Distributed Real-time Embedded Systems

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

6.1.1. The TSH multi-criteria scheduling heuristic

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 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]. By running TSH on a single problem instance, we are able to provide the Pareto front for this instance in 3D, therefore exposing the user to several tradeoffs between the power consumption, the reliability and the execution time. Thanks to extensive simulation results, we have shown how TSH behaves in practice. Firstly, we have compared TSH versus an optimal Mixed Linear Integer Program on small instances; the experimental results show that TSH behaves very well compared to the ILP. Secondly, we have compared TSH versus the ECS heuristic (Energy-Conscious Scheduling [84]); 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.

6.1.2. Automating the Addition of Fault Tolerance with Discrete Controller Synthesis

In collaboration with Emil Dumitrescu (INSA Lyon), Hervé Marchand (VERTECS team from Rennes), and Eric Rutten (SARDES team from Grenoble), we have defined a complete framework for the automatic design of fault tolerant embedded systems, based on discrete controller synthesis (DCS) [88]. Its interest lies in the ability to obtain automatically systems satisfying by construction formal properties specified a priori. Our aim is to demonstrate the feasibility of this approach for fault tolerance. We start with a fault intolerant program, modeled as the synchronous parallel composition of finite labeled transition systems. We specify formally a fault hypothesis, state fault tolerance requirements and use DCS to obtain automatically a program having the same behavior as the initial fault intolerant one in the absence of faults, and satisfying the fault tolerance requirements under the fault hypothesis. Our original contribution resides in the demonstration that DCS can be elegantly used to design fault tolerant systems, with guarantees on key properties of the obtained system, such as the fault tolerance level, the satisfaction of quantitative constraints, and so on. We have shown with numerous examples taken from case studies that our method can address different kinds of failures (crash, value, or Byzantine) affecting different kinds of hardware components (processors, communication links, actuators, or sensors). Besides, we have shown that our method also offers an optimality criterion very useful to synthesize fault tolerant systems compliant to the constraints of embedded systems, like power consumption or execution times. In summary, our framework for fault tolerance has the following advantages [67]:

- The automation, because DCS produces automatically a fault tolerant system from an initial fault intolerant one.
- The separation of concerns, because the fault intolerant system can be designed independently from the fault tolerance requirements.
- The flexibility, because, once the system is entirely modeled, it is easy to try several fault hypotheses, several environment models, several fault tolerance goals, several degraded modes, and so on.
• The safety, because, in case of positive result obtained by DCS, the specified fault tolerance properties are guaranteed by construction on the controlled system.

• The optimality when optimal synthesis is used, modulo the potential numerical equalities (hence a non strict optimality). We consider weights cumulated along bounded-length paths. We have adapted our models in order to take into account the additive costs of, e.g., execution time or power consumption, and adapting synthesis algorithms in order to support the association of costs with transitions, and the handling of these new cost functions in the optimal synthesis [59].

We therefore combine, on the one hand, guarantees on the safety of the execution by tolerating faults, and on the other hand, guarantees on the worst cumulated consumption of the resulting dynamically reconfiguring fault tolerant system. Recently, we have incorporated multi-criteria optimization results in this work, to take into account several weight functions: for instance the execution costs of several tasks, the execution of which must be controlled thanks to DCS. We therefore propose several synthesis algorithms, to aggregate the costs into a single cost function, to hierarchize the costs (e.g., to reflect the priorities of the tasks), or to compute the Pareto front of non-dominated solutions.

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. [92] 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 [88]. The first attempts to compute discrete abstractions for hybrid systems were based on traditional systems behavioral relationships such as simulation or bisimulation, initially proposed for discrete systems most notably in the area of formal methods. These notions require inclusion or equivalence of observed behaviors which is often too restrictive when dealing with systems observed over metric spaces. For such systems, a more natural abstraction requirement is to ask for closeness of observed behaviors. This leads to the notions of approximate simulation and bisimulation introduced in [63].

These 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 [22] we have presented a novel class of multiscale discrete abstractions for incrementally stable switched systems that allows us to deal with fast switching while keeping the number of states in the abstraction at a reasonable level. We assume that the controller of the switched system has to decide the control input and the time period during which it will be applied before the controller executes again. In this context, it is natural to consider abstractions where transitions have various durations. For transitions of longer duration, it is sufficient to consider abstract states on a coarse lattice. For transitions of shorter duration, it becomes necessary to use finer lattices. These finer lattices are effectively used only on a restricted area of the state-space where the fast switching occurs.

These abstractions allow us to use multiscale iterative approaches for controller synthesis as follows. An initial controller is synthesized based on the dynamics of the abstraction at the coarsest scale where only transitions of longer duration are enabled. An analysis of this initial controller allows us to identify regions of the state-space where transitions of shorter duration may be useful (e.g., to improve the performance of the controller). Then, the controller is refined by enabling transitions of shorter duration in the identified regions. The last two steps can be repeated until we are satisfied with the obtained controller.
In [21] we propose a technique for the synthesis of safety controllers for switched systems using multi-scale abstractions. We present a synthesis algorithm that exploits the specificities of multi-scale abstractions. 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 provide experimental results that show drastic improvements of the complexity of controller synthesis using multi-scale abstractions instead of uniform abstractions.

6.2.2. Modular Discrete Controller Synthesis

Discrete controller synthesis (DCS) [88] 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 [57]. This work is done in collaboration with Hervé Marchand (VERTECS team from Rennes) and Éric 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.

This language has been used in different contexts. In [15], BZR is used for the generation of discrete handlers of real-time continuous control tasks, in the framework of the ORCCAD tool. BZR has also been used in a case-study of a Fractal designed HTTP server [19]. The purpose of the synthesized controller is to control the automatic reconfigurations of the system (e.g., start of new components, migrations of some components from one computing element to another), in order to preserve some properties (either functional, e.g., exclusivity of activities of two components, or non-functional, e.g., bounded overall load of the system).

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 [45]). 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

---

31 Open Robot Controller Computer-Aided Design
the same location. Compared to the PhD of Gwenaël Delaval [55], [56], the goal is to achieve modular distribution. By modular, we mean that we want to compile each function of the program into a single function capable of running on any computing location. We make use of our uniform type system to express the computing locations as first-class abstract types, exactly like clocks, which allows us to compile a typed variable (typed by both the clock and the location calculi) into \texttt{if \ldots then \ldots else \ldots} 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, preserving only explicitly specified properties, we have developed a quantitative clock calculus to (1) describe timing properties of the architecture’s clock domain, and (2) describe the properties of the synchronous program to be preserved. The clock calculus is inspired by the network calculus [83], 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.

We aim to extend the analysis to account for shared resources and synchronization protocols, so as to distribute synchronous programs preserving specified properties.

### 6.4. New Programming Languages for Embedded Systems

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

#### 6.4.1. The DSystemJ programming language

In collaboration with Avinash Malik (IBM Watson) and Zoran Salcic (University of Auckland), we have designed the \textsc{SystemJ} programming language [9], which implements the Globally Asynchronous Locally Synchronous (GALS) Model of Computation (MoC) over JAVA. In a nutshell, \textsc{SystemJ} uses the notion of clock domains (CD) to design portions of the system that must operate at unrelated clocks. CDs communicate with each other via asynchronous rendez-vous. Then, a CD consists of one or several reactions, which react synchronously in lock-step and communicate with each other via synchronous broadcast of signals. Finally, all the data computations are implemented in JAVA.

We have further extended \textsc{SystemJ} to allow programmers to design dynamic GALS systems: this is the new language \textsc{DSystemJ} [27], [12], aimed at dynamic distributed systems that use socket based communication protocols for communicating between components. \textsc{DSystemJ} allows the creation and control at runtime of CDs, their mobility on a distributed execution platform, as well as the runtime reconfiguration of the system’s functionality and topology. We have defined the formal semantics of \textsc{DSystemJ}, based on the Dynamic GALS MoC: it offers very safe mechanisms for implementation of distributed systems, as well as potential for their formal verification. The runtime support is implemented in the \textsc{SystemJ} language, which can as such be considered as a static subset of \textsc{DSystemJ}.

This work has been done within the AFMES associated team with the Electric and Computer Engineering Department of the University of Auckland.
6.4.2. The PRET-C programming language for time-predictable systems

Typical safety critical embedded applications, ranging from complex aircraft flight controllers to embedded health devices require worst case guarantees on their timing behavior. The problem is that general-purpose processors, being highly speculative, are intrinsically non-deterministic, and thus are not ideally suited for implementing such systems: either the computed worst-case execution time is highly pessimistic, or heroic efforts are required to accurately model the caches, pipeline, and speculative execution [93]. For similar reasons, using an RTOS to guarantee the determinism of a program’s behavior, along with temporal guarantees, is not feasible. The ability to analyze temporal bounds is dependent on the selected programming language, compiler tool chain, operating system, and the target hardware.

To alleviate these problems, we have defined a synchronous variant of C called PRET-C, together with Sidharta Andalam and Partha Roop (University of Auckland). PRET-C offers constructs for reactive inputs/outputs; it supports a notion of logical time, synchronous concurrency, and preemption [40]. We have also designed the ARPRET architecture for efficient and predictable execution of PRET-C. ARPRET inherits from the long lasting research effort on reactive processors conducted at the University of Auckland. Finally, all timing constraints are precisely verified using a Worst Case Reaction Time (WCRT) analyzer. While there has been a considerable body of work on the timing analysis of procedural programs [93], such analysis for synchronous programs has received less attention. Current state-of-the-art analyses for synchronous programs use integer linear programming (ILP) combined with path pruning techniques to achieve tight results. These approaches first convert a concurrent synchronous program into a sequential program. ILP constraints are then derived from this sequential program to compute the longest tick length. For PRET-C, we have proposed an alternative approach based on model checking [16]. Unlike conventional programs, synchronous programs are concurrent and state-space oriented, making them ideal for model checking based analysis. Our analysis of the abstracted state-space of the program is combined with expressive data-flow information, to facilitate effective path pruning. We have demonstrated through extensive experimentation that the proposed approach is both scalable and about 67% tighter compared to the existing approaches (namely Protothreads [60] and SC [94]).

This overall framework provides an ideal platform for designing and verifying precision timed real-time systems. It has been conducted within the AFMES associated team with the Electric and Computer Engineering Department of the University of Auckland, and is the topic of the PhD of Sidharta Andalam.

6.4.3. Analysis and Scheduling of Parametric Data-Flow 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) model of computation (MoC) for dynamic streaming applications [23], [32], [36], [34], [35]. SPDF extends the standard data flow model by allowing rates to be parametric (e.g., of the form $2xy$). 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 considering other kinds of analyses for this new data-flow MoC. The objective of these analyses is to generate distributed schedules optimizing both the power consumption and the execution time of applications. The targeted hardware is P2012, a new embedded many-core platform designed by STMicroelectronics consisting of several clusters (9 in the current implementation) interconnected through a 2D mesh asynchronous NoC. Each cluster comprises 16 identical computing cores and is equipped with
a hardware mechanism for DVFS (dynamic voltage and frequency scaling). As a first step, we have studied
energy efficient scheduling of simple data-flow graphs for that platform \[ 81 \]. The next step is to extend the
approach to SPDF.

This line of research will be followed in the PhD thesis of Vagelis Bebelis which has just started. It will be
conducted in collaboration with STMicroelectronics.

6.4.4. Translating Hybrid Data-Flow Languages to Hybrid Automata

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 SIMULINK, 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 § 8.1.4 ), we studied how to translate the
ZELUS hybrid data-flow language \[ 43 \] 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 has been accepted to
the conference HSCC’2012 (Hybrid Systems: Computation and Control).

This work is part of the PhD thesis of Peter Schrammel.

6.5. Static Analysis and Abstract Interpretation

Participants: Alain Girault, Bertrand Jeannet [contact person], Lies Lakhdar-Chaouch, Peter Schrammel,
Pascal Sotin.

6.5.1. Numerical and logico-numerical abstract acceleration

Acceleration methods are used for computing precisely the effects of loops in the reachability analysis of
counter machine models. Applying these methods to synchronous data-flow programs with Boolean and
numerical variables, e.g., LUSTRE programs, firstly requires the enumeration of the Boolean states in order
to obtain a control graph with numerical variables only. Secondly, acceleration methods have to deal with the
non-determinism introduced by numerical input variables.

Concerning the latter problem, we pushed further the work presented in \[ 90 \] that extended the concept of
abstract acceleration of Gonnord et al. \[ 69 \], \[ 68 \] to numerical input variables, and we wrote a journal
version \[ 13 \]. The original contributions of \[ 13 \] compared to \[ 91 \] is abstract backward acceleration (for
backward analysis) and a detailed comparison of the abstract acceleration approach with the derivative closure
approach of \[ 39 \], which is related to methods based on transitive closures of relations.

We then worked more on the first point, which is to apply acceleration techniques to data-flow programs
without resorting to an exhaustive enumeration of Boolean states. To this end, we introduced (1) logico-
numerical abstract acceleration methods for CFGs with Boolean and numerical variables and (2) partitioning
methods that make logical-numerical abstract acceleration effective. Experimental results showed that
incorporating these methods in a verification tool based on abstract interpretation provides not only significant
advantage in terms of accuracy, but also a gain in performance in comparison to standard techniques. This
work was published in \[ 28 \].

This line of work is part of the PhD thesis of Peter Schrammel.

6.5.2. Improving dynamic approximations in static analysis

Abstract interpretation \[ 51 \] formalizes two kind of approximations that can be done in the static analysis of
programs:
6.5.2.2. Policy Iteration.

6.5.2.1. Widening with thresholds.

This work was published in [30].

we provided an efficient method to integrate the concept of policy in the logico-numerical abstract domain.

In particular we considered the verification of programs manipulating Boolean and numerical variables, and of Policy Iteration in static analysis to be considerably extended.

We first investigated the integration of the concept of Policy Iteration in a generic way into existing numerical abstract domains. We implemented it in the A library (see module 5.4). This allows the applicability of Policy Iteration in static analysis to be considerably extended.

The problem addressed here is that the extrapolation performed by widening often loses crucial information for the analysis goal.

6.5.2.1. Widening with thresholds.

A classical technique for improving the precision is “widening with thresholds”, which bounds the extrapolation. The idea is to parameterize the extrapolation operator called widening and denoted with $\nabla$. This results in a sequence $Z_0 = \bot_A, Z_{n+1} = \nabla G(Z_n)$ that converges to a post-fixpoint $Z_\infty \sqsupseteq G(Z_\infty)$. For instance, for many numerical abstract domains (like octagons [86] or convex polyhedra [75]) the “standard” widening $\nabla : A \times A \rightarrow A$ consists in keeping in the result $R = P\nabla Q$ the numerical constraints of $P$ that are still satisfied by $Q$.

The benefit of this refinement strongly depends on the choice of relevant thresholds. In [33], [26] we proposed a semantic-based technique for automatically inferring such thresholds, which applies to any control graph, be it inprocedural, interprocedural or concurrent, without specific assumptions on the abstract domain. Despite its technical simplicity, we showed that our technique is able to infer the relevant thresholds in many practical cases.

6.5.2.2. Policy Iteration.

Another direction we investigated for solving the fix-point equation $Y = G(Y)$, $Y \in A$ is the use of Policy Iteration, which is a method for the exact solving of optimization and game theory problems, formulated as equations on min max affine expressions. In this context, a policy $\pi$ is a strategy for the min-player, which gives rise to a simplified equation $X = F^\pi(X)$, $F^\pi \geq F, X \in \mathbb{R}^n$ which is easier to solve that the initial equation $X = F(X), X \in \mathbb{R}^n$. Policy iteration iterates on policies rather than iterating the application of $F$ (as in Kleene iteration), using the property that the least fixpoint of $F$ corresponds to the least fixpoint of $F^\pi$ for some $\pi$.

[50] showed that the problem of finding the least fixpoint of semantic equations on some abstract domains can be reduced to such equations on min max affine expressions, that can then be solved using Policy Iteration instead of the traditional Kleene iteration with widening described above.

We first investigated the integration of the concept of Policy Iteration in a generic way into existing numerical abstract domains. We implemented it in the APRON library (see module 5.4). This allows the applicability of Policy Iteration in static analysis to be considerably extended.

In particular we considered the verification of programs manipulating Boolean and numerical variables, and we provided an efficient method to integrate the concept of policy in the logico-numerical abstract domain BDDAPRON that mixes Boolean and numerical properties (see module 5.4). This enabled the application of the policy iteration solving method to much more complex programs, that are not purely numerical any more. This work was published in [30].
6.5.3. Analysis of imperative programs

We also studied the analysis of imperative programs. Even if it is preferable to analyze embedded systems described in higher-level languages such as synchronous languages, it is also useful to be able to analyze C programs. Moreover, it enables a wider diffusion of the analysis techniques developed in the team.

6.5.3.1. Inferring Effective Types for Static Analysis of C Programs

This work is a step in the project of connecting the C language to our analysis tool INTERPROC/CONCURINTERPROC (see section 5.5.4). The starting point is the connection made by the industrial partner EADS-IW in the context of the ANR project ASOPT (§ 8.1.2) from a subset of the C language to INTERPROC. This translation uses the NEWSPEAK intermediate language promoted by EADS [77].

```c
typedef struct {
    int n;
} t;

int main()
{
    t x; t* y;
    int *p,*q;
    y = alloc(t); p = &(y->n);
    y = &x; q = &(y->n);
    *p = 1; *q = 2; *p = *p < 1;
    return *p;
}
```

Initial program.

```c
typedef enum {
    10=0,11=1,12=2
} e;

t typedef struct {
    e n;
} t;

int main()
{
    t x; t* y;
    e *p,*q;
    y = alloc(t); p = &(y->n);
    y = &x; q = &(y->n);
    *p = l1; *q = l2; *p = (*p==l0)?l1:l0;
    return *p;
}
```

Transformed program.

*Figure 3. Inferring finite types in C programs*

The problem addressed here is that the C language does not have a specific Boolean type: Boolean values are encoded with integers. This is also true for enumerated types, that may be freely and silently cast to and from integers. On the other hand, our verification tool INTERPROC that infers the possible values of variables at each program point may benefit from the information that some integer variables are used solely as Boolean or as enumerated type variables, or more generally as finite type variables with a small domain. Indeed, specialized and efficient symbolic representations such as BDDs are used for representing properties on such variables, whereas approximated representations like intervals and octagons are used for larger domain integers and floating-points variables.

Driven by this motivation, we proposed in [25] a static analysis for inferring more precise types for the variables of a C program, corresponding to their effective use. The analysis addresses a subset of the C99 language, including pointers, structures and dynamic allocation. The principle of the method is very different from type inference techniques used in functional programming languages such as ML, where the types are inferred from the context of use. Instead, our analysis can be seen as a simple points-to analysis, followed by a disjunction version of a constant propagation analysis, and terminated by a program transformation that generates a strongly typed program. Fig. 3 illustrates this process. On this example, we discover that the program is a finite-state one, to which exact analysis technique can be applied.

6.5.3.2. Interprocedural analysis with pointers to the stack

This work addresses the problem of interprocedural analysis when side-effect are performed on the stack containing local variables. Indeed, in any language with procedures calls and pointers as parameters (C, Ada) an instruction can modify memory locations anywhere in the call-stack. The presence of such side effects...
breaks most generic interprocedural analysis methods, which assume that only the top of the stack may be modified. In [29] we presented a method that addresses this issue, based on the definition of an equivalent local semantics in which writing through pointers has a local effect on the stack. Our second contribution in this context is an adequate representation of summary functions that models the effect of a procedure, not only on the values of its scalar and pointer variables, but also on the values contained in pointed memory locations. Our implementation in the interprocedural analyzer PINTERPROC (see §5.5.4) results in a verification tool that infers relational properties on the value of Boolean, numerical, and pointer variables.

6.6. Component-Based Construction

Participants: Lacramioara Astefanoaei, 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. Causality Analysis in Contract Violation

Establishing liabilities in case of litigation is generally a delicate matter. It becomes even more challenging when IT systems are involved. Generally speaking, a party can be declared liable for a damage if a fault can be attributed to that party and this fault has caused the damage. The two key issues are thus to establish convincing evidence with respect to (1) the occurrence of the fault and (2) the causality relation between the fault and the damage. The first issue concerns the technique used to log the relevant events of the system and to ensure that the logs can be produced (and have some value) in court. The second issue is especially complex when several
faults are detected in the logs and the impact of these faults on the occurrence of the failure has to be assessed. In [6] we have focused on this second issue and proposed a formal framework for reasoning about causality. A system based on this framework could be used to provide relevant information to the expert, the judge, or the parties themselves (in case of amicable settlement) to analyze the origin of the failure of an IT system.

The notion of causality has been studied for a long time in computer science, but with very different perspectives and goals. In the distributed systems community, causality (following Lamport’s seminal paper [82]) is seen essentially as a temporal property. In our context, the temporal ordering contributes to the analysis, but it is obviously not sufficient to establish the logical causality required to rule on a matter of liability: the fact that an event $e_1$ has occurred before an event $e_2$ does not imply that $e_1$ was the cause for $e_2$ (or that $e_2$ would not have occurred if $e_1$ had not occurred).

Our formal model is based on components interacting according to well identified interaction models [5]. Each component is associated with an individual contract which specifies its expected behavior. The system itself is associated with a global contract which is assumed to be implied by the composition of the individual contracts.

In [6] we have defined several variants of logical causality. The first variant, necessary causality, characterizes cases when the global contract would not have been violated if the local contract had been fulfilled. The second variant, sufficient causality, characterizes cases when the global contract would have been violated even if all the other components had fulfilled their contracts. In other words, the violation of its contract by a single component was sufficient to violate the global contract.

We are currently extending to framework to other models of computation and communication, in particular, to timed automata.

### 6.6.3. Realizability of Choreographies for Services Interacting Asynchronously

Choreography specification languages describe from a global point of view interactions among a set of services in a system to be designed. Given a choreography specification, the goal is to obtain a distributed implementation of the choreography as a system of communicating peers. These peers can be given as input (e.g., obtained using discovery techniques) or automatically generated by projection from the choreography. Checking whether some set of peers implements a choreography specification is called realizability. This check is in general undecidable if asynchronous communication is considered, that is, services interact through message buffers.

In [24] we consider conversation protocols as a choreography specification language, and leverage a recent decidability result [54] to check automatically the realizability of these specifications by a set of peers under an asynchronous communication model with a priori unbounded buffers.

### 6.6.4. A Theory of Fault Recovery for Component-Based Models

In [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 define 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. We identify corrector (atomic or composite) components whose presence in a model is essential to guarantee recovery after the occurrence of faults. We also formalize component based models that effectively separate recovery from functional concerns.

### 6.7. Aspect-Oriented Programming

**Participants:** Henri-Charles Blondeel, Pascal Fradet [contact person], Alain Girault, Marnes Hoff.

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 [10]. Our categories of aspects comprise, among others, observers, abortors, 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 ASCOLAINRIA 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 Synchrone. 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 $k$ faults within $n$ clock signals”. The language semantics as well as the fault model have been formalized in Coq. The next step is to express standard fault tolerance techniques as program transformations and prove that they allow to tolerate all faults of a given model.

6.7.3. Refinement Aspects

Chemical programming describes computation in terms of a chemical solution in which molecules (representing data) interact freely according to reaction rules (representing the program). Solutions are represented by multisets of elements and reactions by rewrite rules which consume and produce new elements according to conditions. This paradigm makes it possible to express programs without artificial sequentiality in a very abstract way. It bridges the gap between specification and implementation languages.

A drawback of chemical languages is that their very high-level nature usually leads to very inefficient programs. We have proposed a refinement oriented approach where the basic functionality is expressed as a chemical program whereas efficiency is achieved separately by:

- structuring the multiset with a data type defining neighborhood relations;
- describing the selection of elements according to their neighborhood;
- specifying the evaluation strategy (i.e., the application of rules and termination).
Using these three implementation aspects (data structure, selection and strategy), the chemical program can then be refined automatically into an efficient low-level program. The crucial methodological advantage is that logical issues are decoupled from efficiency issues.

This research, that takes place within the AUTOCHEM project (see Section 8.1.1), is done in collaboration with Jean-Louis Giavitto (Ircam, Paris). It is the subject matter of Marnes Hoff’s PhD thesis.
6. New Results

6.1. Models and Verification Techniques

6.1.1. The BCG Format and Libraries

Participants: Hubert Garavel, Frédéric Lang, Wendelin Serwe.

BCG (Binary-Coded Graphs) is both a file format for the representation of explicit graphs and a collection of libraries and programs dealing with this format. Version 1.0 of the BCG format was recently replaced by version 1.1, which can exploit the capabilities of 64-bit addressing.

In 2011, we continued to enhance the BCG libraries as follows:

- We extended the BCG_READ application programming interface with three new primitives so as to increase symmetry with the OPEN/CÆSAR application programming interface (see § 6.1.2).
- We fixed a memory corruption problem occurring with very long label strings; this problem would cause random crashes of the DISTRIBUTOR tool (see § 6.1.5).

6.1.2. The OPEN/CÆSAR and CÆSAR_SOLVE Libraries

Participants: Iker Bellicot, Hubert Garavel, Yann Genevois, Frédéric Lang, Radu Mateescu, Wendelin Serwe.

OPEN/CÆSAR is an extensible, modular, language-independent software framework for exploring implicit graphs. This key component of CADP is used to build simulation, execution, verification, and test generation tools.

In 2011, a bug in the CÆSAR_TABLE library has been corrected, which would cause segmentation faults when certain primitives of the Application Programming Interface were invoked on a bounded table.

CÆSAR_SOLVE is a generic software library based on OPEN/CÆSAR for solving boolean equation systems of alternation depth 1 (i.e., without mutual recursion between minimal and maximal fixed point equations) on the fly. This library is at the core of several CADP verification tools, namely the equivalence checker BISIMULATOR, the minimization tool REDUCTOR, and the model checkers EVALUATOR 3.5 and 4.0. The resolution method is based on boolean graphs, which provide an intuitive representation of dependencies between boolean variables, and which are handled implicitly, in a way similar to the OPEN/CÆSAR interface [4].

In 2011, we improved the parallel resolution algorithm of CÆSAR_SOLVE (see § 6.1.5).

6.1.3. The EVALUATOR Tool

Participants: Iker Bellicot, Hubert Garavel, Yann Genevois, Radu Mateescu.

EVALUATOR is a model checker that evaluates a temporal logic property on a graph represented implicitly using the OPEN/CÆSAR environment. EVALUATOR works on the fly, meaning that only those parts of the implicit graph relevant to verification are explored. The model checking problem is reformulated in terms of solving a boolean equation system. A useful feature of EVALUATOR is the generation of diagnostics (examples and counterexamples) explaining why a formula is true or false.

In version 3.5 of EVALUATOR, properties are described in regular alternation-free μ-calculus, a logic built from boolean operators, possibility and necessity modalities containing regular expressions denoting transition sequences, and fixed point operators without mutual recursion between least and greatest fixed points. The input language of the tool also enables the user to define parameterized temporal operators and to group them into separate libraries.
In version 4.0 of EVALUATOR (5,000 lines of SYNTAX code, 40,500 lines of LOTOS NT code, and 13,100 lines of C code), properties are written in MCL (Model Checking Language) [ 18 ], an extension of the regular alternation-free μ-calculus of EVALUATOR 3.5 with data-handling and fairness operators. In particular, EVALUATOR 4.0 can handle modalities and fixed point operators with data parameters, regular expressions extended with counters, operators inspired from programming languages (“if-then-else”, “for”, etc.), and operators (of alternation depth two) allowing to characterize complex infinite sequences.

In 2011, we continued the extensive testing of EVALUATOR 3.5 and 4.0 using our test base of 10,000 BCG graphs and 3,800 MCL formulas. This revealed three errors in EVALUATOR 4.0, which have been corrected.

We also brought the following enhancements to MCL and EVALUATOR 4.0:

- We extended the set of MCL operators of alternation depth two with parameterized versions of the infinite looping and saturation operators, which allow to succinctly encode the presence (respectively, the absence) of accepting cycles in generalized Büchi automata. The evaluation of these parameterized operators is translated into parameterized boolean equation systems, which are instantiated into plain boolean equation systems and solved on the fly using the algorithms A3 and A4 (extended with marked cycle detection) of the CÆSAR_SOLVE library. This evaluation procedure has a complexity linear in the size of the degeneralized Büchi automaton, which is represented by the boolean equation system obtained after instantiation.

- We enhanced MCL by adding a data type for manipulating sets of natural numbers. This data type, equipped with the classical set operations (union, intersection, difference, insertion, deletion, membership, etc.), enables a succinct specification of temporal properties referring to the past, such as the fact that a certain set of events (represented by natural numbers) occurred on the transition sequences leading from the initial state to the current state.

- We added a new option to EVALUATOR 4.0 for displaying a set of regular expressions that over-approximate the set of visible actions (transition labels in the LTS) satisfying the action predicates occurring in the MCL formula. This feature enables to improve the efficiency of verification by hiding the set of actions other than those produced from the MCL formula, minimizing the LTS modulo a weak equivalence relation compatible with the formula, and then verifying the MCL formula on the minimized LTS. This may increase the efficiency of verification by one order of magnitude, as reported in [ 40 ].

EVALUATOR 4.0 was officially integrated in CADP in March 2011. MCL and EVALUATOR 4.0 were used successfully for analyzing mutual exclusion protocols (see § 6.3.1 ) and hardware architectures (see § 6.3.2 ).

6.1.4. Compositional Verification Tools

**Participants:** Frédéric Lang, Radu Mateescu.

The CADP toolbox contains various tools dedicated to compositional verification, among which EXP.OPEN 2.1, PROJECTOR 3.0, BCG_MIN 2.0, and SVL 2.2 play a central role. EXP.OPEN explores on the fly the graph corresponding to a network of communicating automata (represented as a set of BCG files). PROJECTOR implements behavior abstraction [ 58 ], [ 64 ] by taking into account interface constraints. BCG_MIN minimizes behavior graphs modulo strong or branching bisimulation and their stochastic extensions. SVL (Script Verification Language) is both a high level language for expressing complex verification scenarios and a compiler dedicated to this language.

In 2011, we corrected one bug in PROJECTOR, two bugs in EXP.OPEN, and four bugs in SVL. We also enhanced these tools as follows:

- The generalized parallel composition operator proposed in [ 13 ], including the support for “m among n” synchronization, has been added to SVL, leading to version 2.3 of SVL.

- Together with Pepijn Crouzen (Saarland University), we pursued our work on the so-called “smart reduction” techniques for compositional verification. An article about smart reduction was published in an international conference [ 32 ].
We improved smart reduction for stochastic branching bisimulation so as to cut stochastic transitions (according to the “maximal progress” assumption) as early as possible in intermediate parallel compositions, thus yielding state space reductions. With this new optimization, EXP.OPEN detects when some action (usually, an output) offered by some process can synchronize with corresponding actions (usually, inputs) offered by the other processes in all their states; if so, all stochastic transitions in choice with this action can safely be cut in every intermediate composition. This situation occurs frequently with Input/Output Interactive Markov Chains.

Additionally, we studied an alternative compositional verification approach named partial model checking \cite{47}. Given a temporal logic formula $\varphi$ to be evaluated on a set $S$ of concurrent processes, partial model checking consists in transforming $\varphi$ into another equivalent formula $\varphi'$ to be evaluated on a subset of $S$. Formula $\varphi'$ is constructed incrementally by choosing one process $P$ in $S$ and incorporating into $\varphi$ 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.

We developed a prototype implementation of this approach using the generic software components of CADP:

- We extended the definition of quotienting given by \cite{47} to support all features of the input language of EXP.OPEN 2.1, which enables networks of labeled transition systems to be described using parallel composition operators borrowed from various process algebras such as CCS, CSP, and LOTOS, including also LOTOS NT parallel composition and “$m$ among $n$” synchronisation operators \cite{13}.
- We gave an executable definition of quotienting in terms of a synchronous product between a graph representation (called formula graph) of the formula $\varphi$ and the process $P$, thus enabling quotienting to be implemented efficiently in EXP.OPEN. We extended EVALUATOR 3.5 to automatically generate the formula graph corresponding to a temporal logic formula.
- We proposed and implemented efficient formula simplifications by combining reductions modulo bisimulations and partial formula evaluation computed using boolean equation systems.

We used this prototype implementation to verify 28 temporal logic properties on the TFTP avionics protocol \cite{14}. For several of these properties, partial model checking uses hundreds of times less memory than on-the-fly model checking using EVALUATOR. This work led to a publication in an international conference \cite{38}.

6.1.5. Parallel and Distributed Verification Tools

Participants: Iker Bellicot, Hubert Garavel, Rémi Hérilier, Radu Mateescu.

The CADP toolbox contains several components designed to take advantage of distributed computing facilities (such as clusters of machines) to perform large-scale verifications, namely: CÆSAR_NETWORK, a network communication library used by the other tools, DISTRIBUTOR and BCG_MERGE, two companion tools \cite{10} that perform reachability analysis using a distributed state space exploration algorithm, and BES_SOLVE \cite{35}, a tool that solves boolean equations systems using the various resolution algorithms provided by the CÆSAR_SOLVE library (see § 6.1.2), including a distributed on-the-fly resolution algorithm.

In 2011, we continued enhancing these tools, taking advantage of the valuable feedback provided by Eric Madelaine (INRIA Sophia Antipolis) who used CADP on the PACAGRID cluster. We brought the following improvements:

- We performed careful code reviews of the CÆSAR_NETWORK library, and corrected nine bugs. We equipped this library with logging primitives that proved to be helpful for debugging distributed algorithms programmed above the CÆSAR_NETWORK library.
- We fixed three bugs in DISTRIBUTOR and two bugs in BCG_MERGE.
- We pursued the intensive testing campaign undertaken in 2010 for BES_SOLVE, using up to 100 concurrent processes running on the PIPOL and GRID5000 platforms. We focused our efforts on the distributed resolution algorithm for boolean equation systems of CÆSAR_SOLVE/BES_SOLVE,
which was tested extensively on examples of boolean equation systems represented explicitly as
text files or generated randomly according to various parameters; the resolution results being cross-
checked against the sequential resolution algorithms provided by CÆSAR _SOLVE/BE_SOLVE.
Several bugs (affecting memory leaks, handling of early termination, diagnostic generation, collect-
ing of statistical information about the resolution) were fixed. Changes were also carried out on the
code to simplify its structure, increase modularity, and improve readability. The convergence of the
distributed resolution algorithm was accelerated by backward propagation of constants as soon as
they have been discovered.

• The EVALUATOR 4.0 tool was extended with a new prototype algorithm allowing the distributed
verification of an MCL formula on a graph using several machines connected by a network. This
functionality was implemented by connecting the tool to the distributed resolution algorithm for
boolean equation systems and experimented out on clusters of machines.

• Finally, we added to CADP four new tools named PBG_CP, PBG_INFO, PBG_MV, and PBG_RM.
These tools respectively enable to copy, query, move, and delete the PBG (Partitioned BCG Graph)
collection of files generated and used by DISTRIBUTOR and BCG_MERGE.

6.1.6. Other Tool Developments
Participants: Hubert Garavel, Yann Genevois, Rémi Hérilier, Frédéric Lang, Radu Mateescu, Wendelin
Serwe.

To support the usage of CADP in industry and academia, we pursued our efforts to master the software quality
of CADP:

• We added support for Mac OS X 10.7 (“Lion”) and enhanced the documentation for Mac OS X.

• We corrected one bug in the INSTALLATOR installation assistant, two bugs in the Tst platform-
checking command, and brought two bug fixes and one usability enhancement in the EUCALYP-
tUS graphical user-interface. We also provided a workaround for supporting recent versions of
UBUNTU.

• We continued building a comprehensive validation framework, based on non-regression testing and
semantical checking for the CADP tools. This framework allows functional testing of individual tools
as well as integration testing for several CADP tools used together to perform complex verification
scenarios on various computing platforms and using various compilers.

Other research teams took advantage of the software components provided by CADP (e.g., the BCG and
OPEN/CÆSAR environments) to build their own research software. We can mention the following develop-
ments:

• the KMELIA tools for component-based systems [ 50 ], developed at the University of Nantes
(France);

• the VERCORS tool for unifying architectural and behavioral specifications of distributed components
[ 52 ], developed at INRIA Sophia-Antipolis;

• the DAMASCO (Discovery, Adaptation and Monitoring of Context-Aware Services and Components)
framework for composition and adaptation based on model transformation [ 54 ], developed at the
University of Málaga (Spain);

• the SCOOP tool for symbolic optimizations of probabilistic processes [ 62 ], [ 74 ], developed at RWTH
Aachen (Germany);

• the SLCO (Simple Language of Communicating Objects) environment [ 49 ], developed at Eind-
hoven University of Technology (The Netherlands);

• the MOTOR tool for probabilistic analysis of embedded systems [ 68 ], developed at the Embedded
Systems Institute (Eindhoven, The Netherlands);

• the ALVIS modeling language for design and verification of embedded systems [ 72 ], developed at
the AGH University of Science and Technology (Krakow, Poland).
6.2. Languages and Compilation Techniques

6.2.1. Compilation of LOTOS

Participants: Hubert Garavel, Wendelin Serwe.

The CADP toolbox contains several tools dedicated to the LOTOS language, namely the CÆSAR.ADT compiler \cite{3} for the data type part of LOTOS, the CÆSAR compiler \cite{11} for the process part of LOTOS, and the CÆSAR.INDENT pretty-printer.

In 2011, in addition to fixing four bugs in the CÆSAR and CÆSAR.ADT compilers, we improved the LOTOS-dedicated tools of CADP as follows:

- We revised the predefined type libraries and C code generated by CÆSAR and CÆSAR.ADT to suppress warnings emitted by recent versions of GCC.
- We enhanced the format in which values of singleton and tuple types are displayed to end users.
- We modified the predefined libraries for natural (unsigned) and integer (signed) types so that users can now indicate the precise number of bits (between 1 and 64) to be used for the machine representation of these types, and can also determine the precise range (lower and upper bound) to be used for values of these types.
- We further modified these two libraries to enable (optional) overflow and underflow checking during arithmetic operations on the natural and integer types.

6.2.2. Compilation of LOTOS NT

Participants: Hubert Garavel, Frédéric Lang, Christine McKinty, Vincent Powazny, Wendelin Serwe.

Regarding the LOTOS NT language — a variant of E-LOTOS created by the VASY project team — we worked along two directions:

- We continued enhancing the TRAIAN compiler (see § 5.2 ), which generates C code from LOTOS NT data type and function definitions. TRAIAN is distributed on the Internet and used intensively within the VASY project team as a development tool for compiler construction \cite{8}.

In 2011, TRAIAN was essentially in maintenance mode. We updated its documentation and added to its predefined library a conversion function that was missing.

- The LNT2LOTOS, LNT.OPEN, and LPP tools convert LOTOS NT code to LOTOS, thus allowing the use of CADP to verify LOTOS NT descriptions. These tools are officially part of CADP since 2010 and have been used successfully for many different systems (see § 6.3.1 , § 6.3.2 , § 6.3.5 , § 6.3.5 , and § 6.3.3 ).

In 2011, we continued enhancing these tools, of which we delivered four new releases. In addition to 13 bug fixes, the following enhancements have been brought:

  - The LOTOS NT language was extended with range types (which are interval subtypes of character, integer, or natural types) and predicate types (which are subtypes of existing types, a boolean predicate being used to determine the domain of each subtype).
  - The LOTOS NT language was enriched with the concept of “module pragmas”, which specify implementation constraints for predefined types such as naturals, integers, and strings. Also, the predefined operations “first” and “last” have been added for enumerated types.
  - The LNT2LOTOS translator was made semantically stricter by adding checks for overflow and underflow when doing natural and integer arithmetics, checking that range type bounds and array type bounds belong to the domain of admissible values for their parent types, and adding additional checks for type pragmas.
– The LOTOS code generated by the LNT2LOTOS translator was optimized by handling equation premises (in the data part) and boolean guards (in the behavior part) that are always false or always true. Other optimizations have been added for process definitions whose bodies are empty or only contain a call to another process, for “case” statements that are followed by no instruction or only a “null” instruction, and for “while” loops with an empty body.
– The speed of processing LOTOS NT specifications containing several modules has been made between two and three times faster.
– The error and warning messages issued by the LOTOS NT tools have been enhanced.
– The reference manual has been corrected, reorganized and comprehensively edited. Two new appendices have been added, one that lists all the predefined functions, and another one (20 pages) giving the formal semantics of LOTOS NT.

6.2.3. Source-Level Translations between Concurrent Languages

Participants: Hubert Garavel, Rémi Hérilier, Frédéric Lang, Radu Mateescu, Gwen Salaün, Wendelin Serwe, Damien Thivolle.

Although process algebras are, from a technical point of view, the best formalism to describe concurrent systems, they are not used as widely as they could be [2]. Besides the steep learning curve of process algebras, which is traditionally mentioned as the main reason for this situation, it seems also that the process algebra community scattered its efforts by developing too many languages, similar in concept but incompatible in practice. Even the advent of two international standards, such as LOTOS (in 1989) and E-LOTOS (in 2001), did not remedy this fragmentation. To address this problem, we started investigating source-level translators from various process algebras into LOTOS or LOTOS NT, so as to widen the applicability of the CADP tools.

In 2011, in addition to the LNT.OPEN tool suite (see § 6.2.2), we worked on the following translators:

- We continued our work on the FLAC tool, which translates a FIACRE program into a LOTOS program automatically, for verification using CADP. In 2011, 2 bugs reported by users of FLAC were corrected. Those corrections led to revisions 74 and 75 of the FLAC code, which is available on the development forge dedicated to FIACRE compilers 1. We collected new examples of FIACRE code to enhance our test suite, which now comprises 79 examples.
- BPEL (Business Process Execution Language) [61] is a language inspired by the π-calculus [67] and standardized by the OASIS consortium (led by IBM and Microsoft) to describe the orchestrations of Web services. BPEL depends on other W3C standard XML-related languages: XML Schema for data types, XPATH for data expressions, and WSDL for declaring the interfaces (communication links and link functions) of a Web service.

Following interest expressed by research teams at MIT and the Polytechnic University of Bucharest, we designed translation rules from BPEL to LOTOS NT in order to formally verify BPEL services with CADP. We began to develop an automated translator.

In 2011, following a remark by Charles Pecheur (Université Catholique de Louvain, Belgium) who spotted an error in the translation of BPEL processes into LOTOS NT, we corrected the translation of exception handling so that it no longer interferes with the atomicity mechanism. The complete translation algorithm is given in Damien Thivolle’s PhD thesis [23]. We pursued the implementation of our BPEL to LOTOS NT translator and finalized the translation of XML Schema types and WSDL definitions.

- We considered the π-calculus [67], a process algebra based on mobile communication. We proposed a general method for translating the finite control fragment of the π-calculus (obtained by forbidding recursive invocations of an agent through parallel composition operators) into LOTOS NT. The mobile communication is encoded using the data types of LOTOS NT, each channel name

1 http://gforge.enseeiht.fr/projects/fiacre-compil
being represented as a value of an enumerated data type. The binary synchronization of π-calculus is enforced by associating a LOTOS NT gate to each parallel composition operator present in the π-calculus specification and by tagging each synchronization with the unique identifiers of the sender and receiver agents. The translation preserves the operational semantics by mapping each transition of a π-calculus agent to a single transition of the resulting LOTOS NT term.

In 2011, we have extended the π-calculus with data-handling features, with the goal of widening its possible application domains. This was done by extending the language grammar and the translation to support typed variables and data expressions. As language for describing data, we chose LOTOS NT: indeed, the data types and functions used in the π-calculus specification can be described in LOTOS NT and directly incorporated to the LOTOS NT code produced by translation. This results in an applied π-calculus, such as the variant of the calculus proposed in [45] for the verification of security properties.

The Ptc2LNT translator was extended accordingly. It now consists of 2,100 lines of SYNTAX code, 3,100 lines of LOTOS NT code, and 500 lines of C code. The tool was tested on 234 examples of π-calculus specifications, including most of the examples provided in the Mobility Workbench distribution.

6.3. Case Studies and Practical Applications

6.3.1. Mutual Exclusion Protocols

Participants: Radu Mateescu, Wendelin Serwe.

Mutual exclusion protocols are an essential building block of concurrent systems to ensure proper use of shared resources in the presence of concurrent accesses. Many variants of mutual exclusion protocols exist for shared memory, such as Peterson’s or Dekker’s well-known protocols. Although the functional correctness of these protocols has been studied extensively, relatively little attention has been paid to their performance aspects.

In 2011, we considered a set of 27 mutual exclusion protocols for up to sixteen processes with a shared memory and coherent local caches. We specified each protocol in LOTOS NT, using a set of generic modules to describe shared variables, the cache protocol, and the overall architectures (in total, 13,600 lines of LOTOS NT code). Then, we compositionally added Markov delays modeling the latencies of read/write accesses on shared variables, so as to obtain the Interactive Markov Chain (IMC) corresponding to each protocol (up to 1.6 billion states and 2.7 billion transitions for the black-white bakery protocol [73] for four processes).

We verified functional properties using the same set of MCL [18] formulas for each protocol (in total, 380 lines of MCL). The mutual exclusion property was easy to express as an MCL formula, but other properties (livelock and starvation freedom, independent progress, and unbounded overtaking) turned out to be quite involved because they belong to the μ-calculus fragment of alternation depth two; fortunately, we succeeded in expressing them using the infinite looping operator of MCL, which can be checked in linear time. In particular, it was challenging to express these properties using a parameter N for the number of processes.

Finally, using the performance evaluation tools of CADP (i.e., BCG_STEADY for small numbers of processes and CUNCTATOR for larger numbers of processes), we computed the steady-state throughputs of critical section accesses by varying several parameters (relative speeds of processes, ratio between the time spent in critical and non-critical sections, etc.).

These experiments enabled us to compare the protocols according to their efficiency (steady-state throughputs) and study also their scalability for an increasing number of processors. We observed that symmetric protocols are more robust when the difference in execution speed between processes is large, which confirms the importance of the symmetry requirement originally formulated by Dijkstra [56]. The quantitative results corroborated those of functional verification: the presence of (asymmetric) starvation of processes, detected using temporal formulas, was clearly reflected in their steady-state throughputs. Our results also corroborate experimental measures found in the literature [48].
6.3.2. The Platform 2012 Architecture

Participant: Wendelin Serwe.

In the context of the MULTIVAL contract (see § 7.1), STMicroelectronics studied PLATFORM 2012, a many-core programmable multi-cluster platform fabric targets a range of emerging video, imaging, and next-generation immersive multimodal applications. Configurability options include the number of clusters, the number and type of processing elements (PE) per cluster, specialization of the architecture and instruction-set of the PEs, and finally, support of hardware-accelerated PEs. The platform is supported by a rich programming environment which embodies a range of platform programming models.

In 2011 we focused on the DTD (Dynamic Task Dispatcher) hardware block that assigns a set of application tasks on a set of PEs. It is called dynamic because each task itself might add tasks to the set of those to be dispatched by the DTD. The DTD is synthesized from a C++ model, optimized to generate an efficient hardware block. Due to the intrinsic complexity of the DTD, STMicroelectronics was interested in the co-simulation of this C++ code with a formal model of the DTD.

In a first step, we generalized the LOTOS NT model of the DTD developed in 2010 to allow the handling of an arbitrary number of PEs (1, 200 lines of LOTOS NT). We also modeled as LOTOS NT processes the different sets of tasks corresponding to various applications. To express the operations provided by the DTD, we had to include a call-stack in the model of each PE, as a means of circumventing the static-control constraints of CÆSAR forbidding recursion over parallel composition. STMicroelectronics judged LOTOS NT to be essential in modeling the DTD, because using LOTOS instead would have been extremely difficult, requiring complex continuations with numerous parameters. We also wrote twelve scenarios (1, 000 lines of LOTOS NT) describing applications to be dispatched by the DTD. For each scenario and for up to six PEs, we generated the corresponding LTS (up to 100 million states and 500 million transitions).

Then, for each generated LTS, we verified several properties, such as the correctness of assertions inserted in the model (by checking the set of transition labels), the termination of the scenario, or that each task is executed exactly once. We expressed the latter properties using the MCL language [18] and verified them using the EVALUATOR 4 model checker. This allowed us to point out a difference between our implementation and the one from the architect, highlighting a high sensibility on the order of terms in an equation, revealing an under-specified mechanism. We also verified the correctness of a complex optimization.

Having gained confidence in the LOTOS NT model, we applied the EXEC/CÆSAR framework to co-simulate the C++ and LOTOS NT models of the DTD, a challenge being the connection of the asynchronous LOTOS NT model with the synchronous C++ model, because one step of the C++ model corresponds, in general, to several transitions of the LOTOS NT model.

This case study enabled us to discover and correct a few bugs in CADP and led to a publication in an international conference [39].

6.3.3. The Self-configuration Protocol

Participant: Gwen Salaün.

Cloud computing emerged a few years ago as a major topic in modern programming. It leverages hosting platforms based on virtualization, and promises to deliver resources and applications that are faster and cheaper with a new software licensing and billing model based on the pay-per-use concept.

Distributed applications in the cloud are composed of a set of virtual machines (VMs) running a set of interconnected software components. However, the task of configuring distributed applications is complex as each VM includes many parameters either for local configuration (e.g., pool size, authentication data) or remote interconnection (e.g., IP address and port to access a server). Existing deployment solutions are often specific to certain applications and rarely take into account these configuration parameters, which are usually managed by dedicated scripts that do not work fully automatically.
Together with Xavier Etchevers, Thierry Coupaye (Orange labs), Fabienne Boyer, and Noël de Palma (INRIA Grenoble), we worked on the verification of an innovative self-configuration protocol [34] that automates the configuration of distributed applications in the cloud without requiring any centralized server nor a scripting effort. The high degree of parallelism involved in this protocol making its design difficult and error-prone, we decided to specify the protocol using LOTOS NT and to verify it with CADP. So doing, we detected a major bug, which was corrected in the reference JAVA implementation. The LOTOS NT specification also served as a workbench to experiment with several possible communication models, which helped us to avoid an erroneous design.

These results have been published in [44].

6.3.4. Realizability of Choreographies
Participants: Matthias Güdemann, Gwen Salaün.

The specification and the analysis of interactions among distributed components play an important role in service-oriented computing. In order to facilitate the integration of independently developed components (named peers) that may reside in different organizations, it is necessary to provide a global contract that the peers participating in a service composition should adhere to. Such a contract is called choreography. One important problem in a top-down development process is figuring out whether a choreography specification can be implemented by a set of peers that communicate via message passing. Given a choreography specification, it would be desirable to generate peers automatically by projecting the global choreography specification to each peer ignoring all messages that are not sent or received by that peer. However, generation of peers that precisely implement a choreography specification is not always possible, i.e., there are choreographies that are not implementable by a set of distributed peers. This problem is known as realizability.

In 2011, we considered the following aspects of the realizability problem:

- In collaboration with Gregor Gössler (INRIA Grenoble) we studied the realizability of choreographies for peers interacting asynchronously through message buffers. Although this problem is generally undecidable for unbounded buffers, we proposed techniques to check whether peers interacting asynchronously with finite buffers can realize a choreography, and if so, for which buffer sizes. These results have been published in [36].
- In collaboration with Pascal Poizat (LRI, Orsay), we proposed an approach to check the realizability of choreographies using the interaction model of BPMN (Business Process Modeling Notation) 2.0. While being a standard for the abstract specification of business workflows and collaboration between services, BPMN has only been recently extended into BPMN 2.0 to support choreographies. Our approach is based on a model transformation into LOTOS NT and the use of equivalence checking. We implemented a prototype of our approach using the ECLIPSE BPMN 2.0 editor and CADP. These results have been published in [43].
- In collaboration with Meriem Ouederni (LINA, Nantes), we studied the automatic synthesis of monitors to enforce realizability, using CADP to check equivalence between the choreography and an automatically obtained distributed implementation.

6.3.5. Other Case Studies
Participants: Hubert Garavel, Frédéric Lang, Radu Mateescu, Gwen Salaün, Wendelin Serwe, Damien Thivolle.

- In the context of the TOPCASED project (see §7.2), we studied how CADP can be used to verify avionic protocols. In 2011, we prepared two lectures summarizing our prior results on four avionic protocols, namely a ground/plane communication protocol based on TFTP (Trivial File Transfer Protocol) [14], the BITE (Built In Test Equipment)/CMS (Central Maintenance Function), the ATC (Air Traffic Control) system, and the AFN (Air Traffic System Facilities Notification).
- Our prior work (2009–2010) with Fabienne Boyer and Olivier Gruber (Université Joseph Fourier Grenoble) on modeling and verification using LOTOS NT and CADP of the SYNERGY reconfiguration protocol led to a publication in an international conference [31].
Other teams also used the CADP toolbox for various case studies. To cite only recent work not already described in previous VASY activity reports, we can mention:

- behavior analysis of malware by rewriting-based abstraction [51];
- safety verification of fault-tolerant distributed components [46];
- verification of mobile ad hoc networks [57];
- model checking ERLANG applications [59];
- model-checking dataflow in service compositions [63];
- verification of a key chain based TTP transparent CEM protocol [65];
- semantics tuning of UML/SYSML [69];
- atomicity maintenance in EPCREPORT of ALE [70];
- cost analysis of semantic composability validation [71];
- rigorous development of prompting dialogs [75];
- scalably verifiable cache coherence [76].
6. New Results

6.1. Modeling

6.1.1. Simulation of electrical circuits as nonsmooth dynamical systems

Participants: Vincent Acary, Olivier Bonnefon, Bernard Brogliato.

DC-DC converters are usually difficult to simulate with classical tools like SPICE because of the highly nonlinear behaviour of some components and the frequent occurrence of intrinsically generated switching events.

The simulation of such circuits modelled as nonsmooth systems has been successfully achieved with a clear advantage over several SPICE simulators and a simulator belonging to the hybrid modelling approach [1] [48].

6.1.2. Spiking neuronal networks dynamics

Participant: Arnaud Tonnelier.

Precise spatiotemporal sequences of spikes are observed in many neural systems and are thought to be involved in the neural processing of sensory stimuli. In [58] we examine the capability of spiking neural networks to propagate stably spatiotemporal sequences of spikes. We derive some analytical results for the wave speed and show that the stability of simple waves is determined by the Schur criteria. The transmission of a sequence of several spikes is related to the existence of stable composite waves, i.e. the existence of stable spatiotemporal periodic traveling waves. We show that the stability of composite waves is related to the roots of a system of multivariate polynomials.

A fundamental aspect that shapes the properties of traveling waves in networks is the underlying lattice-structure of the space. Discreteness has a strong effect on propagating activity patterns and, for instance, anisotropy or propagation failure can be observed. Numerical simulations and analytical calculations have been carried out to characterize more precisely these properties [47].

6.1.3. Computational Toxicology

Participant: Arnaud Tonnelier.

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 assessment using a physiologically based toxicokinetic model [28].

6.1.4. 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 [50] are able to capture the bending and twisting deformations of thin elastic rods while strictly and robustly avoiding stretching deformations. In this work we are exploring new reduced-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. We have already investigated the clothoidal case \( (d = 1) \) in the 2d case [51], relying on Romberg numerical integration, and a general approach in 3d based on power series expansion was formulated in the master thesis of R. Casati, for a single element. R. Casati is currently extending the method to a chain of linked elements as well as to an arbitrary degree \( d \) of the curvature function.
6.1.5. **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 [29] we started to extend to 3d our 2d method introduced in [8] for automatically converting a smooth sketched curve into a dynamic curve at stable equilibrium under gravity. The main challenge in 3d amounts to converting an input curve into a continuous piecewise helix. Using a least-squares optimization approach is a natural option, however it may suffer from both robustness and computational issues due to the presence of multiple local minima in the objective function. To overcome these issues, we have recently proposed to reformulate the problem as a geometric interpolation problem. 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, \textit{i.e.,} a configuration that is compatible with the interpolation by a helix. Our method proves to be much more robust and faster compared to the global optimization approach. We plan to publish these results in 2012.

6.1.6. **Multiple impacts modelling**  
**Participants:** Bernard Brogliato, Hongjian Zhang, 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. Results are in [42].

6.1.7. **Simulating contact with Coulomb friction in fiber assemblies**  
**Participants:** Florence Bertails-Descoubes, Gilles Daviet.

We have developed a new frictional contact solver in [21] which is able to robustly and efficiently handle large fiber problems composed of thousands self-contacting rods with exact Coulomb friction. The solver relies on a Gauss-Seidel iterative approach, where each local one-contact solver is based on a hybrid strategy. The solution to the one-contact problem is first searched for using a nonsmooth Newton method based upon a generalized Fischer-Burmeister formulation. This primary solver manages to solve the local problem in 99.9% of the cases. When the solver fails to converge to an acceptable solution, the method switches to a more costly but exact solver, based on the \(\alpha\)-formulation introduced in [39]. This hybrid strategy experimentally allows us to always find a solution to the local problem, which greatly contributes to improve the robustness of the global solver. We have compared our new solver against other solvers of the literature (\textit{e.g.,} damped Newton solvers relying on the Alart-Curnier function) and observed a noticeable gain, both in terms of robustness and computational efficiency.

6.2. **Optimization**

6.2.1. **Nonsmooth analysis and optimization on matrix manifolds**  
**Participant:** Jérôme Malick.

Optimization on matrix manifolds is an emerging fields of research in optimization, driven by applications in robotics. We have contributed on two different projects.

- **Numerical efficiency of optimization methods.** Newton method on manifolds would require to compute a geodesic (that is to solve a ODE). It is clear that replacing classical differential-geometric objects with certain approximations, resulting in faster and possibly more robust algorithms. With our colleague P.-A. Absil from the Department of Mathematical Engineering of the École Polytechnique de Louvain (Belgique), we propose in [16] a way to construct “retractions” (a key step when applying optimization algorithms on matrix manifolds) by projecting onto the submanifold. We show
that the operation remains a retraction if the projection is generalized to a projection-like procedure that consists of coming back to the submanifold along “admissible” directions. This theory offers a framework in which previously-proposed retractions can be analyzed, as well as a toolbox for constructing new ones. Illustrations are given for projection-like procedures on some specific manifolds for which we have an explicit, easy-to-compute expression.

- **Towards the application of matrix optimization techniques to spectral manifolds.** Spectral sets are sets of matrices that depend only on the constraints on the eigenvalues: \( S = \lambda^{-1}(C) \) with \( C \) a subset of \( \mathbb{R}^n \). A spectral set \( S \) inherits from properties of the underlying set \( C \), such as convexity. We prove in [46] that the spectral sets associated to smooth manifolds in \( \mathbb{R}^n \) (having some local symmetry) are themselves manifolds in the space of matrices. This result looks simple but generalizes several useful particular cases, and was extremely difficult to prove: we brace together tools from nonsmooth analysis, differential geometry, group theory and spectral analysis.

### 6.2.2. 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 [26], [54]. An interesting feature is that the final accuracy level is controlled by a real parameter acting like a cursor. This gives ways 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 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.

- **Heaviest k-subgraph problems.** Our algorithm [26] takes advantage of the new bounds 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). In practice, our method works particularly fine on the most difficult instances (with a large number of vertices, small density and small k).

- **Max-cut.** We are working on extending our algorithm to max-cut problems [53]. It dynamically manages polyedral and semidefinite relaxations to outperform the state-of-the-art solver ([56]) on the large test-problems.

### 6.2.3. Marginal prices in electricity production

**Participants:** Claude Lemaréchal, Jérôme Malick, Welington Oliveira, Sofia Zaourar.

Two subjects were involved this year in our ongoing collaboration with EdF.

- **Stabilizing prices.** 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 [52]. 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 [59]. Our illustrations on the daily electricity production optimization of EDF show a striking stabilization at a negligible cost.

- **Accelerating** the solution phase by the so-called disaggregation technique [49], using the fact that (see Activity Report of 2010) the dual objective function is the sum of two terms: one coming from primal cost, one coming from valorization of constraints (plus possibly a third term when price stabilization is present). The resulting CPU time is drastically improved, sometimes divided by 10.
6.3. Control

6.3.1. Digital sliding mode control

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

The problem of digital sliding mode controllers is a long-standing issue not yet satisfactorily solved. We propose ideas which are inspired from the numerical methods of contact mechanics [2] and which permit a) to suppress the numerical chattering, b) to obtain a smooth stabilization on the sliding surfaces. The work is continued together with Yury Orlov in more general cases where the system is acted upon by disturbances and a disturbance estimation is added [19].

6.3.2. Discrete-time discontinuous systems

**Participants:** Vincent Acary, Bernard Brogliato, Carmina Georgescu, Scott Greenhalg, Thorsten Schindler.

We focus on some classes of discontinuous dynamical systems like relay systems, linear complementarity systems. The objectives are to show that the time-stepping numerical schemes like Moreau’s algorithm can be used to successfully simulate such systems (like in the case of biological systems like gene networks), and also to study the properties of these schemes for finite nonzero time steps (like preservation of dissipativity properties). Results are in [23], [35]. Further work deals with timestepping schemes for nonsmooth dynamical systems. So far, these schemes are locally of order one both in smooth and nonsmooth segments. This is inefficient for applications with few events like circuit breakers. To consistently improve the behavior during smooth episodes, the traditional schemes are being embedded in time discontinuous Galerkin methods. After establishing the correct mathematical setting, a Petrov-Galerkin distributional differential inclusion is outlined. The bouncing ball example illustrates its capabilities.

6.4. Locomotion analysis

6.4.1. Synchronous imitation of human motion by a humanoid robot

**Participants:** Mehdî Benallegue, Pierre-Brice Wieber.

Interactions between humans and robots require that each one is able to understand and interpret each other’s actions. From the point of view of the robot, this means: (i) to move in a way that can be naturally interpreted by humans and (ii) to be able to understand the humans’ actions. Studies in Neuroscience in the case of interactions between humans indicate that these two abilities might be tightly linked in the human’s brain: we understand actions when we map the observed action onto our motor representation of the same action [57]. In this work, we consider that the “motor representation” of a task is the control law, and “mapping an observed action” means finding the corresponding control parameter, in an observer-based approach.

Considering a correspondence between two different control laws can be seen a modeling error. This modeling error can be seen as an unknown arbitrary perturbation on the modeled system, or an unknown input sent to the observed system. We developed an observer that can cancel the effects of unknown inputs on the dynamics of discrete time linear systems with unknown inputs. To do so, the observer has to satisfy a delayed invertibility condition and use delayed outputs. In other words, the observer has to wait for several measurement after a given instant to collect enough data to reconstruct the state at that instant.

6.4.2. Hierarchic QP solver

**Participant:** Pierre-Brice Wieber.
We are working in collaboration with the LAAS-CNRS and the CEA-LIST on solving multi-objective Quadratic Programs with Lexicographic ordering: Hierarchic QPs [25]. The focus this year has been on the regularization of the problem when the Quadratic Program approaches singularities. There is indeed a problem of discontinuity of the solution when reaching such singularities, what’s not a rare event in robotic applications. This discontinuity has been related to the fact that Lexicographically ordered QPs correspond to the limit of weighed multi-objective QPs when weights go singular, and that the regularization is itself a weighting of objectives which goes to a limit when approaching singularities, and those two limit processes interfere. The solution we found so far is to cancel the first limiting process and move back from strict Hierarchic QPs to weighted QPs staying at a small distance from singularity [33]. But this solution is not really satisfying and we have to find a better one.

6.4.3. Numerical modeling of muscle contraction under FES


We have been working in collaboration with the EPI DEMAR in Montpellier on modeling muscle contraction under Functional Electric Stimulation (FES). With respect to the literature in the domain, our contributions are mostly linked to the model of the contractile element, through the introduction of the recruitment at the fibre scale, formalizing the link between FES parameters, recruitment and Calcium signal paths. The resulting controlled model is able to reproduce both short term (twitch) and long term (tetanus) responses. It also matches some of the main properties of the dynamic behaviour of muscles, such as the Hill force-velocity relationship or the instantaneous stiffness of the Mirsky-Parmley model. The specific contribution of the BIPOP team has been on the numerical implementation of the contraction model as a Linear Complementarity Problem (LCP) allowing fast and precise numerical simulations [22].

6.4.4. Modeling of human balance in public transports


In our ongoing collaboration with the IFSTTAR (previously INRETS) on modeling human balance in public transports, we have aggregated biomechanical studies and numerical models proposed in robotics, and compared how they match or mismatch in situations of strong perturbations requiring a step to recover balance. We began developing a specific Model Predictive Control scheme for the prediction of recovery step locations with adaptive step timings, reproducing various balance recovery strategies as observed in humans. Initial results for stepping predictions have been validated against a balance recovery scenario found in the literature [45].

6.4.5. Model Predictive Control for Biped Walking


We improved our Linear MPC-based walking motion generator by incorporating explicitly the robot’s kinematic constraints: polyhedral constraints on the position of the CoM ensure the kinematic feasibility of the generated walking motions for arbitrary vertical motions of the CoM. This more precise kinematic model within the LMPC allowed considering toe rotations in a safer way, considerably improving energy efficiency, naturalness of the motion, and maximal speed.

We proposed a formulation of dynamic constraints for 3D motion through simple bounds on the variables, leading to faster resolution of the corresponding optimization problem. This allowed generating three-dimensional walking on non-planar ground in real-time. Thanks to specifically enforcing leg compliance, this scheme managed additionally to reproduce the natural profiles of the CoM and the contact forces observed in human walking.

We finally refined our numerical scheme for solving Linear MPC problems in walking motion generation. We switched the underlying QP solver to enable reductions of the number of iterations through warmstart, non-empty initial active sets, and obtaining feasible solutions at each iterations, what considerably reduced the computation time, allowing 1 ms feedback loops [30], [32].
6.5. Software development

6.5.1. MECHE toolbox

**Participants:** Florence Bertails-Descoubes, Gilles Daviet.

The main tool developed in 2011 in the MECHE software was the hybrid iterative solver for Coulomb friction, published in [21]. In 2011, the MECHE software was extensively used to validate this new solver on large data consisting of thousands interacting fibers (subject to tens of thousands frictional contacts). Code parallelization and optimization were performed so as to speed up computations.

6.5.2. Platform development: Siconos

**Participants:** Vincent Acary, Olivier Bonnefon, Maurice Brémond, Franck Pérignon.

The main achievements for the SICONOS platform are:

1. Automatic serialization of the whole set of classes in SICONOS
2. Improvements and development of a full auto-generated Python wrapper in the SICONOS/Front-End
3. Development of the Siconos/Multi-body library and validation on industrial examples (C60 circuit breaker of Schneider Electric)
4. New algorithms for the resolution of the discrete frictional contact problem
5. Development of $(\theta/\gamma)$-schemes for first order dynamical systems
6. Development of routines for sliding mode control

6.5.3. AMELIF framework

**Participants:** Pierre-Brice Wieber, François Keith, Jory Lafaye.

The main improvements to the AMELIF framework developed this year are:

- A new package specific to torque control has been developed, that contains the algorithms required to realize a given motion with a humanoid robot: estimation of contact forces and torque computation (feedforward), feedback methods ensuring the contact force convergence. These algorithms have been tested for two humanoid platforms: the robot Romeo and the robot HRP-2.
- The dynamics algorithm has been improved, based on the expertise coming from the HuMAnS toolbox. Besides, inverse dynamic algorithms and Runge Kutta integration methods have been added.
- Finally, the bridge with the stack-of-tasks framework [55], that computes the inverse kinematics and the inverse dynamics of humanoid systems, has been enhanced to handle the binding with Python. With this framework, it is possible to use the Model Predictive Control algorithm aforementioned and to simulate the behaviour of a humanoid in a dynamic simulation realized by AMELIF. Tests are still in progress.
5. New Results

5.1. Mixture models

5.1.1. Taking into account the curse of dimensionality

**Participant:** Stéphane Girard.

**Joint work with:** Bouveyron, C. (Université Paris 1), Celeux, G. (Select, INRIA).

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].

In collaboration with Gilles Celeux and Charles Bouveyron, we have designed an automatic selection of the discrete parameters of the model [12]. Also, the description of the R package is submitted for publication [44].

5.1.2. A new family of multivariate heavy-tailed distributions with variable marginal amounts of tailweight: Application to robust clustering

**Participants:** Florence Forbes, Darren Wraith.

We proposed a family of multivariate heavy-tailed distributions that allow variable marginal amounts of tailweight. The originality comes from the eigenvalue decomposition of the covariance matrix in the traditional Gaussian scale mixture representation. By contrast to most existing approaches, the derived distributions can account for a variety of shapes and have a simple tractable form with a closed-form probability density function whatever the dimension. We examined a number of properties of these distributions and illustrate them in the particular case of Pearson type VII and t tails. For these latter cases, we provided maximum likelihood estimation of the parameters and illustrated their modelling flexibility on clustering examples for several simulated and real data sets.

5.2. Markov models

5.2.1. 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 fMRI within-subject analysis, two steps are generally performed separately: detection and estimation. Because these two steps are inherently linked, we proposed in this work a joint detection-estimation procedure. We adopt the so-called region-based Joint Detection Estimation (JDE) framework that deals with spatial dependencies between voxels belonging to the same functionally homogeneous parcel in the mask of the 3D brain. After building a spatially adaptive General Linear Model, prior information is introduced and a hierarchical Bayesian model is established. In contrast to previous works that use Markov Chain Monte Carlo (MCMC) techniques to approximate 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. It follows a new algorithm that exhibits interesting properties 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. Corresponding papers [27], [38], [26].

5.2.2. Adaptive experimental condition selection in event-related fMRI

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

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

Standard Bayesian analysis of event-related functional Magnetic Resonance Imaging (fMRI) data 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 areas. Criteria are not always available to select the relevant conditions or stimulus types (e.g. visual, auditory, etc.) prior to estimation and the unnecessary inclusion of the corresponding events may degrade the results. To face this issue, we propose within a Joint Detection Estimation (JDE) framework, a procedure that automatically selects the conditions according to the brain activity they elicit. It follows an improved activation detection that we illustrate on real data.

5.2.3. Finding Audio-Visual Events in Informal Social Gatherings

Participant: Florence Forbes.

Joint work with: Xavier Alameida-Pineda and Radu Horaud from the INRIA Perception team.

In this work [21] we addressed the problem of detecting and localizing objects that can be both seen and heard, e.g., people. This may be solved within the framework of data clustering. We proposed a new multimodal clustering algorithm based on a Gaussian mixture model, where one of the modalities (visual data) is used to supervise the clustering process. This was made possible by mapping both modalities into the same metric space. To this end, we fully exploited the geometric and physical properties of an audio-visual sensor based on binocular vision and binaural hearing. We proposed an EM algorithm that is theoretically well justified, intuitive, and extremely efficient from a computational point of view. This efficiency makes the method implementable on advanced platforms such as humanoid robots. We described in detail tests and experiments performed with publicly available data sets that yield very interesting results.

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

Participants: Lamiae Azizi, Florence Forbes, Senan James Doyle.

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

We recast the disease mapping issue of automatically classifying geographical units into risk classes as a clustering task using a discrete hidden Markov model and Poisson class-dependent distributions. The designed hidden Markov prior is non standard and consists of a variation of the Potts model where the interaction parameter can depend on the risk classes. The model parameters are estimated using an EM algorithm and the mean field approximation. This provides a way to face the intractability of the standard EM in this spatial context, with a computationally efficient alternative to more intensive simulation based Monte Carlo Markov Chain (MCMC) procedures. We then focus on the issue of dealing with very low risk values and small numbers of observed cases and population sizes. We address the problem of finding good initial parameter values in
this context and develop a new initialization strategy appropriate for spatial Poisson mixtures in the case of not so well separated classes as encountered in animal disease risk analysis. Using both simulated and real data, we compare this strategy to other standard strategies and show that it performs well in a lot of situations. Corresponding papers and communications [43], [24], [37], [25].

5.2.5. Probabilistic model definition for physiological state monitoring

Participants: Laure Amate, Florence Forbes.

Joint work with: Catherine Garbay, Julie Fontecave-Jallon and Benoit Vettier from LIG.

Assessing the global situation of a person from physiological data is a well-known difficult problem. In previous work, we proposed a system that does not produce a diagnosis but instead follows a set of hypotheses and decides of an alarming situation with this information. In this work [22], we focus on data processing part of the system taking into account the complexity and the ambiguity of the data. We propose a statistical approach with a global model based on Hidden Markov Model and we present data models that rely on classical physiological parameters and expert’s knowledge. We then learn a model that depends on the person and its environment, and we define and compute confidence values to assess the plausibility of hypotheses.

5.2.6. Solder Paste Inspection


This is joint work with VI-Technology.

The majority of defects in PCB manufacture are attributed to the stencil printing process. Stencil printing is the process where solder paste bricks are deposited on the PCB pads. Solder paste deposition is required to be accurate and repeatable, however complex physical process make this problematic. Components are placed, and their leads are pushed into the solder paste. The solder paste is then melted using, for example, reflow soldering.

Inspection can be performed before the solder paste is melted, and it is more economical to identify defects at this stage.

The evaluation of solder paste joint quality involves the analysis of a number of indicative measurements. From these measurements, potential faults are identified and inspected manually. The general challenge is to reduce of the number of potential faults by better analyzing the indicative factor measurements. That is, to improve the first pass yield (FPY) which is the percentage of total solder deposits that are good, and that do not require manual inspection. However, the ability to catch defects must be retained. Another aspect to consider is the temporal nature of the process; The mechanism for identifying faults needs to be retrained after a period of time, and so a solution must be capable of using a small training dataset.

It is important to understand and identify the factors that influence quality. The industry standard factor for measuring quality is solder volume. The precise volume is not directly observable, and so is estimated. Often, height is used as a proxy measure for solder bricks of equal area and shape. There are many other contributing factors, however not all of these can be measured directly, making accurate quality determination difficult.

Stencil printing process control attempts to adjust machine parameters according to informative factors. Online printing process control faces a similar challenge of using a limited number of measurements to inform on the quality of solder paste deposition.

We used statistical techniques to analyze such measurements. The exact nature of the work is confidential.

5.2.7. PCB defect detection

Participants: Florence Forbes, Kai Qin, Huu Giao Nguyen.

This is joint work with VI-Technology.

The objective is to detect defective components in PC Boards from image data. The exact nature of the work is confidential.
5.2.8. Statistical characterization of tree structures based on Markov Tree Models and multitype branching processes, with applications to tree growth modeling.

**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 modelled 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 a Poissonian Generalized Linear Model for the contingency tables of the counts of joint children state configurations. Then, parametric families of distributions were implemented and compared statistically to provide probabilistic models compatible with the estimated independence graph.

This work was carried out in the context of Pierre Fernique’s Master 2 internship (Montpellier 2 University and AgroParisTech). 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 5.2.9). This work will be continued during Pierre Fernique’s PhD thesis, with extensions to other fruit tree species and other strategies to build probabilistic graphical models and parametric discrete multivariate distributions including covariates and mixed effects.

5.2.9. 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), Evelyne Costes and Baptiste Guitton (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.

To model alternation of flowering at AS scale, a second-order Markov tree model was built. The ASs were of two types: flowering or vegetative. Generalized Linear Mixed Models (GLMMs) were used to model the effect of year, replications and genotypes (with their interactions with year or memories of the Markov model) on the transition probabilities. This work was the continuation of the Master 2 internship of Jean Peyhardi (Bordeaux 2 University) and was carried out in the context of the PhD thesis of Baptiste Guitton.

This PhD thesis also comprised the study of alternation in flowering at individual scale, with annual time step. To relate alternation of flowering at AS and individual scales, indices were proposed to characterize alternation at individual 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 on the base effect random slopes specific to genotype.
and replications), which translated into mixed effect modelling. Different indices of alternation were then computed on the residuals. Clusters of individuals with contrasted patterns of bearing habits were identified. Our models highlighted significant correlations between indices of alternation at AS and individual scales. The roles of local alternation and asynchronism in regularity of flowering were assessed using an entropy-based criterion, which characterized asynchronism.

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 5.2.8).

5.3. Semi and non-parametric methods

5.3.1. Harmony Search with Differential Mutation Based Pitch Adjustment

Participants: Kai Qin, Florence Forbes.

Harmony search (HS), as an emerging metaheuristic technique mimicking the improvisation behavior of musicians, has demonstrated strong efficacy of solving various numerical and real-world optimization problems. This work [36] presents a harmony search with differential mutation based pitch adjustment (HSDM) algorithm, which improves the original pitch adjustment operator of HS using the self-referential differential mutation scheme that features differential evolution - another celebrated metaheuristic algorithm. In HSDM, the differential mutation based pitch adjustment can dynamically adapt the properties of the landscapes being explored at different searching stages. Meanwhile, the pitch adjustment operator’s execution probability is allowed to vary randomly between 0 and 1, which can maintain both wild and fine exploitation throughout the searching course. HSDM has been evaluated and compared to the original HS and two recent HS variants using 16 numerical test problems of various searching landscape complexities at 10 and 30 dimensions. HSDM consistently demonstrates superiority on most of test problems.

5.3.2. Dynamic Regional Harmony Search Algorithm with Opposition and Local Learning

Participants: Kai Qin, Florence Forbes.

To deal with the deficiencies associated with the original Harmony Search (HS) such as premature convergence and stagnation, a dynamic regional harmony search (DRHS) algorithm incorporating opposition and local learning is proposed [35]. DRHS utilizes the opposition-based initialization, and performs independent HS with respect to multiple groups that are randomly recreated on a fixed period basis. Besides the traditional harmony improvisation operators, an opposition based harmony creation scheme is introduced to update the group memory. Any prematurely converged group will be restarted with the doubled size to further augment its exploration capability. Local search is periodically applied to exploit promising regions around top-ranked candidate solutions. The performance of DRHS has been evaluated and compared to HS using 12 numerical test problems at 10D and 30D, which are taken from the CEC2005 benchmark. DRHS consistently demonstrate superiority to HR over all the test problems at both 10D and 30D.

5.3.3. Evolutionary algorithms with CUDA

Participants: Kai Qin, Federico Raimondo.

Evolutionary algorithms (EAs), inspired by natural evolution processes, have demonstrated strong efficacy for solving various real-world optimization problems, although their practical use may be constrained by their computation efficiency. In fact, EAs are inherently parallelizable due to the operations at the individual element level and population-wise evolution. However, most of the existing EAs are designed and implemented in the sequential manner mainly because hardware platforms supporting parallel computing tasks and software platforms facilitating parallel programming tasks are not prevalently available.
In recent years, the graphics processing unit (GPU) has emerged as a powerful general-purpose computation device that can favorably support massively data parallel computing tasks carried out on its hundreds of cores. The compute unified device architecture (CUDA) technology invented by NVIDIA provides an intuitive way to express parallelism and to implement parallel programs using some popular programming languages, such as C, C++ and FORTRAN. Accordingly, we can simply write a program for one data element, which gets automatically distributed across hundreds of cores for thousands of threads to execute. Although the CUDA programming model is easy-to-use, the computation efficiency of CUDA parallel programs crucially depends on careful consideration of hardware characteristics of GPUs during algorithmic design and implementation, especially about memory utilization and thread management (to maximize the occupancy of streaming multi-processors). Without proper considerations, the parallel programs may even run slower than their sequential counterparts.

The objectives of our project are to: 1. Redesign state-of-the-art EAs using CUDA under thorough consideration of GPU’s hardware characteristics. 2. Develop a generic hardware-self-configurable EA framework, which allows automatically configuring available hardware computing resources to maximize the computation efficiency of the EA.

Currently, we had developed a memory-efficient parallel differential evolution algorithm, which features maximally utilizing the available shared memory in GPU while maximally reducing the use of the global memory in GPU considering its very limited access bandwidth. Compared with two recent parallel differential evolution algorithms implemented with CUDA in 2010 and 2011, our algorithm demonstrated significantly faster computation speed. We had also investigated the parallel implementation of test problems and provided a guideline on how to implement any user-defined test problem and combine it with an existing parallel EA framework. To the best of our knowledge, this is the first research work on this topic.

5.3.4. Modelling extremal events

Participants: Stéphane Girard, Laurent Gardes, Jonathan El-methni, El-Hadji Deme.

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

We introduced a new model of tail distributions depending on two parameters \( \tau \in [0, 1] \) and \( \theta > 0 \) [16]. 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 an 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 \( \tau \). This permits the construction of new estimators of extreme quantiles. The results are submitted for publication [48]. 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 \( \rho \) (see paragraph 3.3.1). We proposed a new family of estimators encompassing the existing ones (see for instance [62], [61]). 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 [46].

5.3.5. Conditional extremal events

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

Joint work with: J. Carreau, A. Lekina, Amblard, C. (TimB in TIMC laboratory, Univ. Grenoble I) and Daouia, A. (Univ. Toulouse I)

The goal of the PhD thesis of Alexandre Lekina is to contribute to the development of theoretical and algorithmic models to tackle conditional extreme value analysis, ie the situation where some covariate information \( X \) is recorded simultaneously with a quantity of interest \( Y \). In such a case, 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 [58] with extreme-value methods in order to obtain efficient
estimators of the conditional tail index and conditional extreme quantiles. When the covariate is random (random design) and the tail of the distribution is heavy, we focus on kernel methods [14]. We extension to all kind of tails in investigated in [45].

Conditional extremes are studied in climatology where one is interested in how climate change over years might affect extreme temperatures or rainfalls. In this case, the covariate is univariate (time). Bivariate examples include the study of extreme rainfalls as a function of the geographical location. The application part of the study is joint work with the LTHE (Laboratoire d’étude des Transferts en Hydrologie et Environnement) located in Grenoble.

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 [64]. The PhD theses of Jonathan El-methni and Gildas Mazo should address this issue too.

5.3.6. Level sets estimation

Participants: Stéphane Girard, Laurent Gardes.

Joint work with: Guillou, A. (Univ. Strasbourg), Stupfler, G. (Univ. Strasbourg), P. Jacob (Univ. Montpellier II) and Daouia, A. (Univ. Toulouse I).

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 [41]: 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. The results are submitted for publication [51].

5.3.7. Quantifying uncertainties on extreme rainfall estimations

Participants: Laurent Gardes, Stéphane Girard.

Joint work with: Carreau, J. (Hydrosciences Montpellier) 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 obtained in [14]. It is then of primary importance to study the sensitivity of the extreme rainfall estimation to the estimation method considered.

The obtained results are published in [13].

5.3.8. 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 (e.g., 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). We have also defined an adaptive version of the method which is able to deal with block-wise evolving data streams [28].

### 5.3.9. 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 overestimate 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 to model the total consumption of the circuit from the probability distribution of consumption of each individual gate. Our preliminary results are published in [18].
6. New Results

6.1. Algorithms for molecular modeling and simulation

6.1.1. Interactive quantum chemistry

Participants: Maël Bosson, Caroline Richard, Antoine Plet, Sergei Grudinin, Stéphane Redon.

We have proposed what appears to be the first algorithm for interactive quantum chemistry simulation at the Atom Superposition and Electron Delocalization Molecular Orbital (ASED-MO) level of theory. When drawing and editing molecular systems, interactive quantum chemistry provide immediate, intuitive feedback on chemical structures. Our method is based on the divide-and-conquer (D&C) approach, which we show is accurate and efficient for this non-self-consistent semi-empirical theory. The errors induced by the D&C approach have been studied empirically and via a theoretical study of two toy models. With this method, we have demonstrated interactive quantum chemistry simulations for systems up to a few hundred atoms on a current multicore desktop computer. As the number of cores on personal computers increases, and larger and larger systems can be dealt with, we believe such interactive simulations – even at lower levels of theory – should thus prove most useful to effectively understand, design and prototype molecules, devices and materials. This result has been published in Journal of Computational chemistry [4]. Figure 7 illustrates an interactive modeling session with a benzene molecule.

![Figure 7. Interactive modeling session. After breaking a benzene cycle, the user moves a hydrogen atom closer to the top carbon atom to force them to bond (a). Then, the user pulls on a carbon atom to form a fulvene molecule (b-d). Interactive electronic structure calculations allow the user to easily build plausible topologies, and get immediate feedback on the chemical structure.](image)

6.1.2. Adaptively Restrained Particle Simulations

Participants: Svetlana Artemova, Stéphane Redon.

Particle simulations are widely used in physics, chemistry, biology [17], [21], and even computer graphics [13]. However, many important problems still constitute significant computational challenges, including molecular docking, protein folding, diffusion across bio-membranes, fracture in metals, ion implantation, etc. Numerous methods have been developed to accelerate particle simulations, by e.g. increasing the simulation’s time step [18], [9], [14], [26], [27], [24], improving the computational complexity of the simulation [32], [15], [8], [10], [31], or simplifying the system under study [19], [29], [28], [16], [11], [31], in particular via coarse-graining methods [20], [33], [30] or multiscale and multiresolution methods [25], [22], [23], [12].
We have introduced a novel, general approach to speed up particle simulations that we call Adaptively Restrained Particle Simulations (ARPS).

Our approach works by adaptively switching positional degrees of freedom on and off repeatedly during a simulation, while letting momenta evolve. The benefits of this approach are that (a) it is mathematically grounded and is able to produce stable, energy-preserving 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, unbiased statistics can be obtained.

We have illustrated ARPS on several numerical experiments, including a shock propagation example and a polymer-in-solvent study. The shock propagation example demonstrates how ARPS make it possible to smoothly trade between precision and speed (Fig. 8).

![Figure 8. Simulating a shock propagation 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.](image)

The polymer-in-solvent study shows how one may collect unbiased statistics with ARPS, and demonstrates that it can be done faster than with usual (reference) simulations. The results are shown in Fig. 9.

These results have been submitted for publication.

### 6.1.3. Adaptive interactive quantum chemistry
**Participants:** Maël Bosson, Sergei Grudinin, Stéphane Redon.

We are now working on applying the adaptive paradigm to the quantum chemistry methods, to allow for interactive editing of systems of any sizes and shapes. We are developing new methods and criteria to adaptively focus the computational resources on the most relevant parts of the system. Figure 10 illustrates our recent results in this direction. In this framework, we can already achieve interactive rates and efficient virtual prototyping for systems of size up to thousand atoms on a current desktop computer.

### 6.1.4. Interactive molecular modeling with haptic feedback
**Participants:** Aude Bolopion, Barthélémy Cagneau, Stéphane Regnier, Stéphane Redon.
Figure 9. Computing the hydrodynamic radius $R_H$ of a solvated polymer. Traditional simulations reduce the variance more at each time step (top), but adaptively restrained (AR) simulations perform many more time steps, so that they reduce the variance faster in wall-clock time (bottom). In this example, for any target precision, AR simulations compute the hydrodynamic radius four times faster than reference simulations.

Figure 10. Interactive electronic structure calculations in SAMSON. In this example the system is divided in four subsystems (for which the bounded boxes are displayed). The electronic structure is adaptively updated and the geometry is being optimized while the user edits the molecular system. Because the user pulls on one atom in the left part of the system, the electronic structure is accurately recomputed for the most left subsystem (atoms in red). In the neighboring subsystem, the electronic structure is updated with a cheaper model (carbon atoms are in black and hydrogen in white). In the right part of the system, the user force do not have a sufficiently large impact and atoms as well as the electronic structure are frozen (frozen atoms are displayed in blue).
In collaboration with ISIR in Paris and LISV in Versailles, we have developed a new approach for haptic interaction with molecular systems.

Molecular interactions typically have a high dynamic range (HDR), combining short-range stiff repulsive effects with long-range, soft attractive and repulsive terms. As a result, faithful haptic rendering of such molecular interactions is both important and difficult, in particular in applications where the precise perception of molecular forces is necessary (e.g. in molecular docking simulations). Traditionally, teleoperation coupling using constant gain control schemes have limited applications since they are unable to transmit to users low attractive forces without truncating repulsive ones. Furthermore, constant scaling displacement induces either instability or time-consuming experiments (displacements are slow), which deteriorates the ease of manipulation. We have described a variable gain haptic coupling method specifically designed to render high dynamic range (molecular) forces. The proposed method has been evaluated by user tests on an experiment involving two water molecules. We have observed that variable force amplification is widely appreciated, whereas variable displacement scaling is appropriate only for users that are already familiar with haptic manipulation. A complex experiment on a HIV molecule has been carried out using this variable gain system. This approach has been published in the proceedings of the 2011 World Haptics Conference [7]. Figure 11 shows SAMSON being used with a haptic interface at ISIR.

Figure 11. Haptic interaction with the HIV protease in SAMSON. Virtual environment setup at ISIR.

6.2. Algorithms for molecular docking

6.2.1. Prediction of Interface Water Molecules Using a Knowledge Base

Participants: Georgiy Derevyanko, Sergei Grudinin.

We developed a method to predict positions for interface water molecules as part of the predicted protein-protein complex. For this purpose we used a previously developed knowledge-base scoring methodology. First, we constructed a training set of non-homologous protein complexes with interfacial water. Then, we deduced the water-protein interaction energy using this training set. And finally, we positioned water molecules around a test protein complex on a regular grid and optimized their positions according to the knowledge-based water-protein interaction energy. This method was validated in a recent CAPRI competition. Figure 12 illustrates our method on a test protein.
Figure 12. Densities of interface water molecules around a test protein computed using our water-protein knowledge-based potential.
6.2.2. Development of a Knowledge-Based Scoring Function  
**Participants:** Georgiy Derevyanko, Sergei Grudinin.

We developed a new method to obtain a knowledge-based potential function for protein-protein interactions. To derive such a potential, we formulated a convex quadratic programming problem with about 1,000,000 of linear constraints and developed a fast iterative solver to solve it. We validated this scoring function in the CAPRI competition Round 24, where our prediction of the Target 50 was ranked 4th. Figure 13 shows the use of Legendre polynomials to fit statistics obtained on the knowledge base.

![Figure 13](image1.png)  
*Figure 13. Using Legendre polynomials to fit protein-protein interaction statistics.*

### 6.2.3. Development of a Local Knowledge-Based Potential for Structure Optimization and Prediction of Point Mutations in a Protein  
**Participants:** Petr Popov, Sergei Grudinin.

We developed and validated a method that reconstructs the shape of the binding potential function between two proteins by locations of its global minima. After, we used the obtained potential function for optimization of positions of two docking partners. We demonstrated that using our method we can significantly improve the quality of predictions of such widely-used docking algorithms as HEX and ZDOCK. We validated this method in the CAPRI competition Round 26, where our re-scoring prediction of the Target 53 was ranked 3rd.

We have also developed a method to predict the influence of point mutations on the binding affinity constant of a protein complex. First, we made point mutations and reconstructed the sidechain of the mutated residue. Then, we repacked the sidechains that are within a certain cutoff distance from the mutated residue. After, we optimized the structure of two proteins using a smooth pair-additive knowledge-based potential function. We iteratively repeated the two previous steps until convergence of the binding energy. Finally, we converted the obtained binding energy into the binding affinity constant of two proteins. We validated this method in the CAPRI competition Round 26 with the Targets 55 and 56. Figure 14 shows an interactive docking session using a knowledge-based potential for CAPRI Round 26 Target 53.

![Figure 14](image2.png)  
*Figure 14 shows an interactive docking session using a knowledge-based potential for CAPRI Round 26 Target 53.*

### 6.3. Software engineering  
#### 6.3.1. SAMSON's architecture  
**Participant:** Stéphane Redon.
The data model of SAMSON has been expanded. The goal is to represent a nanosystem as the union of several interacting models: structural models (geometrical and topological information, to define relationships between structural elements), dynamical models (to define kinematical and dynamical relationships between structural components), interaction models (to define physical interactions between dynamical components, e.g. forces between atoms or rigid bodies), and visual models (visual representations, for user interaction).

All models are part of the data graph, which contains all the information related to the system being modeled. The referencing system has been significantly expanded, with data structures to safely handle objects creation and deletion. An event mechanism has been designed and added to SAMSON, so that nodes of the data graph may send messages to each other. These messages can be related to topological changes, structural changes, dynamical changes, etc.

6.3.2. SAMSON’s software engineering process

Participants: Jocelyn Gaté, Stéphane Redon.

SAMSON’s software development process has been much improved. CMAKE is used to ensure that all parts of SAMSON may easily be built on several platforms (Windows, Mac and Linux). Thanks to CMAKE, a variety of Integrated Development Environments may be used (Visual Studio, Eclipse, XCode, etc.).

CTEST and CDASH are used to test SAMSON, and the Pipol platform has been used to perform nightly builds.

6.3.3. Graphical User Interface design

Participants: Jocelyn Gaté, Stéphane Redon.

Several functionalities have been added to the graphical interface of SAMSON, including customizable toolbars (that plug-in developers will be able to modify), as well as a data graph view (Figure 15).
Also, because plug-ins might have complex interfaces and settings, a mechanism to save and load custom presets has been developed (Fig. 16).

**Figure 16.** SAMSON allows users to save and load plug-in presets.

### 6.4. Applications

Methods and tools developed in our group have been used in the following studies:

**6.4.1. Building Blocks of Bacterial Chemoreceptor Arrays**

**Participant:** Sergei Grudinin.
Bacterial chemoreceptors are known to cluster at the cell poles where they form partially hexagonally ordered arrays. This clusterisation is important for the function of chemotaxis system. In this study, we performed an analysis of the known structural and biochemical information on the components of chemoreceptor arrays: chemoreceptors themselves, histidine kinases and adapter proteins. Based on this analysis, we proposed a set of basic interactions within the chemotaxis system (the array building blocks) and constructed their atomistic models. The models resulting from these blocks are in agreement with experimental information and provide a basis for understanding the atomic-level structural organization of chemoreceptor arrays.

### 6.4.2. A Novel Dimerization Interface of Cyclic Nucleotide Binding Domain

**Participant:** Sergei Grudinin.

Cyclic nucleotide binding domain (CNBD) is a ubiquitous domain of effector proteins involved in signalling cascades of prokaryota and eukaryota. In this study, we described a novel CNBD dimerization interface found in crystal structures of bacterial CNG channel MlotiK1 and mammalian second messenger cAMP-activated guanine nucleotide-exchange factor Epac2. Using computational tools we demonstrated that the found interface is stable, in contrast to the dimerization interface reported previously. Comparisons with cN-bound structures of CNBD showed that the dimerization is incompatible with second messenger cAMP binding. Thus, the cAMP-dependent monomerization of CNBD may be an alternative mechanism of the cAMP sensing. Based on these findings, we proposed a model of the bacterial CNG channel modulation by cAMP.
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.

This work is the topic of the PhD thesis of N. Cardoso de Castro. 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 [54], [66]. Indeed, as we had pointed out in the review paper [67], 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 some intermediate radio-modes, which consume more energy than ‘Sleep’ but allow to reach the transmitting mode consuming less energy in the transition. 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. This work is described in [66] and in a journal paper (in preparation). A further research direction is the exploration of receding-horizon techniques (Model Predictive Control), to solve a slightly modified formulation of the same problem. This research is in collaboration with Dr. Daniel Quevedo, senior lecturer at the University of Newcastle, Australia, in particular during a three-months visit of N. Cardoso de Castro at University of Newcastle.

6.1.2. Adaptive Delta Modulation in Networked Controlled Systems with bounded disturbances

Participants: C. Canudas de Wit [Contact person], F. Gomez-Estern [University of Sevilla], F. R. Rubio [University of Sevilla].

In the context of communication and control co-design for networked systems, this work investigates the closed-loop properties of the differential coding scheme known as Delta Modulation when used in feedback loops within the context of linear systems controlled through a communication network [19]. We propose a new adaptive scheme with variable quantization step, by defining an adaptation law exclusively in terms of information available at both the transmitter and receiver. With this approach, global asymptotic stability of the networked control system is achieved for a class of controllable (possibly unstable) linear plants. Moreover, thanks to the globally defined switching policy, this architecture enjoys a disturbance rejection property that allows the system to recover from any finite–time unbounded disturbance or communication loss.

6.1.3. Control, communication, computation (3C) co-design: Multi-level classification and formulation

Participants: C. Canudas de Wit [Contact person], A. Farhadi [University of Melbourne].

We introduce here an integration framework for Control/Communication/Computation (3C) co-design based on the motivating example of fleet control of Autonomous Underwater Vehicles (AUVs) [35], [75]. Specifically, we address the problem of almost sure stability of an unstable system with multiple observations over packet erasure channel, with emphasis on coding computational complexity. We look at the tradeoff between duty cycle for feedback channel use, coding computational complexity, and performance. We compare coding computational complexity and performance for two cases: a) No feedback channel at all, and b) Feedback channel all the time. It is shown that the strategy of using feedback channel results in a better performance.
6.2. Collaborative distributed consensus algorithms for control and estimation

6.2.1. Distributed Control

**Participants:** A. Seuret [Contact person], C. Canudas de Wit, L. Briñón Arranz, G. Rodrigues de Campos, K. H. Johansson [KTH].

The first contribution in this area deals with the source-seeking problem in which the task is to locate the source of some signal using a fleet of autonomous underwater vehicles. The objective is here to use the underwater vehicles equipped with appropriate sensors as a mobile sensors network. In [28] and [29], we present a method which allows estimating the gradient of the signal propagation using a distributed consensus filters [27]. To do so, we consider a group of vehicles uniformly distributed in a fixed circular formation. We then show that this distributed consensus algorithm converges to good approximation of the gradient of the signal propagation. The algorithm takes into account the communication constraints and depends on direct signal measurements. Our approach is based on the previous results in formation control to stabilize the fleet to elastic formations which can be time-varying [29] and in a collaborative source-seeking algorithm proposed earlier by members of the team. The results are supported through computer simulations.

The second contribution on collaborative control concerns the design and analysis of a distributed algorithm whose goal is symmetric robot deployment. This activity results from the collaboration between INRIA and KTH provided by the visit of G. Rodrigues de Campos (PhD student) at KTH during six month. The objective is here to propose a hierarchical control strategy composed of two stages. The first one corresponds to an algorithm for swarm dispersion and a second concerns the design of a additional algorithm which minimizes the inter-agent angles. In this context, the behavior of each vehicle depends only on the relative positions of agents it can sense. The article submitted to ICRA’12 [84], presents some simulation examples for different configuration support the derived theoretical results.

6.2.2. Distributed Estimation

- Collaborative protocols for estimation and control
  **Participants:** A. Kibangou [Contact person], A. L. F. Almeida [Universidade Federal do Ceara].

  In wireless communication systems, spatial diversity plays a key role in combating signal fading arising from multipath propagation. As long as the transmitter is equipped with multiple antennas, it is well known that spatial diversity can be exploited further at the transmitter by means of space-time coding [88]. In contrast to conventional (single-user) space-time coding/decoding, when dealing with cooperative wireless networks, spatial diversity must resort to distributed space-time coding/decoding, where a collection of distributed antennas belonging to multiple terminals work in a coordinated way to encode/decode the transmitted information [85].

  For this purpose, we have extended the Khatri-Rao Space time coding method proposed in [86] to cooperative networks (see Fig. 5). For cooperating nodes having a single antenna, these nodes constitute a virtual antenna array at both transmitting and receiving front-end. At each node, the received data can be viewed as slices of a third-order tensor. Therefore, retrieving the informative data is achieved by means of a CP tensor decomposition using an Alternating Least Squares (ALS) algorithm for example. When all the slices cannot be gathered at the same node, for storage resources limitations for example, a distributed ALS method can be used as in [77], which is an average consensus based method. In a consensus problem, a group of network nodes try to reach agreement on a given quantity of interest that depends on their local values [79]. Instead of using a standard consensus method where convergence is achieved asymptotically, we have proposed a finite time average consensus approach that relies on the knowledge of the graph topology. The proposed algorithm and its performance evaluation by means of simulations are described in [37].

- Kalman filtering based distributed fault detection and isolation
  **Participants:** A. Kibangou [Contact person], F. Garin, S. Hachour, A. Esna Ashari.

  This year, we have started a research activity related to distributed fault detection and isolation. Our first work has been the master thesis of S. Hachour on the monitoring of a solar farm.
Indeed, stimulated by increasing energy demand and ecological concerns, clean energy production with renewable resources is a key research topic that presents a largely unexplored potential of development.

For this purpose, solar farms constitute power plants of the future. In such systems, electricity is produced thanks to the combined action of a large number of interconnected modules (solar panels). Each module individually produces energy, but only their interconnection allows reaching the global task of a relevant energy production. Due to the interconnection topology, a local fault on a given module can induce damageable effects on the whole network. In order to detect and localize a fault, a sensor network can be deployed over the farm. Thanks to the recent advances in wireless communications, the sensors can be equipped with wireless devices, creating a network of communicating sensors. A classical way to exploit such a network would be to create a hierarchical (tree-like) structure which conveys all measurements to a centralized computer which would analyze all data. However, a failure in a communication link or in the centralized computer would result in breakdown of the whole fault detection system, which is an unacceptable risk in an application domain of strategic importance such as a power plant. Therefore we have proposed a decentralized approach that relies on local data aggregation using the computing and communicating resources of the sensor nodes. As a consequence, nodes cooperation produces a global decision, available at each point of the network, and computable even in the case where a few sensors or links are unavailable.

The monitoring procedure is achieved in a distributed way using a Kalman filtering approach. Now, by considering the sensor network as the system of interest, we try to derive distributed estimation methods that are robust to malicious entities and monitoring methods to detect anomalies in the system due to these malicious entities or malfunctioning of the network. These issues are addressed by A. Esna Ashari during his post-doctoral stay in our team.

![Cooperative communication system](image)

**Figure 5. Cooperative communication system.**

### 6.2.3. Distributed Consensus

- Finite-time distributed average consensus on sensor networks
  
  **Participant:** A. Kibangou [Contact person].

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 algorithm as it is the case for the Distributed Kalman filter [80] and the Distributed Alternating Least Squares algorithm [77]. Obviously, asymptotic convergence is not suitable for these kinds of distributed methods. Even though, speed convergence of consensus algorithm have been explored in [78] and [91] with the goal to derive fast consensus algorithms, running standard consensus in finite-time constitute a source of errors not easily quantifiable. Sometimes, bounds can be derived. 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 [38], [48]. Moreover, the number of iterations is equal to the number of distinct nonzero eigenvalues of the graph Laplacian matrix. Then, by periodically restarting the consensus algorithm, we have also shown that, in the noisy case, exact average consensus can be achieved asymptotically.

- Quadratic indices for performance evaluation of consensus algorithms

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

Traditional analysis of linear average-consensus algorithms studies, for a given communication graph, the convergence rate, given by the essential spectral radius of the transition matrix (i.e., the second largest eigenvalues’ modulus). For many graph families, such analysis predicts a performance which degrades when the number of agents grows, basically because spreading information across a larger graph requires a longer time. However, when considering other well-known quadratic performance indices (involving all the eigenvalues of the transition matrix), the scaling law with respect to the number of agents can be different. This is consistent with the fact that, in many applications, for example in estimation problems, it is natural to expect that a larger number of cooperating agents has a positive, not a negative effect on performance. It is natural to use a different performance measure when the algorithm is used for different purposes, e.g., within a distributed estimation or control algorithm. Examples of various relevant costs can be found in the book chapter [50] 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 [18] 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 symmetries. To do so, we exploit the analogy between reversible Markov chains and resistive electrical networks. This latter work is part of the Ph.D. thesis of E. Lovisari at University of Padova, Italy.

- Distributed averaging over digital noisy networks

**Participants:** F. Garin [Contact person], R. Carli [Università di Padova], G. Como [MIT], P. Frasca [Politecnico di Torino].
We study iterative distributed averaging algorithms for networks whose nodes can communicate through memoryless erasure broadcast channels. In order to compare the performance of different algorithms, we define suitable complexity measures, which account for the number of channel transmissions (communication complexity), and, respectively, of in-node computations (computational complexity) required to achieve a desired precision. These performance measures are particularly relevant, as they allow for directly estimating the energy consumption of such distributed computation systems, as well as their time-complexity.

The algorithms we propose combine the classical iterative linear consensus algorithm (where at each iteration, each agent receives the states of its neighbors and takes a suitable convex combination of them), with source-channel coding schemes for the reliable transmission of real numbers on noisy channels. Our algorithms involve a sequence of transmission phases, of increasing duration, in which the agents attempt to broadcast their state, i.e. their current estimate of the global average, to their neighbors, alternated to averaging steps, in which the agents’ states are updated. These algorithms are fully distributed, and they do not require the agents to have any global knowledge of the network structure or size. Our main result shows that such algorithms drive the agents to state agreement (consensus) which can be made arbitrarily close to the true average. The number of channel transmissions and in-node computations is shown to grow at most poly-logarithmically in the desired precision. We also show how communication feedback, when available, allows one to modify the algorithms, achieving asymptotic average consensus (i.e., state agreement on the average of the initial observations), and reducing the computational and communication complexities. This work is presented in the paper [15]. In the paper [22], we present and analyze a modified algorithm, which can be used when source coding (compression) and channel coding (error correction) are performed by two separate encoders. Such algorithm takes into the account the fact that, even without any channel feedback, the part of the error which is due to lossy compression and not to channel noise is perfectly known by the transmitter, and a compensation can be introduced, thus improving performance.

- **Distributed Consensus algorithms**
  **Participant:** A. Seuret [Contact person].

Concerning this problem, we address the classical issues of the stability analysis of consensus algorithm in continuous-time. The objective of the present work is to show that the performances classical consensus algorithms can be improved using an appropriate memory of the controlled state. We want to design a novel type of consensus algorithm which uses not only the current state of the algorithm but also a sampled version of it. The key problems are here the design of the best parameters, i.e., the sampling period and the ratio between the contributions of the current and the sampled states. It has to be noticed that a usual intuition is to say that using past values of the state a reduction of performances or to instability. However, our contributions show that this is not always in single and double integrator consensus algorithms [41]. These contributions is based on an LMI framework and based on algebraic communication matrix structure. The efficiency of the method is tested for different network communication schemes.

### 6.2.4. Distributed real-time Simulation of numerical models

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

To allow real-time simulation of high fidelity engine models, different techniques have to be applied in order to fulfill the real-time constraints. Real-time simulation involves trade-offs between several aspects, such as real-time constraints, models computational complexity and integration accuracy. Traditionally HIL designers consider that every step of the simulation must be real-time and deterministic, leading to strongly synchronized systems, at the cost of ineffective computation burdens. It has been shown that adequately splitting the plant’s model into weakly synchronized sub-systems allows for efficiently using variable steps numerical integrators, simulation speed-ups and subsequent effective parallel versions of HIL systems [25].
6.3. Stability and control design of asynchronous interconnected systems

6.3.1. New approaches for stability conditions design

- Stability for asynchronous sampled data systems
  **Participants:** A. Seuret [Contact person], C. Briat [KTH], J. Gomès Da Silva Jr. [UFRGS], W. Jiang, M. M. Peet [Illinois Institute of Technology].

  During the last year an important effort has been devoted to controlled systems under communication constraints. In particular a novel approach to assess stability of continuous linear systems with sampled-data inputs has been provided for the first time in [21]. The main contribution of this article is to make the bridge between the discrete-time and the continuous-time approaches to ensure stability of the closed loop system. The interest of the method remains in the application of the discrete-time Lyapunov theorem using the continuous-time model without introducing exponential. This method suggests the introduction of particular types of functionals of several shapes: using an adaptation of classical time-delay functionals [21]; using a discretization method [26]; or using SOS [44].

  Then extensions to uncertain systems, time-varying parameter systems [21]; or non linear systems (for instance with saturations [36], [43]) become straightforward in comparison to the discrete-time approaches. The stability conditions are expressed in terms of linear matrix inequalities. Sufficient conditions for asymptotic and exponential stability are provided dealing with synchronous and asynchronous samplings and uncertain systems. An additional stability analysis is provided for the cases of multiple sampling periods and packet losses in [21]. Moreover this method has also been extended to the case of sampled-data systems with additional input delay [42], [46] and to the case of impulsive systems (several papers are submitted on this topic, for example [65]).

- Stability of control under weakened real-time constraints
  **Participants:** D. Simon [Contact person], A. Seuret, P. Andrianiaina [AIRBUS].

  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 [46]. 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 [24], [63].

6.3.2. Control for asynchronous sampled data systems

- Control architecture and tools
  **Participants:** D. Simon [Contact person], R. Pissard [SED], S. Arias [SED].

  During the development of control systems, hardware-in-the-loop that is showed in Fig. 6 takes place between design level simulations and costly experiments with the real plant. Using the prototype of ORCCAD V4 several HIL real-time simulators have been set up. These simulators combine multi-threaded/multi-rate real-time controllers running control algorithms synchronized with a variable step numerical integrator running a model of the plant [49]. These simulators have been further used to implement and test several kind of feedback/flexible scheduling schemes related to the FeedNetBack project, in particular real-time controllers subject to (m,k)-firm scheduling, Kalman filters modified to account for data loss and varying sampling controllers [62]. Finally a collaboration with SARDER about the integration of discrete (logical) control loops on top of continuous control tasks has been carried on. In this architecture ORCCAD is used to design the low level continuous controllers while the reactive parts are designed and synthesized using the BZR language [23].
Event-based control algorithms

Participants: A. Seuret [Contact person], N. Marchand [Contact person], S. Durand.

Asynchronicity is becoming more and more meaningful in modern control architectures and some new control strategies are being developed by some research teams in the world. The principle of these control laws is to compute a new control signal only when some event occur, where a event characterizes a change in the system and therefore a need for a new control. These approaches are supposed to reduce the number of times the control is computed (and consequently the CPU utilization) and to remove the real-time hard constraint on the computational system. Some works around this domain have been proposed by some members of the NeCS team.

In [45], one may look at the problem of reducing the amount of information to be sent to the actuators through the Network. Indeed the controller may be able to trigger the information to be sent. The main idea is to let the controller decide if the system needs an update of its control input. This class of control algorithms is called event-triggered. An algorithm is suggested to sample the control input based on the behavior of a Lyapunov-like function in [45]. This algorithm is event-triggered since the Lyapunov-like function directly depends on the state of the systems.

In [69] and [73], we firstly proposed to remove the safety limit condition introduced by K-E. Årzén in his event-based PID controller [93]. In this paper, the control signal is updated only when required from a performance point of view, that is when the measurement crosses a given level. Årzén also suggested to enforce an event when the sampling interval achieves a given maximal amount of time. This safety limit was added to prevent the system to be sampled less than what Shannon theorem requires but, in fact, we showed that the Shannon sampling condition is no more consistent in the context of event-based systems. Moreover, a practical implementation of a cruise control mechanism on a small radio-controlled vehicle was recently suggested [34]. Moreover, we are interested on updating the control signal only when required from a stability point of view. Such a solution consists for instance in enforcing an event when a Lyapunov function crosses a given level. Based on the seminal work from M. Velasco in [90], we suggested a simple Lyapunov sampling in [33].

Control with adaptive sampling

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

Control and real-time computing have been associated for a long time, for the control of industrial plants and in embedded or mobile systems, e.g. automotive and robotics. However both parts, control and computing, are often designed with poor interaction and mutual understanding. We propose here an integrated control and scheduling co-design approach, where closing the loop between the control
performance and the computing activity is promising for both adaptivity and robustness issues. We
developed during the last years a variable sampling control methodology based on the LPV (Linear
Parameter Varying) framework and $H_\infty$ control synthesis, where the sampling interval is used as a
known and controlled variable [8]. Few assumptions about sampling are needed for this control
design: the main point is that the control interval is known and lies between the predefined bounds
$[h_{\text{min}}; h_{\text{max}}]$, whatever the origin of the control interval variations, its speed and its frequency.
Another approach is proposed now to design sampling varying gain-scheduled controller for LPV
systems, based on the Linear Fractional Representation (LFR). This method has already been
studied concerning the synthesis of discrete-time gain-scheduled controller, depending only on the
sampling period. The method has been extended to deal with the design of a gain scheduled LFT
controller w.r.t the sampling interval and w.r.t system’s parameters, given a discrete-time Linear
Fractional Representation (LFR) of the LPV varying sampling model [13]. The approach comes
from the robust control theory and consists in separating the LTI part $P$ (not depending on the set
of parameters) from the varying part $\Delta$ (parameters or uncertainties), as shown on Fig. 7. The
matrix $\Delta$ represents the influence of the set of parameters $\rho(\cdot)$ on the plant. From this model, a gain
scheduled controller can be computed, depending on the same set of parameters $\Delta$, or a subset of $\Delta$.
The LFT approach proposed in [83] has been extended to set a LFR model that accounts for system
and sampling parameters. In that case the uncertainty matrix is as depicted in Fig. 7 on the right,
where $\delta$ is the deviation of the sampling interval w.r.t. the nominal sampling rate and $\Delta$ account
for the plant’s uncertainty. A controller is then synthesized using the $H_\infty$ framework, where weighting
functions allow to shape the control system’s response [82].

In the framework of the FeedNetBack IST project, the LFR formulation previously presented is
applied to an Autonomous Underwater Vehicle (AUV) for the control of its altitude $z$ [40]. The
global control structure is presented on Fig. 8. To control the altitude, two controllers are considered
($\hat{K}_z(\delta)$ and $\hat{K}_\theta(\delta, \rho)$) based on two models ($\hat{G}_z(\delta)$ and $\hat{G}_\theta(\delta, \rho)$). The LFR formulation is used here
to keep some varying parameters into the model formulation. Indeed, in previous works, the limits
of a simple linearization around a fixed equilibrium point have appeared. When the pitch angle is too
far from 0 (the value chosen for $\theta$ at equilibrium), the linearized model becomes too different from
the linearized one, which leads to bad performances. The $\Delta$ block contains the varying part of the
model, which depends on the linearization point ($\theta_{eq}$). The two considered varying parameters are
$\rho_1 = \cos(\theta_{eq})$ and $\rho_2 = \sin(\theta_{eq})$. The model is then discretized and the sampling period is added
to the $\Delta$ parameter block. Compared with the previous approach using a linearized plant model, the
new one shows (in simulation) improved altitude tracking and a better utilization of the available
range of the actuators [82].

---

![Figure 7. System under LFT form.](image-url)
6.4. Vehicular transportation systems

6.4.1. Traffic modelling, estimation and control

Participants: C. Canudas de Wit [Contact person], C. Irinel-Constantin [CRAN], D. Pisarski, L. Leon Ojeda.

This part is related to the developed work within the Network of Excellence HYCON2 (Highly-Complex and Networked Control Systems). It is interested in problems of modelling, estimation and control in traffic.

In [81], the problem of equilibrium sets for the Cell Transmission Model is studied. The objective is to design the homogeneous distribution of density on the freeway, where the input flows are the decision parameters to be determined. For the proper design of the balanced density, the extensive analysis on the structures of equilibria is crucial. The analysis is carried out for the two different cases, where all sections of the freeway are assumed to be free or congested, respectively. The necessary conditions for the existence of balanced equilibria are formulated. These conditions show that the key for the design of the balanced states may be the variable speed limiting, which strictly cooperates with the ramp metering. The computational algorithm for the input flows in case of free balance is proposed. In order to illustrate the results, the numerical example is provided.

In [39], the authors are interested in the highway traffic model-based density estimation. A strategy is proposed for real-time density estimation for traffic networks. To this aim, we introduced a deterministic constrained macroscopic model which reduces the number of possible affine dynamics of the system and preserves the number of vehicles in the network. This model is used to recover the state of the traffic network and precisely localize the eventual congestion front. The state of the network is recovered using what we call forward/backward observers. We pointed out that during unobservable modes the estimation error is preserved due to vehicle conservation law. Numerical simulations show the efficiency of the proposed strategy.

In [30], the problem of front congestions control is treated. For this, we have introduced a new traffic lumped model with only two cells (one free, and another congested) the cells have variable length, and a variation law for the front congestion completes the 3-dimensional model. In opposition to fixed-length cell models that are commonly represented by a set of linear state-dependent switching systems, our model results in a lower dimensional nonlinear system which solutions are continuous. Based on this model, we have designed a “best-effort” control strategy using variable speed limits. The notion of best effort control is here linked to the physical variable speed limit constraints which limits its size and as well as its rate variation. This results in a relative simple control in closed-form that can be implemented by using only information about the front congestion location.

Other work is under development and is related to the traffic show case application and the achievements reached that correspond to the operation of the freeway network around the Grenoble area (Grenoble South Ring). We started by designing the general network architecture, after specifying sensors and actuators locations along the highway and finally setting the platform of an interface between Matlab and our micro simulator "Aimsun". We have also carrying out some simulations from a real life application on Grenoble South ring of a deterministic state estimation technique. Using constrained macroscopic model which in fact reduces the number of dynamic states and preserves the conservation law (number of vehicles in the network).
6.4.2. Vehicle control

- Tele-operated control
  Participants: C. Canudas de Wit [Contact person], W. Jiang, J. Dumon, O. Sename.
  A mathematical driver model in the spatial equation form has been introduced for analysis of drivers’ behavior [31]. In the model, a previewed distance is taken into account. First, optimal control is applied. For the ideal case without driver’s reaction delay concerned, for both long distance preview and shorter one, the vehicle tracks well the path. Whereas, when time-delay added into the system, too short preview distance cause the instability of the system. The simulation result corresponds with the real driving experience. Then, Lyapunov-Krasovskii functional approach is applied deal with stability problem with the driver’s delay. In this case, when the delay becomes greater, the longer preview distance will be needed. The average derivation of the steering input is calculated for different preview distance as well as the time-delay cases, which well verifies our driver model. The main contribution of this model is that the preview effect only depends on the path information and it does not affected by the vehicle speed, so the result is more neutral

- Electric Power Steering Systems
  Participants: C. Canudas de Wit [Contact person], V. Ciarla, J. Tordesillas Illán [UJF].
  This part presents several aspects of modeling, observation and control towards a new generation of Electrical Power Steering (EPS) systems [47]. In particular we design an optimal control to reject oscillations of the steering column, then we device a new observer to estimate the internal state variables of the steering column, the driver applied torque (steering wheel torque), and the load torque (tire/ground contact friction). Finally, we also revisited the LuGre tire dynamic friction model by improving the transient behavior between the sticking phases and the dynamic ones. Simulation of the proposed control and observer are shown at the end of the paper using the improved LuGre-tire friction model. Index Terms—Electric Power Steering systems (EPSs), LQ control, LuGre friction model, observer.

6.5. Energy-aware control of systems on chip

Participants: N. Marchand [Contact person], S. Durand.

Achieving a good compromise between computing power and energy consumption is one of the challenges in embedded architectures of the future. This management is especially difficult for 45 nm or 32 nm known to be at the limit of the scalability, i.e. with a high process variability. This is a key point in the ARAVIS project. Automatic control loops have therefore to be designed to minimize the energy consumption still making the performance fit the requirement in a context of highly technological uncertainties of the chip. This issue is notably discussed in [92]. Finally, the main objective is to dynamically control the computing activity and the energy consumption using the voltage and the frequency according to the requirements of the operating system. In this way, a robust control law was developed in [72] and [32] in order to minimize the high voltage running time with a predictive technique, i.e. to minimize the energy consumption while ensuring good computational performance. This control was initially done for one node (i.e. a processor) but in ARAVIS SoC, the chip is composed of several clusters with several nodes each (see Fig. 9). Thus, the energy controller has to manage the voltage level (one voltage domain by cluster) and the frequency for all nodes: a maximal frequency is performed for critical node and then a ratio of this frequency could be apply for the other nodes. Thus, a multicore control strategy with low computational needs was also proposed [71]. This work yields two patents: the first one for the monocore case [70] and the second one for the multicore case [74].
Figure 9. Architecture du SoC ARAVIS.
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].

Concerning road traffic, the research activity during 2011 focused on the mathematical analysis of traffic flow models on road networks or subject to unilateral constraints. In particular, [34] is devoted to a hyperbolic 2nd order model for traffic flow with local flux constraint. We describe two admissible Riemann solvers and we construct ad hoc finite volume numerical schemes to compute these solutions. The paper [59] 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. Finally, a general traffic flow model with phase-transition is proposed and described in [28].

M.L. Delle Monache just started her doctoral thesis in the same topic. More precisely she will study hyperbolic models of traffic flow and associated optimization problems.

6.1.2. Crowd motion

Participants: Nora Aïssiouene, Régis Duvigneau, Nader El Khatib, Jihed Joobeur, Paola Goatin, Massimiliano D. Rosini [ICM, Warsaw University, Poland].

Concerning pedestrian 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 a macroscopic model for crowd movements consisting in a scalar conservation law accounting for mass conservation 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 [54]). We are now studying existence and uniqueness of solutions of the corresponding initial boundary value problem. From the numerical point of view, we are implementing the model 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 will be completed soon by Nora 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. Moreover, jointly with the PULSAR team, we have supervised J. Joobeur’s internship, which was devoted to pedestrian data collection from real-word video recordings (Turin metro station). The density data will serve to validate the model.

The above researches were partially funded by the ERC Starting Grant "TRAM3 - Traffic management by macroscopic models".

6.2. Optimum design 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.
6.2.1. Cooperation and competition in multidisciplinary optimization

Participants: Étienne Baratchart [ENSEIBB MATMÉCA], Jean-Antoine Désidéri, Régis Duvigneau, Adrien Zerbinati.

The framework of our research aims to contribute to numerical strategies for PDE-constrained multiobjective 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 testcase of aero-structural aircraft wing shape optimization, in which the eigen-split-based optimization reveals clearly superior [ 33 ].

While the two-discipline method is currently being applied to various complex physical multiobjective situations (see in particular 6.2.2, 6.2.6, 6.2.7, 6.2.8), 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 [ 53 ], see Fig. 3.

Figure 3. 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 multiobjective problem in which all the gradients are assumed to be known. The theorem [58] states that in the convex hull of the gradients, there exists a unique vector of minimal norm, $\omega$; if it is nonzero, the vector $\omega$ 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 $\omega$ 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 6.2.3.

6.2.2. Virtual games for coupling global to local shape optimization

Participant: Régis Duvigneau.

In several engineering problems, the system to optimize is characterized by some parameters that define global shape properties, while remaining parameters define local shape modifications. Of course, these two sets of parameters do not play the same role and have not the same impact on the cost functional value. Therefore, we are studying how to construct an efficient optimization strategy that takes benefit of this global / local splitting of parameters.

A typical aerodynamic shape optimization problem has been studied, that consists of a lift-constrained drag minimization for a transonic wing, whose sections are defined by two B-Spline curves whereas global shape characteristics are defined by five parameters (span, root-tip length ratio, angle of attack, twist angle, sweep angle). It has been found that the naive simultaneous optimization of all parameters failed, due to the multimodality of the problem. Alternatively, the use of a virtual game strategy, based on a splitting between the local and global parameters, yields a satisfactory result for a moderate cost [47].

6.2.3. Multiple-Gradient Descent Algorithm (MGDA)

Participants: Jean-Antoine Désidéri, Régis Duvigneau, Adrien Zerbinati.

6.2.3.1. Basic experiments and validation

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 [50], MGDA has been tested over a number of classical multiobjective-optimization testcases, and found successful to converge to Pareto-optimal solutions in situations of either convex or concave Pareto sets. Additionally, MGDA and PAES [61] were found to have complementary merits, making a hybrid method promising.

6.2.3.2. Metamodel-supported 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 metamodels of the functionals of interest (lift, drag, etc) and to calculate approximate gradients by local finite differences. These metamodels 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 is currently being tested and extended to an internal flow optimization problem related to an automobile air-conditioning system and governed by the Navier-Stokes equations. This more difficult problem has been proposed by Renault within the OMD2 ANR project.
6.2.3.3. MGDA in functional setting

One aspect of the theoretical result concerning the minimal-norm element $\omega$ is that, regardless the possibly-functional setting of the problem in case of a distributed system, the descent-direction $\omega$ is identified in the standard $n$-dimensional vector space $\mathbb{R}^n$ ($n$: the number of objective functions).

This observation has led to examine the application of MGDA in the functional setting of domain-decomposition methods (DDM) in which a functional criterion and a functional control can be defined at each interface independently permitting to formulate the DDM problem as a multi-objective optimization. On-going research in this area is related to the necessary preconditioning, or normalization procedure, of the gradients.

6.2.4. Flow control

Participants: 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 flow separation. 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 solution 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. First results show the efficiency of such an approach for laminar flow problems [31],[44].

6.2.5. Optimum shape design in aerodynamics by the adjoint method

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

At ONERA, compressible flow simulations governed by the Euler or Navier Stokes (RANS) equations are conducted with the software elsA [57] that admits both structured and unstructured-grid formulations. Local aerodynamic optimizations are made with a version that includes the calculation of the shape gradient via the solution of an adjoint equation. The discrete adjoint is calculated formally step-by-step to include the various derivative terms involved, and is being enhanced gradually to account for more complex models. In particular, for RANS computations, this gradient today includes the differentiation of the turbulence model.

6.2.5.1. Metamodels including derivative information.

In this context, to alleviate the cost of an optimum-shape design in aerodynamics, M. Bompard in his thesis [26], has examined how metamodels, firstly based on functional values only, could be used to determine shortcuts in the convergence process. Second, when the gradient w.r.t. the design parameters is known, the gradients of functionals of interest, that is, most commonly, aerodynamic coefficients, are calculated. Thus, these derivative informations can also used to construct more elaborate metamodels. Such constructions have also been studied systematically and used efficiently in global optimizations [37]; in particular co-Kriging and Support-Vector Regression, for which a technique to adjust automatically the free parameters has been proposed based on a simplification of the leave-one-out test.

6.2.5.2. Parameterization-free local optimization

When the derivatives of the functionals w.r.t. the volume geometry, $dJ/dX$, have been calculated, it is also possible to calculate the gradient w.r.t. surface coordinates, $dJ/DS$. Since the surface deformation steers the entire mesh movement, often through analytical dependencies, M. Bompard [26] has also examined how could $dJ/DS$ be used directly in a local aerodynamic optimization. However, it is well-known that the distribution of $dJ/DS$ is very irregular, and its usage in the optimization loop necessitates that adequate smoothing procedures be elaborated. Partial success was achieved in this area, still subject to research.
6.2.6. Aero-structural optimization

**Participants:** Gérald Carrier [Research Engineer, ONERA/DAAP], Jean-Antoine Désideri, 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 are applied to evaluate accurately the various physical contributions to drag. As a result, the numerical tools necessary to conduct a very realistic aircraft wing optimization are now set up and are being exploited [38]. It is also envisaged to conduct a two-objective optimization (drag and mass reduction) via a Nash game using our optimization platform FAMOSA.

6.2.7. 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’s 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 [60] for the calculation of the Whitham function corresponding to the lowest-boom (axisymmetric) shape. The method is currently being extended to account for more general geometries. Besides, aero-acoustic optimizations have been realized successfully by coupling the aerodynamic optimizer (based on Euler calculations by the elsA sofware) 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.

6.2.8. 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 has recently started at ONERA a CIFRE thesis supported by EUROCOPTER, Marignane. 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 thesis work includes the set up of a hierarchy of models from low to high fidelity, in order to calibrate appropriate functional criteria. Secondly, our Nash game approach to competitive optimization will be implemented, using our optimization platform FAMOSA, and comparisons with the results by A. Dumont will be made.

6.2.9. 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 modeling of hull shapes by parametric surfaces. The objective is to be able to modify existing hull shapes by controlling a small number of parameters, that are meaningful for naval architects. Two testcases are considered: the bow shape for trawler ships (see Fig. 4) and the whole hull shape for canoes, in collaboration with the Fédération francaise de Canoe-Kayak.

Figure 4. Initial shape (left) and deformed shape to generate a bow (right) for the trawler ship, to reach two line targets (in red).

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 application of this approach for the case of beverage cans which undergo elastoplastic deformation under high pressure 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.
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].

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 \[27\] is extended in order to include the non-homogeneous wound healing or non-homogeneous dorsal closure \[52\].

6.5. Particular applications of simulation methods

6.5.1. Analysis of a two-level parameterization optimization for antenna design

**Participants:** Benoît Chaigne [Doctoral student, 2007-2010], Jean-Antoine Désideri.

Similar to the discretization of ordinary or partial differential equations, the numerical approximation of the solution of an optimization problem is possibly subject to numerical stiffness. In the framework of parametric shape optimization, hierarchical representations of the shape can be used for preconditioning, following the idea of Multigrid (MG) methods. By analogy with the Poisson equation, which is the typical example for linear MG methods, we have addressed a parametric shape inverse problem. The ideal cycle of a two-level algorithm can be defined and adapted to shape optimization problems that require appropriate transfer operators. With the help of a symbolic calculus software we have shown that the efficiency of an optimization MG-like strategy is ensured by a small dimension-independent convergence rate. Numerical examples are worked out and corroborate the theoretical results. Applications to antenna design have been realized. Finally, some connections with the direct and inverse Broyden-Fletcher-Goldfarb-Shanno preconditioning methods have been shown \[29\].

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 adaption. 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.

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.

6.5.4. Nash game approach to image processing

**Participants:** Abderrahmane Habbal, Rajae Aboulaich [Mohamed V University of Rabat], Maher Moakher [University of Tunis], Moez Kallel [University of Tunis], Anis Theljani [University of Tunis].
We have started in 2011 to study the application of game modeling to image processing problems. We propose an original game theory approach to simultaneously restore and segment noisy images \[ 56 \]. We define two players: one is restoration, with the image intensity as strategy, and the other is segmentation with contours as strategy. Cost functions are the classical relevant ones for restoration and segmentation, respectively. The two players play a static game with complete information, and we consider as solution to the game the so-called Nash Equilibrium. For the computation of this equilibrium we present an iterative method with relaxation. The results of numerical experiments performed on some real images show the relevance and efficiency of the proposed algorithm. Based on a similar idea, we formulated well known data completion (Cauchy) problems for Laplace equation as Nash games \[ 55 \] and obtained results of existence, uniqueness and stability of a Nash equilibrium which turns out to be the Cauchy solution when the Cauchy data are compatible. With A. Theljani, we study the extension of the Nash data completion approach to nonlinear parabolic equations with application to image inpainting.

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 research axes have been developed in collaboration with GALAAD Project-Team:

- Methods for adaptive parameterization including a posteriori error estimate for elliptic problems \[ 36 \], \[ 35 \], \[ 42 \];
- Numerical schemes based on Spline functions for 2D inviscid compressible flow simulations;
- Optimization methods for structural elasticity, based on shape-gradient concept, and fluid-structure interactions \[ 48 \] (in collaboration with Technical University of Munich).

6.7. **Resilient workflows for distributed multidiscipline optimization**

**Participants:** Toan Nguyen, Laurentiu Trifan.
A distributed platform based on the YAWL workflow management system has been designed and implemented to deploy HPC applications on the Grid5000 network infrastructure. The goal is to provide a generic environment for the design of complex applications that require HPC resources for large-scale fault-tolerant applications, see Fig. 2 and [39].

The platform provides application-level fault-tolerance, i.e., resilience, in order to restart the workflow execution whenever abnormal behavior or system-level errors occur. This allows a variety of errors to be taken into account, ranging from execution time-outs to out-of-bounds parameter values to be managed, with the help of user intervention when necessary [40].

The error management procedure uses exception handlers in YAWL to trigger the appropriate corrective actions, which are defined by rules invoking the adequate compensating workflows. Once defined, this can be made transparent to the users [41].

An original scheme based on asymmetric checkpoints has been designed in order to reduce overhead in both checkpointing and application restarts. It minimizes the number of required checkpoints created based on default rules and user-specific needs.

The platform is currently developed in Java on Linux workstations and should be portable on Windows and MacOS, although this has not been tested yet.

Examples are deployed on the Grid5000 national network infrastructure using the OMD2 test-cases (e.g., vehicle air-conditioner pipe optimization). The goal is here to provide a demonstrator platform that deploys large-scale optimization applications involving several (typically over five) HPC clusters distributed on the Grid5000 network. The coarse-grain definition of the application is defined by a workflow that monitors the distributed execution of the parallel component codes on the various clusters, providing resilience capabilities in case of system and application errors, see Fig. 5.

![Figure 5. Application definition using YAWL.](image-url)
6. New Results

6.1. Sex-specific impact of meiotic recombination on nucleotide composition

Meiotic recombination is an important evolutionary force shaping the nucleotide landscape of genomes. For most vertebrates, the frequency of recombination varies slightly to considerably between the sexes (heterochiasmy). We extended the examination of the evolutionary impact of heterochiasmy beyond primates to include four additional eutherian mammals (mouse, dog, pig, and sheep), a metatherian mammal (opossum), and a bird (chicken). We compared sex-specific recombination rates with nucleotide substitution patterns evaluated on transposable elements. Our results, based on a comparative approach, revealed a great diversity of the relationship between heterochiasmy and nucleotide composition. We found that the stronger male impact on this relationship is a conserved feature of human, mouse, dog, and sheep. In contrast, variation in genomic GC content in pig and opossum is more strongly correlated with female, rather than male, recombination rate. We also showed that the sex-differential impact of recombination is mainly driven by the chromosomal localisation of recombination events, not the overall average recombination rate. We proposed a new explanation for the evolutionary impact of heterochiasmy on nucleotide composition. This work has been submitted for publication. This work was done in collaboration with D. Mouchiroud.

6.2. Modelling the influence of karyotype on the distribution of meiotic recombination

Given the important evolutionary role of recombination, recent work has focused on understanding the dynamics of this molecular process. We analysed the variation of recombination rate among species in relation to their karyotype. Specifically, we developed a non-linear model between the total genetic and physical lengths of chromosomes. Our model incorporates important biological knowledge of the recombination process. It further allows the estimation of two main parameters of recombination: the additional recombination rate per Mb and the per-species average strength of interference. Since the model is defined on data from genetic maps, at the global level of the karyotype, it can be applied even on low-resolution data and, hence, can result in the exploration of multiple species. By analysing the variability of our models recombination parameters among species, we showed that the recombination rate and the interference strength are regulated at the Mb scale of the genomes. We found that the genome size is a strong predictor of the recombination rate, while the average physical length of chromosomes is positively correlated with the interference parameter of our model. These relations represent valuable tools for the estimation of recombination parameters even for species lacking genetic maps. This work is in submission. This work was done in collaboration with D. Mouchiroud.

6.3. Finding long and multiple repeats with edit distance

We developed an algorithm, FILMRED, for detecting long similar fragments occurring at least twice in a set of biological sequences (a conference paper [25] has already appeared, a journal version is in preparation). The problem becomes computationally challenging when a non-negligible number of insertions, deletions and substitutions are allowed. The algorithm is exact and manages instances whose size and combination of parameters cannot be handled by other currently existing method. This is achieved by using a filter as a preprocessing step, and then the information that this filter has gathered in the following inference phase. FILMRED can deal with very long repeats (up to a few thousands) occurring possibly several times, with a difference rate (substitutions and indels) of 10% or more. This work was done in collaboration with N. Pisanti and P. Peterlongo. The software will be made available in a near future.
6.4. Genomics of symbiosis

Insect symbioses are model systems for studying the evolution of bacterial genomes. Importantly, evolution is directly related to the type of interaction and may also be influenced by the presence of other symbionts [11]. We are currently studying the genomes of different symbionts. The first one is the genome of the Wolbachia strain that has recently become obligatory for the reproduction of Asobara tabida. First results indicate that this genome may have been recently invaded by mobile elements. The second project concerns the sequencing of the different symbionts that co-exist in the insect Bemisia tabaci. The initial sequences are promising and we hope to close the genome of four different symbionts in this system, which will allow the study of the complementation between symbionts and lateral gene transfers among symbionts sharing the same intracellular arena. This work is done in collaboration among others with L. Mouton.

6.5. Bacterial syntenies

The automatic identification of syntenies across multiple species is a key step in comparative genomics that helps biologists shed light both on evolutionary and functional problems. We developed a versatile algorithm to extract all syntenies from multiple bacterial species based on a clear-cut and very flexible definition of the synteny blocks that allows for gene quorum, partial gene correspondence, gaps, and a partial or total conservation of the gene order [8]. We then applied this algorithm to two different kinds of studies. The first one is a search for functional gene associations. In this context, we compared our algorithm to a widely used heuristic - I-ADHORE - and showed that at least up to ten genomes, the problem remains tractable with our exact definition and algorithm. The second application was linked to evolutionary studies: we verified in a multiple alignment setting that pairs of orthologs in synteny are more conserved than pairs outside, thus extending a previous pairwise study. We then showed that this observation is in fact a function of the size of the synteny: the larger the block of synteny is, the more conserved the genes are. This work was done in collaboration with F. Boyer.

6.6. Spatial synteny in Eukaryotes

Folding and intermingling of chromosomes has the potential of bringing close to each other loci that are very distant genomically or even on different chromosomes. On the other hand, genomic rearrangements also play a major role in the reorganisation of loci proximities. Whether the same loci are involved in both mechanisms has been studied in the case of somatic rearrangements, but never from an evolutionary standpoint. From the joint study of the network of spatial proximities of human genomic loci and a dataset of evolutionary breakpoints between human and mouse, we were able to provide evidence that evolutionary breakpoints tended to cluster spatially in human cells, which led us to propose the new notion of spatial synteny, which generalises the concept of genomic synteny. This work was submitted in 2010 and was accepted in 2011 [24]. It was done in collaboration with C. Lemaitre.

6.7. Chimeric transcripts in Eukaryotes

In the framework of the extension of the ENCODE project to chromosomes 21 and 22, we had the opportunity to identify a new category of transcripts, which we call chimeric transcripts, since they contain exons from different genes, which can themselves be located far apart on the same chromosome, or on different chromosomes [9]. We further found that the network formed by these connected genes is enriched in cliques of sizes 3 and 4, which seems to indicate that transcription and/or splicing of these sets of genes co-occur in time and space, as is confirmed by the confrontation of our expression dataset to a dataset indicating the co-localisation of DNA fragments in 3D. This work was done in collaboration with, among others, R. Guigó.

6.8. KisSplice: de-novo calling alternative splicing events from RNA-seq data

We addressed the problem of identifying polymorphisms in RNA-seq data when no reference genome is available, and avoiding an assembly. Based on the fundamental idea that each polymorphism will correspond to a recognisable pattern in a De Bruijn graph constructed from the RNA-seq reads, we propose a general...
model for all polymorphisms in such graphs. We then introduce an exact algorithm to extract alternative splicing events and show that it enables to identify more correct events than current transcriptome assemblers. Additionally, when we applied our method on a 50M reads dataset from human, we were able to identify 3884 events, out of which 57% are not present in the annotations, which confirms recent estimates showing that the complexity of alternative splicing has been largely underestimated so far. This work has been submitted to publication. This work was done in collaboration with P. Peterlongo.

6.9. 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. Through transcriptomic approaches (Sanger sequencing of mRNA, in vitro substraction of transcriptomes), we have established a first reference transcriptome of this insect. The analyses done demonstrate that *Wolbachia* interferes with different host pathways, and notably regulation of oxidative stress, apoptosis and autophagy, which are known to be involved in host-pathogen interactions. RNAseq has now been performed on this system and analyses are underway, which will allow a finer investigation of the interaction using the algorithm KISSPLICE (see above) developed in the EPI for the analysis of NGS data without reference genome.

6.10. Navigating through unexplored pre-miRNA candidates

The computational search for novel miRNA precursors often involves some sort of structural analysis with the aim of identifying which type of structures are prone to being 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 candidate pre-miRNAs that can be identified in single-genome approaches, even after applying several filters for robustness and stability, a conventional structural clustering approach is unfeasible. We presented a method, MINDIST, to represent candidate structures in a feature space which summarises key sequence/structure characteristics of each candidate. We demonstrated 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. Our method was compared to another single-genome method (TRIPLET-SVM) in two datasets, showing better performance in one and comparable performance in the other. Additionally, we showed that our approach allows for a better interpretation of the results. The MinDist method is available upon request and will be made available online. This work has been submitted to publication. This work was done in collaboration with A. T. Freitas and R. Backofen.

6.11. Exploration of the genetic network of Buchnera aphidicola

Aphids are important agricultural pests which can grow and reproduce thanks to their intimate symbiosis with the γ-proteobacterium *Buchnera aphidicola* that furnishes them with essential amino acids lacking in their phloem sap diet. We investigated how *B. aphidicola*, with its reduced genome containing very few transcriptional regulators, responds to variations in the metabolic requirements of its host by concentrating attention on the leucine metabolic pathway [23]. We showed that leucine is a limiting factor for aphid growth and displays a stimulatory feeding effect. Our metabolic analyses demonstrated that symbiotic aphids are able to respond to leucine starvation or excess by modulating the neosynthesis of this amino acid. Taken together, our data showed that the response of *B. aphidicola* to the leucine demand of its host is multimodal and dynamically regulated, providing new insights concerning the genetic regulation capabilities of this bacterium in relation to its symbiotic functions.
6.12. Annotation database system to ease the development and update of BioCyc databases

In recent years, genomes from an increasing number of organisms have been sequenced, but their annotation remains a time-consuming process. The BioCyc databases offer a framework for the integrated analysis of metabolic networks. The PATHWAY TOOL SOFTWARE SUITE allows the automated construction of a database starting from an annotated genome, but it requires prior integration of all annotations into a specific summary file or into a GenBank file. To allow the easy creation and update of a BioCyc database starting from the multiple genome annotation resources available over time, we developed an ad hoc data management system that we called Cyc Annotation Database System (CycADS) [22]. The CycADS pipeline for annotation management was used to build the ACYPiCyc database for the pea aphid Acyrthosiphon pisum, TRICAyc for Tribolium castaneum and DROMEcyc for Drosophila melanogaster. This work will be extended to create a database for other arthropods. This work was done in collaboration among others with S. Collela.

6.13. Representation and curation of metabolic pathways: UniPathway

UniPATHWAY (http://www.unipathway.org) is a manually curated database for the representation and annotation of metabolic pathways developed in collaboration with the Swiss Institute of Bioinformatics (Swiss-Prot group). UniPATHWAY provides explicit chemical representations of enzyme-catalysed and spontaneous chemical reactions, as well as a hierarchical representation of metabolic pathways. This hierarchy uses linear subpathways as the basic building block for the assembly of larger and more complex pathways, including species-specific pathway variants. All of the pathway data in UniPATHWAY has been extensively cross-linked to existing pathway resources such as KEGG and METACYC, as well as sequence resources such as the UniProt KnowledgeBase (UniProtKB). UniPATHWAY has been used to provide a controlled vocabulary for pathway annotation within UniProtKB records since UniProt release 14.7 (January 2009). In release 2011_08 of UniProt, UniPATHWAY provides annotation for 118,390 distinct Swiss-Prot protein records and 783,299 TrEMBL protein records. On the UniProtKB web site, each of these records is linked to the appropriate pathway description in the UniPATHWAY web site. In 2011, the complete description of the UniPATHWAY database has been published in Nucleic Acids Research (Jan. 2012 Database Issue) and has been chosen by the editors of Nucleic Acids Research to appear on the Featured Articles page (top 5% of NAR papers: http://www.oxfordjournals.org/our_journals/nar/featured_articles.html) [16].


RHEA (http://www.ebi.ac.uk/rhea) is a project developed in collaboration with the Swiss Institute of Bioinformatics (SIB) and the European Institute for Bioinformatics (EBI). It aims at providing a comprehensive resource of expert-curated biochemical reactions. RHEA provides a non-redundant set of chemical transformations for use in a broad spectrum of applications, including metabolic network reconstruction and pathway inference. RHEA includes enzyme-catalysed reactions (covering the IUBMB Enzyme Nomenclature list), transport reactions and spontaneously occurring reactions. RHEA reactions are described using chemical species from the Chemical Entities of Biological Interest ontology (ChEBI) and are stoichiometrically balanced for mass and charge. They are extensively manually curated with links to source literature and other public resources on metabolism including enzyme and pathway databases. This cross-referencing facilitates the mapping and reconciliation of common reactions and compounds between distinct resources, which is a common first step in the reconstruction of genome scale metabolic networks and models. The complete description of the database will appear in the Jan. 2012 NAR Database issue [5].

6.15. Metabolic reconstruction of Klebsiella pneumoniae str. Kp13

Klebsiella pneumoniae str. Kp13 is a multidrug resistant pathogen involved in nosocomial outbreaks in Brazil. The objectives of this study still underway are two-fold: (1) to perform a graph-based metabolic reconstruction of the small-molecules network of strain Kp13 and (2) from the reconstructed network evaluate what makes this pathogen so successful in colonising its human host. Manual annotation of the network was performed
and a choke-point analysis was carried out, which yielded interesting targets such as L-rhamnose biosynthesis enzymes that may be related to the virulence of this bacterium. The MetAnnot platform within MetExplore was used for the manual curation of the network and graph export/import. A paper is in preparation. This work is being done in collaboration with A. T. Vasconcelos and M. Nicolás.

6.16. Clustering of elementary modes and metabolic modules identification

While it is commonly admitted that metabolism is modular, the identification of metabolic modules remains an open topic. In fact, what remains open comes even upstream of any identification problem, and refers instead to the question of defining a good model for modules in metabolic networks. One would hope that such a model might enable, for instance, to automatically derive the metabolic pathways that have been discovered and painstakingly established by biochemists over the years. Elementary modes, that are informally defined as metabolic subnetworks that can function at steady state, meaning that all internal metabolites are produced and consumed in equal rates (that is, nothing accumulates internally), represent one starting point for a definition of modules that has been considered. There are two difficulties related to this however. One is that enumerating elementary modes has itself been proven (by members of the EPI) to be a hard problem, while the second is that even small networks (around 100 reaction nodes) can have millions of elementary modes. Clustering them based on, for instance, the amount of overlap, that is of shared reactions, is one idea that has been used. We attempted another definition of modules using elementary modes that is related to a form of node-covering. This has been submitted to publication. The corresponding software is available on request. This work was done in collaboration with C. E. Ferreira, E. Moreno, and P. Crescenzi.

6.17. Enumeration of metabolic stories

In many cases, we are interested in understanding how metabolism reacts when an organism is submitted to some environmental stress, that is, to establish which metabolic processes are involved in an organism’s adaptation to such stress. In order to do it, elements of the metabolism such as metabolites are monitored to determine which are over- or under-produced during the stress response as compared to the organism’s state in normal conditions. Such quantitative and qualitative measurements are called metabolomics. The affected part may represent only a small portion of the network, that is, involve a small subset of metabolites. The aim then is to identify subnetworks that enable to link together all elements in this subset. More formally, given such subset and a metabolic network represented as a digraph where nodes are metabolites and edges link two metabolites when one is the input of a reaction that produces the other, we are interested in identifying all maximal directed acyclic graphs that cover all the metabolites in the subset of interest, and have no sources or targets that are not one of these metabolites. Such maximal DAGs are called metabolic stories. We established already that finding one metabolic story is easy (paper submitted) and used the algorithm developed, G OBBOLINO (see the Softwares Section), for practical purposes (second publication in preparation). This work was done in collaboration with P. Crescenzi, A. Marchetti-Spaccamela, L. Stougie, A. Marino, F. Jourdan, and L. Cottret.

6.18. Identification of all the minimal precursor sets for a given set of targets

Once the metabolic network of an organism has been defined, the question of how are produced the essential metabolites for the organism arises. In particular, it is important to know which are the metabolites that the organism needs to obtain from its environment to produce those essential metabolites. In the case of symbiosis, this environment could be the host, and determining such metabolites one way of exploring the dialog that is established between different organisms that entertain a close and often long term relation. We call such metabolites that must be obtained from the environment precursors. We had already in 2008 established the complexity of the problem of, given a network and a set of targets of interest, enumerating all minimal sets of precursors enabling to produce the targets, and given one first algorithm. The algorithm has since been much improved (journal paper submitted). The algorithm has been developed into a software called PITUFO (see the Softwares Section). This work was done in collaboration with A. Marchetti-Spaccamela, L. Stougie, and L. Cottret.
6.19. Metabolic network comparison

All living organisms have very similar metabolic needs as they all uptake nutrients from their environment, degrade them into basic building blocks such as amino acids and nucleotides, which in turn are essential input for protein and DNA synthesis. A natural expectation is therefore that they share an even reduced core of metabolic functions necessary to carry out this basic small molecule metabolism. 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. This absence of a metabolic core is essentially due to intracellular symbionts. These results were in preparation in 2010 and are now submitted. The work was done in a collaboration with Ludovic Cottret (INSA Toulouse) and Ana Tereza Vasconcelos.

6.20. Wolbachia detection

Wolbachia is a large monophyletic genus of intracellular bacteria, traditionally detected using PCR assays. Its considerable phylogenetic diversity and impact on arthropods and nematodes make it urgent to assess the efficiency of these screening protocols. The sensitivity and range of commonly used PCR primers and of a new set of 16S primers were evaluated on a wide range of hosts and Wolbachia strains [20]. We showed that certain primer sets are significantly more efficient than others but that no single protocol can ensure the specific detection of all known Wolbachia infections. This work was done in collaboration among others with S. Charlat.

6.21. Genetic architecture of parasite infection

The problem here is to understand the genetic architecture of a parasitic invasion by investigating the different phenotypes such invasion produces in the host. One such phenotype is called "cytoplasmic incompatibility". Briefly, when a parasite invades a male host, it induces developmental arrest, ultimately, death of the host’s offspring unless the fertilised embryo carries the same symbiont inherited from its mother, that is, unless the female is also infected. This has been tentatively explained by a toxin/antitoxin model that involves a toxin deposited by the parasites in the male’s sperm that induces the death of the zygote unless neutralised by an antidote produced by the parasites present in the egg. One toxin/antitoxin pair is linked to one gene. Given a set of observed cytoplasmic incompatibilities, the question is how many genes are required to explain it. Formally, and skipping many intermediate modelling steps, this translates into, given a 0/1 matrix M for pairs of male/female (a 0 indicating that either the male is not infected or, if it is, then so is the female meaning that there is no incompatibility, and a 1 indicating that the male is infected while the female is not), what is the minimum number of "rectangles" that enable to cover all the 1s in M? A "rectangle" in this case is a subset of columns and rows such that, if permuted, they can be arranged in the form of a rectangle with only 1s (publication Nor et al., 2010 by the BAMBOO Team and collaborators). One rectangle corresponds to a gene. The model can then be made more complex by considering that genes may have different alleles (different forms), and are expressed in variable quantities. The quantitative version of the problem in particular translates into having to find a minimum number of "triangles" that cover all 1s. All the above problems translate also into different versions of edge covers of a bipartite graph that are for the most part algorithmically original, and always not fully resolved (meaning, there remains open questions, notably regarding complexity). Work on these problems within the Associate Team SIMBIOSI and the results already obtained should lead to a publication in 2012. This work was done in collaboration among others with S. Charlat.

6.22. Population structure and dynamics

We recently started a collaboration with the Pasteur Institute in Cambodia (Dr. P. Buchy) and the CIRAD at Montpellier (Dr. R. Frutos) on viral population structure and dynamics. In this context, we developed an exploratory statistical approach to characterise mutational patterns in viral populations. The basic idea is to use Multiple Correspondence Analysis (MCoA) on a multiple alignment of nucleic sequences. To this purpose, the alignment is encoded as a boolean table where rows correspond to the sequences and columns to the presence/absence of characters. This can be done simply by considering each of the four possible bases as a
different character, or by considering only two possible states: one for the major base and one for all other (minor) bases. This technique turned out to be very effective in representing co-mutation patterns within a population of sequences. As for simple CoA, the plot is quite easy to interpret by biologists, both in terms of proximities of sequences and characters (i.e. mutations). Moreover, it has some strong relationships with parsimony phylogeny that need to be clarified. In [10], we applied this technique to study the structure and time evolution of the Dengue virus (serotype I) population in Cambodia, using heterochronous sequence samples. Beside its methodological aspect, this work also introduced new topics related to population genetics in BAMBOO (e.g. the use of coalescent theory to reconstruct population dynamics). Another approach, widely used in ecology, to measure biodiversity and to determine the species composition of environmental samples is the "DNA-barcoding" technique. In collaboration with the Laboratoire d’Ecologie Alpine (Univ. Joseph Fourier Grenoble), we developed a new software for identifying new barcode markers and their associated PCR primers [19].
5. New Results

5.1. Astrocyte Regulation of Synaptic Depression and Facilitation

**Participants:** Hugues Berry, Maurizio De Pitta, Vladislav Volman, Eshel Ben-Jacob.

Synaptic plasticity is the capacity of a preexisting connection between two neurons to change in strength as a function of neuronal activity. Because it admittedly underlies learning and memory, the elucidation of its constituting mechanisms is of crucial importance in many aspects of normal and pathological brain function. Short-term presynaptic plasticity refers to changes occurring over short time scales (milliseconds to seconds) that are mediated by frequency-dependent modifications of the amount of neurotransmitter released by presynaptic stimulation. Recent experiments have reported that glial cells, especially hippocampal astrocytes, can modulate short-term plasticity, but the mechanism of such modulation is poorly understood. Here, we explore a plausible form of modulation of short-term plasticity by astrocytes using a biophysically realistic computational model. Our analysis indicates that astrocytes could simultaneously affect synaptic release in two ways. First, they either decrease or increase the overall synaptic release of neurotransmitter. Second, for stimuli that are delivered as pairs within short intervals, they systematically increase or decrease the synaptic response to the second one. Hence, our model suggests that astrocytes could transiently trigger switches between paired-pulse depression and facilitation. This property explains several challenging experimental observations and has a deep impact on our understanding of synaptic information transfer [16].

5.2. A theory of rate coding control by intrinsic plasticity effects

**Participants:** J Naudé, J Paz, Hugues Berry, Bruno Delord.

Intrinsic plasticity (IP) is a ubiquitous activity-dependent process regulating neuronal excitability and a cellular correlate of behavioral learning and neuronal homeostasis. Because IP is induced rapidly and maintained long-term, it likely represents a major determinant of adaptive collective neuronal dynamics. However, assessing the exact impact of IP has remained elusive. Indeed, it is extremely difficult to disentangle the complex non-linear interaction between IP effects, by which conductance changes alter neuronal activity, and IP rules, whereby activity modifies conductance via signaling pathways. Moreover, the two major IP effects on firing rate, threshold and gain modulation, remain unknown in their very mechanisms. Here, using extensive simulations and sensitivity analysis of Hodgkin-Huxley models, we show that threshold and gain modulation are accounted for by maximal conductance plasticity of conductance in two separate domains of the parameter space corresponding to sub- and supra-threshold conductance (i.e. activating below or above the spike onset threshold potential). Analyzing equivalent integrate-and-fire models, we provide formal expressions of sensitivities relating to conductance parameters, unraveling unprecedented mechanisms governing IP effects. Our results generalize to the IP of other conductance parameters and allow inference of calcium-gated conductance, yielding a general picture that accounts for a large repertoire of experimental observations. The expressions we provide can be combined with IP rules in rate or spiking models, offering a general framework to systematically assess the computational consequences of IP of pharmacologically identified conductance with both fine grain description and mathematical tractability. We provide an example of such IP loop model addressing the important issue of the homeostatic regulation of spontaneous discharge. Because we do not formulate any assumptions on modification rules, the present theory is also relevant to other neural processes involving excitability changes, such as neuromodulation, development, aging and neural disorders [20].

5.3. Impact of receptor clustering on ligand binding

**Participants:** Hédi A. Soula, Bertrand Caré.
Cellular response to changes in the concentration of different chemical species in the extracellular medium is induced by ligand binding to dedicated transmembrane receptors. Receptor density, distribution, and clustering may be key spatial features that influence effective and proper physical and biochemical cellular responses to many regulatory signals. Classical equations describing this kind of binding kinetics assume the distributions of interacting species to be homogeneous, neglecting by doing so the impact of clustering. As there is experimental evidence that receptors tend to group in clusters inside membrane domains, we investigated the effects of receptor clustering on cellular receptor ligand binding. We implemented a model of receptor binding using a Monte-Carlo algorithm to simulate ligand diffusion and binding. In some simple cases, analytic solutions for binding equilibrium of ligand on clusters of receptors are provided, and supported by simulation results. Our simulations show that the so-called “apparent” affinity of the ligand for the receptor decreases with clustering although the microscopic affinity remains constant. Changing membrane receptors clustering could be a simple mechanism that allows cells to change and adapt their affinity/sensitivity toward a given stimulus [14].

5.4. Illegitimate and Homologous Rearrangements, an intricate relationship?

Participants: David P. Parsons, Guillaume Beslon, Carole Knibbe.

We have introduced homologous horizontal transfer in the aevol model. First results show that this process interacts in a complex way with both non-homologous transfer (that creates homologous sequences) and mutation rate (that destroys homologous sequences). We have shown that homologous transfer is useful only in strict conditions (small mutation rate and small - but non-null - non-homologous rate) [27]. This result confirms the genericity of the indirect selection mechanism we previously shown to be at the origin of scaling laws in genomes and transcriptomes [25].

5.5. An error threshold due to chromosomal rearrangements

Participants: Stephan Fischer, Guillaume Beslon, Carole Knibbe, Samuel Bernard.

Error threshold is a well known theory in evolutionary biology. However, the theory of error threshold only takes into account point mutation rate and states that this rate generates a maximum level of coding sequences in a given genome. Using aevol and mathematical formulations, we have shown other types of mutations are also likely to create error thresholds. In particular we have shown the chromosomal rearrangement (duplications and deletions) generate a very strong threshold.

5.6. Enhanced Stimulus Encoding Capabilities with Spectral Selectivity in Inhibitory Circuits by STDP

Participants: Guillaume Beslon, Hédi A. Soula, Antoine Coulon.

The ability to encode and transmit a signal is an essential property that must demonstrate many neuronal circuits in sensory areas in addition to any processing they may provide. It is known that an appropriate level of lateral inhibition, as observed in these areas, can significantly improve the encoding ability of a population of neurons. We have shown that a homeostatic mechanism by which a spike-timing-dependent plasticity (STDP) rule with a symmetric timing window (swSTDP) spontaneously drives the inhibitory coupling to a level that ensures accurate encoding in response to input signals within a certain frequency range. Interpreting these results mathematically, we find that this coupling level depends on the overlap of spectral information between stimulus and STDP window function. Generalization to arbitrary swSTDP and arbitrary stimuli reveals that the signals for which this improvement of encoding takes place can be finely selected on spectral criteria. We finally show that this spectral overlap principle holds for a variety of neuron types and network characteristics. The highly tunable frequency-power domain of efficiency of this mechanism, together with its ability to operate in very various neuronal contexts, suggest that it may be at work in most sensory areas [15].
6. New Results

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

The production of red blood cells, erythropoiesis, occurs in the bone marrow, where immature erythroid cells differentiate and produce red blood cells. Differentiation and maturation of immature red blood cells occurs in very specific spatial structures called erythroblastic islands. They consists of a macrophage (big white blood cell) surrounded by immature cells and providing them with survival factors. Using a hybrid model, made of a discrete model describing cell-cell interactions and accounting for spatial interactions, and a continuous model describing intracellular protein regulation (deciding for cell fate), we showed the importance of the central macrophage in the erythroblastic island, in order to prevent unstable islands leading either to cell populations extinction or excessive proliferation. This result is actually under review (Journal of Theoretical Biology), partial results have been already published (Math. Model. Nat. Phenom.)

6.2. Modelling of the CD8 T cell Immune Response

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. This model, a system of ordinary and age-structured partial differential equations, allows to describe 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. We are now confronting the model to experimental data being generated by Jacqueline Marvel’s team in Lyon (immunology team). The analysis of the model and the first results have been published in Journal of Biological Systems.

6.3. Particle Dynamics Methods of Blood Flow Simulations

Various particle methods are widely used to model dynamics of complex media. In our work molecular dynamics and dissipative particles dynamics are applied to model blood flows composed of plasma and erythrocytes. The properties of the homogeneous particle fluid are studied. Capillary flows with erythrocytes are investigated.

6.4. Periodic linear cell cycle models

Several results on periodic linear cell cycle models were obtained in collaboration with Frederique Billy (Inria Bang), Jean Clairambault (Inria Bang), Olivier Fercoq (Inria Maxplus), Stéphane Gaubert (Inria Maxplus) and Thomas Ouillon (Ensta). Those results are currently in revision (minor modifications requested). It deals with the property of the growth rate in such models. Several aspects are discussed, among wich: effect of the variability of the division process on the growth rate, fitting coefficients to data obtained by FUCCI methodology and optimization procedure for the growth rate.

6.5. Relaxed cross diffusion models

A general well posedness result for relaxed cross diffusion models was obtained by members of our team in collaboration with Michel Pierre and Guillaume Rolland (IRMAR, Rennes). A paper in this subject has been submitted.
5. New Results

5.1. Qualitative modeling, simulation, analysis, and verification of gene regulatory networks

**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 8.3 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 [15]. 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* (see also [14]). A paper on the use of temporal logic and formal verification in the context of GNA appeared in *Theoretical Computer Science* this year [7], in a special issue associated with the conference Computational Methods in System (CMSB), held in Rostock in 2008.

Notwithstanding the above improvements of the software, most of our efforts in the past year have gone into applications in collaboration with users of GNA. For example, Delphine Ropers has worked with several groups at IST Lisbon on the modeling of the FLR1 network in yeast, resulting in a paper in *IET Systems Biology* [8]. The paper reports on the qualitative modelling and simulation of the transcriptional regulatory network controlling the response of the model eukaryote *Saccharomyces cerevisiae* to the agricultural fungicide mancozeb. The model has allowed the analysis of the regulation level and activity of the components of the mancozeb-induced network controlling the transcriptional activation of FLR1. This gene is proposed to confer multidrug resistance to the cell through its putative role as a drug efflux pump. Formal verification analysis of the network allowed us to confront model predictions with experimental data and to assess the model robustness to parameter ordering and gene deletion. This analysis led to a better understanding of the mechanisms regulating the response of FLR1 to mancozeb and confirmed the need for a new transcription factor to account for the full transcriptional activation of the gene *YAP1*. The result is a model of the response of FLR1 to mancozeb, permitting a quick and cost-effective test of hypotheses prior to experimental validation.

As another example of the use of GNA, Hidde de Jong has contributed to the modeling of the TOL system in *Pseudomonas putida*, carried out at the Spanish National Biotechnology Center (CNB). The gene regulatory network of the TOL plasmid pWW0 of the soil bacterium *Pseudomonas putida* mt-2 for catabolism of m-xylene is an archetypal model for environmental biodegradation of aromatic pollutants. Although nearly every metabolic and transcriptional component of this regulatory system is known in detail, the complexity of its architecture is still perplexing. To gain an insight into the inner layout of this network a PL model of the TOL system was implemented, simulated and experimentally validated by measuring the expression of the genes encoding the regulators XylR and XylS when specific portions of the network were activated with selected inducers (m-xylene, o-xylene, 3-methylbenzyloxy alcohol and 3-methylbenzoate). This analysis made sense of the specific regulatory topology on the basis of an unprecedented network motif in the genetic circuit for m-xylene catabolism. The motif appears to ensure a simultaneous expression of the upper and lower segments of the m-xylene catabolic route that would be difficult to bring about with a standard substrate-responsive single promoter. Furthermore, it is plausible that the motif helps to avoid biochemical conflicts between competing plasmid-encoded and chromosomally-encoded pathways in this bacterium. The analysis of the TOL system has been published in *BMC Systems Biology* [11].
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 are actually 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 Omaya Dudin and Stephan Lacour, the validation of a model of the network of global regulators of transcription by Sara Berthoumieux and Hidde de Jong, and the analysis of the regulation of cAMP levels in the bacterial cell by Claire Villiers.

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. We previously showed, in a paper published in the *IEEE/ACM Transactions on Computational Biology and Bioinformatics* this year, that PL models provide a good approximation of the direct and indirect interactions occurring in gene regulation [10].

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 short, preliminary paper on this work was presented at an invited session of the 18th IFAC World Congress held in Milano [12] and a long paper has been accepted for publication in the *Journal of Theoretical Biology* [2].

### 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 have developed an approximate model of central metabolism of *E. coli*, using so-called linlog functions to approximately describe the rates of the enzymatic reactions. More precisely, linlog models describe metabolic kinetics by means of a linear model of the logarithms of metabolite concentrations. We have used metabolome and transcriptome data sets from the literature to estimate the parameters of the linlog models, a task in principle greatly simplified by the mathematical form of the latter. However, a major problem encountered during parameter estimation was the occurrence of missing data, due to experimental problems or instrument failures. In the framework of her PhD thesis, Sara Berthoumieux has addressed the missing-data problem by developing an iterative parameter estimation approach based on an Expectation-Maximization (EM) procedure. This approach adapted from the statistical literature has the advantage of being well-defined analytically and applicable to other kinds of linear regression problems with missing data. It has been tested on simulations experiments with missing data and performs well compared to basic and advanced regression methods.

On the biological side, we have applied the method to a linlog model of central metabolism in *Escherichia coli*, consisting of some 23 variables. We estimated the 100 parameters of this model from a high-throughput dataset published in the literature. The data consists of measurements of metabolic fluxes and metabolite and enzyme levels in glucose-limited chemostat under 29 different conditions such as wild-type strain and single-gene mutant strains or different dilution rates. Standard linear regression is difficult to apply in this case due to missing data, which disqualifies for 7 reactions too many datapoints, leaving a dataset of size inferior to the
number of parameters to estimate. Application of our approach allows one to compute reasonable estimates for most of the identifiable model parameters even when regression is inapplicable. The method and its application to the linlog model of central metabolism in *E. coli* are the subject of a paper accepted for the ISMB/ECCB conference this year and published in a special issue of *Bioinformatics* [3]. Sara Berthoumieux received the Ian Lawson Van Toch Memorial Award for outstanding student paper at ISMB/ECCB. In the continuation of this work, we are currently preparing for submission a journal paper on the identifiability of linlog models.

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 theses of Stéphane Pinhal and Valentin Zulkower, which started at the end of this year, 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 proposed 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 which we refer to as models with unate-like structure. In Boolean network modeling, unate functions are argued to capture virtually all observable interactions in gene regulatory networks. In our quantitative framework, unate logics are encoded in the structure of the nonlinear synthesis rates of the network proteins. This framework allows us to integrate *a-priori* information on the most likely network structures, and the models enjoy monotonicity properties that can be exploited to simplify the identification task.

As described in previous work, published in *Bioinformatics* in 2010, the key idea is to divide the identification process into two steps. In the first step, different monotonicity properties of different model structures are exploited to discard those structures whose property is falsified by the observed data points (time-lapse protein concentrations and synthesis rates). In the second step, the parameters of the model structures not discarded in the first step are fitted to the data in the search of the simplest structure explaining the data with sufficient accuracy. The procedure was validated on challenging data from the literature.

On the methodological side, in the context of the same international collaboration, the identification approach has been further developed. For important subclasses of unate models, larger sets of network structures can now be discarded in the hypothesis falsification step, based on additional properties other than monotonicity (namely quasi-convexity). These improvements have been presented at the 18th IFAC World Congress held in Milan, and are reported in a journal paper that has been accepted for publication in the *International Journal of Robust and Nonlinear Control*, in a special issue on system identification for biological systems. In the framework of the PhD thesis of Diana Stefan, in collaboration with Eugenio Cinquemani, Stephan Lacour and Omaya Dudin, the method is now being applied to experimental data produced within IBIS for the study of the gene network regulating motility of *E. coli* bacteria.

Woei-Fuh Wang, who defended her PhD thesis carried out under the supervision of Johannes Geiselmann and Chung-Ming Chen in December 2011, addressed a different problem in the structural identification of gene regulatory networks. The inference of the network topology and regulatory mechanisms is complicated by the fact that we usually do not know all relevant genes that need to be taken into account for explaining the observed expression patterns. The aim of the thesis was to detect the presence of such “missing genes”, as
well as their regulatory roles and expression patterns. Using a well-known class of simplified kinetic models, based on power-law approximations of synthesis rate functions, an inference algorithm was developed. The algorithms are based on factor analysis, a well-developed multivariate statistical analysis approach that is used to investigate unknown, underlying features of a dataset, as well as independent component analysis. The proposed method of inferring the expression profile of a missing gene and connecting it to a known network structure has been applied to artificial networks, as well as a real network studied within IBIS: the aces regulatory network in *Escherichia coli*.

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 bacterial populations and even of single cells, with applications in for example biotechnology and medicine.

In the literature, much effort has been devoted to the analysis of stochastic gene expression models derived from biochemical kinetics and specific knowledge of the systems at hand. Less effort has been dedicated to developing general methods for inferring unknown parameter values of these stochastic models from single-cell experimental data. While some strategies have been proposed in the recent literature, no method of general applicability exists. IBIS recently started a new line of research dedicated to the study of stochastic modelling and identification of gene regulatory networks in single cells. This work, coordinated by Eugenio Cinquemani, focuses on simple network modules in bacterial cells. Our reference system is the regulation of the inset of arabinose uptake in *E. coli* upon depletion of glucose.

In the past year we developed a working method for the estimation of unknown network parameters of a simple stochastic model of the arabinose uptake process. The method was tested on simulated data and applied with success to time-lapse fluorescence microscopy data acquired by Guillaume Baptist. This application involved the development by Michel Page of a microscopy data processing program based on a customization of the freely accessible Matlab tool CellTracer. Preliminary results were presented in the poster session of the Conference on Stochastic Systems Biology held in Monte Verità (Switzerland). The work is currently being extended in preparation for a journal publication. A generalization of the method and the investigation of alternative stochastic modelling and identification methodologies are being pursued in parallel. Other ongoing work concerns the study of noise propagation in gene regulatory networks in collaboration with Irina Mihalcescu (Université Joseph Fourier).

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 5.1), 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 Jérome 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.
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.

In 2011, A. Rousseau and J. McWilliams (UCLA) proposed in [62] 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 several international conferences [26], [49].

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 is currently underway (D. Cherel’s 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. Moreover the impact of two numerical schemes (a projection method, and a direct method [67]) on the implementation of the domain decomposition method has been discussed.

Using the direct method in a 2D $x-z$ Navier-Stokes model, D. Cherel has implemented a Schwarz-based domain decomposition method, for which he used the so-called transmission 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 paper is in preparation.

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.

This year, she obtained both numerical and theoretical results on a Laplace equation in a domain that suggests a domain decomposition method with two sub-domains that do not have the same space dimension (see Figure 1). A paper is in preparation.
6.1.3. Numerical schemes for ocean modelling

**Participants:** Laurent Debreu, Jeremie 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 [78], 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 2011, in collaboration with F. Lemarié at UCLA, this work has been extended in order to render the biharmonic diffusion operator scheme unconditionally stable (paper submitted to ocean modelling). This is particularly needed when the slopes between coordinates lines and isopycnals 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 Ph'D, Jeremie 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).

Salinity at 1000 m in the Southwest Pacific ocean is shown in figure 2. The use of traditional upwind biased schemes (middle) exhibits a strong drift in the salinity field in comparison with climatology (left). The introduction of high order diffusion rotated along geopotential surfaces prevents this drift while maintaining high resolution features (right).

6.2. Development of New Methods for Data Assimilation

6.2.1. Variational Data Assimilation with Control of Model Error

**Participants:** Bénédicte Lemieux-Dudon, Arthur Vidard.
One of the main limitations of the current operational variational data assimilation techniques is that they assume the model to be perfect mainly because of computing cost issues. Numerous researches have been carried out to reduce the cost of controlling model errors by controlling the correction term only in certain privileged directions or by controlling only the systematic and time correlated part of the error.

Both the above methods consider the model errors as a forcing term in the model equations. Trémolet (2006) describes another approach where the full state vector (4D field: 3D spatial + time) is controlled. Because of computing cost one cannot obviously control the model state at each time step. Therefore, the assimilation window is split into sub-windows, and only the initial conditions of each sub-window are controlled, the junctions between each sub-window being penalized. One interesting property is that, in this case, the computation of the gradients, for the different sub-windows, are independent and therefore can be done in parallel.

We are implementing this method in a realistic Oceanic framework using OPAVAR/ NEMOVAR as part of the VODA ANR project. An extensive documentation is being produced and should be available along with a first prototype early 2012.

### 6.2.2. Variational Data Assimilation and Control of the Boundary Conditions

**Participant:** Eugène Kazantsev.

A variational data assimilation technique applied to the identification of the optimal discretization of interpolation operators and derivatives in the nodes adjacent to the boundary of the domain is discussed in the context of the shallow water model. It was shown in [8] that control of approximation of boundary derivatives and interpolations can increase the model’s accuracy in boundary regions and improve the solution in general. On the other hand, optimal schemes obtained in this way may not approximate derivatives in a common sense. This may lead to another model physics, violating, for example, impermeability boundary condition.

Experiments with a full non-linear shallow water model in [7] show that controlling the discretization of operators near a rigid boundary can bring the model solution close to observations as in the assimilation window and beyond the 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 [30].
The sensitivity of the shallow water model in the previously described configurations has been studied in detail in [9]. It is shown in both experiments that the 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.

In order to illustrate the influence of optimal discretization of operators near the boundary we compare this influence with now classical data assimilation for identification of the optimal initial conditions of the model. The norm of the difference between the model solution and real observational data is plotted in figure 3. Observed sea surface elevation of the Black sea was assimilated during 50 days (May–June 1992) to identify optimal initial and boundary conditions. After that, models have been integrated forward for 500 days and their solutions have been compared with data. One can see that, starting from optimal initial point, the model remains close to observations during less than 100 days while optimal optimal discretization of operators allows the model to be always closer than the model with default parameters.

Adjoint models, necessary to variational data assimilation have been produced by the TAPENADE software, developed by the TROPICS team.

Figure 3. Evolution of the difference "model–observations".

6.2.3. A Nudging-Based Data Assimilation Method: the Back and Forth Nudging

Participants: Didier Auroux, Jacques Blum, Maëlle Nodet.

The Back and Forth Nudging (BFN) algorithm (see [63]) has been recently introduced for simplicity reasons, as it does not require any linearization, nor adjoint equation, or minimization process in comparison with variational schemes. Nevertheless it provides a new estimation of the initial condition at each iteration.

Previous theoretical results [65] showed that BFN was often ill-posed for viscous partial differential equations. To overcome this problem, we proposed a new version of the algorithm, which we called the Diffusive BFN [2], and which showed very promising results on one-dimensional viscous equations. Experiments on more sophisticated geophysical models, such as Shallow-Water equations and NEMO ocean model are still in progress, in collaboration with University of Nice.

6.2.4. Variational Data Assimilation for locally nested models.

Participants: Eric Blayo, Laurent Debreu, François-Xavier Le Dimet, Emilie Neveu.
The objectives are to study the mathematical formulation of variational data assimilation for locally nested models and to conduct numerical experiments for validation.

The state equations of the optimality system have been written for the general case of two embedded grids, for which several kinds of control (initial conditions, boundary conditions) have been proposed. Both one way and two way interactions have been studied. This last year, we worked on integration of non linear grid interactions in the algorithm. Additionally, the problem of specification of background error covariances matrices has been studied (see [ 85 ]).

In the ANR MSDAG project and Emilie Neveu’s PhD, we continue to work on the subject. Our main interest is on the use of multiscale optimization methods for data assimilation. The idea is to apply a multigrid algorithm to the solution of the optimization problem. The work includes the analysis of the ellipticity of the optimization problem [ 12 ], the comparison of different multigrid methods (Gauss-Newton multigrid method and Full Approximation Scheme) and specific developments for highly non linear problems. To extend previous work on Burgers equation, the Full Approximation Scheme (FAS) and the Newton Multigrid algorithm have been compared in a more complex shallow water model. The results shows good performance of the FAS and also put more interest in the design of the background error covariance matrix.

6.3. Data Assimilation for Ocean Models

6.3.1. Development of a Variational Data Assimilation System for OPA9/NEMO

Participants: Arthur Vidard, Franck Vigilant, Claire Chauvin, Bénédicte Lemieux-Dudon, Pierre-Antoine Bouittier, Laurent Debreu.

We are heavily involved in the development of NEMOVAR (Variational assimilation for NEMO). From 2006, 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), and to get technical and scientific support from Inria Sophia (Automatic adjoint derivation, TROPICS project-team) and ECMWF (Parallelization). This project aimed at avoiding duplication of effort, and at developing a common NEMOVAR platform. It has led to the creation of the VODA (Variational Ocean Data Assimilation for multi scales applications) ANR project.

The project aims at delivering a common NEMOVAR platform based on NEMO platform for 3D and 4D variational assimilation. Following 2009-10 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 3.0 version of NEMO. Two upgrades were done to follow NEMO standard development race. As a consequence, NEMOVAR is also available for NEMO version 3.2, 3.2.1 and 3.2.2 both offer fully parallel features. The local group has developed a python interface engine (PIANO) to perform test and run on NEMO. A constant support to NEMOVAR project is ensured to deliver a focused response on dedicated issue (internal and external interaction).

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 4).

A new topic has been explored this year in the framework of VODA: data assimilation in a framework of nested models. It makes full use of the AGRIF capabilities of NEMO and follows previous work done on a toy model during the PhD thesis of E. Simon. Some early results are available with a global ocean 2 degrees configuration including a 1/2 degree zoom on the Agulhas region (see Figure 5).

As a side project we collaborate with Mercator-Ocean in order to use the adjoint to perform sensitivity analysis with the fourth of a degree global model used for the reanalysis. This collaboration that includes both heavy software developments and challenging scientific investigation, has been going on for 2 years now and is producing interesting results for both part that still need to be published.

Apart from the VODA ANR project, the NEMOVAR working group gets additional financial support by LEFE-Assimilation and the Mercator National Programs.
Figure 4. Surface relative vorticity of a 1/24th of a degree NEMO configuration

Figure 5. Temperature increment at 200m around the zoom on the Agulhas region
6.3.2. Variational data assimilation into highly nonlinear ocean models
Participants: Pierre-Antoine Bouttier, Eric Blayo, Jacques Verron.

The purpose of this study is to explore the behaviour of variational data assimilation methods in a non-linear ocean model. In an eddy-permitting or eddy-resolving ocean model, controlling mesoscale eddies activity is crucial for data assimilation methods. Our goal is to highlight the impact of these non-linearities on the assimilation system. To illustrate this, test experiments are performed with a double-gyre NEMO configuration at different resolutions (1/4°, 1/12°) which mimics Gulfstream-like behaviour in term of eddy system, and an incremental 4D-V AR formulation for the assimilation system.

First, we are mainly interested in observing the impact of the length of assimilation window on the quality of the analyzed trajectory. For that, we are doing twin experiments with 1/4° model, using simulated altimeter data, for different lengths of assimilation window. Helped by diagnoses on error scales, we also attempt to link the non-linear phenomena and error structures observed after assimilation quantitavely and qualitatively.

Then, by increasing the model resolution (and consequently mesoscale eddy activity), we bring to light the sensitivity of our assimilation system to non-linearity by repeating the same experiments on the length of assimilation window, and the same diagnoses about error structures.

6.3.3. Assimilation of Lagrangian Data
Participants: Claire Chauvin, Maëlle Nodet, Arthur Vidard.

When an observation is given at a sequence of positions along the fluid flow, then it can be defined as Lagrangian, from a mathematical point of view. From this sequence of positions (for instance the profiling drifting floats of Argo program), one can deduce important information on the stream that transports the drifters. Such an information has not yet been exploited in an operational framework, although previous works [82] have shown the interest of assimilating this new type of data.

A task of the ANR VODA has thus been defined in order to develop the tools for the variational assimilation of Lagrangian data in the context of NEMOVAR. C. Chauvin is an engineer working on this task. She first constructed the observation operator, which requires the interpolation of the velocity at any point of the domain. This interpolation operator is not linear for the general grids used in NEMO, implying heavy tangent and adjoint operators.

Tangent and adjoint procedures associated to this interpolation method have been developed, as well as the tests of these procedures on the main test configurations GYRE and ORCA2. Their implementation in NEMOVAR has required a specific application, in order to be consistent with the conventions and data structures already present in NEMOVAR. We also performed extensive numerical experiments to assess the impact of Lagrangian data assimilation, and its complementarity with other types of data, and we prepare an article to sum up the results. Preliminary results have been presented at EGU [34].

6.4. Assimilation of Image Data

6.4.1. Direct assimilation of sequences of images

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 enveolps. 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 [86], [90]. 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 [45]. We proved its efficiency and robustness on simulated and experimental geophysical flows [64]. As part of GeoFluids, we are investigating new ways to define distance between each couple of images. One idea was to define this distance as the norm of the apparent motion between two images. This has been done thanks to optical flow methods which turned out to need a specific parametrization for each couple of images. 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.

6.4.2. Assimilation of ocean images

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

In addition with the direct assimilation approach previously described, a particular attention has been given to the cloud occlusion and the representation of the observation errors in the context of ocean image data. Such works will be intensified with the post-doctorate Alexandros Makris that will start his activities in December. The assimilation of images (SST and chlorophyll) provided by geostationary satellites is also studied with the oceanographers of the Laboratoire des Écoulements Géophysiques et Industriels. The objective is here to take benefit from the correlation that exists between image gradients and ocean flow discontinuities that can be exhibited through the computation of Lyapunov coefficients and vectors from numerical ocean models [70].

6.5. Quantifying Uncertainty

6.5.1. Propagation of uncertainties

**Participants:** François-Xavier Le Dimet, Victor Shutyaev.

Basically, geophysical models are suffering of two types of errors:

- errors in the model itself due to approximations of physical processes and their subgrid parametrization and also errors linked to the necessary numerical discretization;
- errors in the observation because of errors of measurements and also errors due to sampling. For instance, many remote sensings observe only radiances, which are transformed into the state variables thanks to complex processes like the resolution of an inverse problem. This is, of course, a source of errors.

Estimating the propagation of errors is an important and costly (in term of computing resources) task for two reasons:

- the quality of the forecast must be estimated
- the estimation of the statistics of errors has to be included in the analysis to have an adequate norm, based on these statistics, on the forecast and also on the observation.

In the variational framework, models, observations, statistics are linked into the optimality system which can be considered as a “generalized” model containing all the available estimation. The estimation of error covariances are estimated both from the second order analysis and the Hessian of the cost function. Numerical experiments have been carried out on a non-linear model [16]. We expect to extent the numerical experiments to a semi-operational model in cooperation with ECMWF.

6.5.2. Sensitivity analysis for West African monsoon

**Participants:** Anestis Antoniadis, Céline Helbert, Clémentine Prieur, Laurence Viry.
6.5.2.1. 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.

6.5.2.2. 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 [91]. 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 [83] in integrated genomic studies which takes into account both aspects. The method uses an $\ell_1$-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 $\ell_2$ 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].

6.5.2.3. 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).

6.5.3. Tracking for mesoscale convective systems

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

6.5.3.1. 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 behaviour of the mesoscale convective systems. We wish to develop stochastic tracking models, allowing for simulating rainfall scenari with uncertainties assessment.

6.5.3.2. 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 [88] 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 [80] or by [79]. The model developed by [88] 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.

6.5.4. Sensitivity analysis for forecasting ocean models

Participants: Eric Blayo, Maëlle Nodet, Clémentine Prieur, Gaëlle Chastaing, Alexandre Janon, Jean-Yves Tissot.

6.5.4.1. 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.
6.5.4.2. Estimating sensitivity indices

A first task is to develop tools for estimated sensitivity indices. Among various tools a particular attention was paid to FAST and its derivatives. In [89], 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.

6.5.4.3. 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 [61], 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 [61]. Alexandre Janon obtained a best poster award on the topic [40]. An ongoing work deals also 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.

6.5.4.4. 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 [75] in the case where input parameters are correlated. A PhD started in October 2010 on this topic (Gaëlle Chastaing). We obtained first results which should be submitted soon, deriving a general functional ANOVA for dependent inputs, allowing defining new variance based sensitivity indices for correlated inputs.

6.5.5. Quantification of uncertainty with Multi-fidelity computer experiments

Participants: Federico Zertuche, Céline Helbert, Anestis Antoniadis.

Propagation of uncertainties through computer codes is a hard task when dealing with heavy industrial simulators. Confidence intervals announced on predictions are often huge because of the lack of data. The context of the study here is the case of simulations when multiple levels of analysis (fast and slow) are available. In most cases the fast (but less trustworthy) and the slow (but more accurate) response values can be obtained independently. Thus, we can learn more about the response by additionally measuring the cheap function(s) on a large number of x ‘s. In most cases the relationship between cheap and expensive responses is modeled by an autoregressive Gaussian regression. This method is a natural extension of the kriging method
in the sense that to build the surrogate one performs a Gaussian regression for the cheap data and one for the difference vector defined by the autoregressive relationship. The prediction error depends on the prediction error of the cheap and expensive surrogates. We can observe that this modeling greatly improves the traditional kriging method when the actual relationship between the cheap and expensive responses is somewhat linear. On another hand, this approach gives worse results when the relation between cheap and expensive is far to be linear. Therefore some improvements must be made on the models to take into account a more precise link between the two levels fidelity of the responses. Some other additional tasks concern the associated numerical designs (must the designs be absolutely nested?) and the allocation of resources between low and fast runs. The work is currently the object of the thesis of Federico Zertuche that has just begun in October 2011.

6.5.6. Impact of the thermodynamics and chemical kinetics parameters at different scales for the models of CO2 storage in geological media

Participant: Céline Helbert.

In collaboration with Bernard Guy and Joharivola Raveloson (Ecole des Mines de Saint-Etienne) we study the water-gas-rock interactions in the case of CO2 storage in geological environment. The focus is 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 of the fluid and influence reaction rates at the pore as well as at the reservoir scale. We propose to evaluate the geostatistical characteristics of local variability thanks to simulations of reactive transport on a small scale in which parameters (namely the equilibrium constants $\log K$ and the rate constant $k$) are perturbed to represent local processes. This contribution is the following of a precedent study of the impact of the reservoir uncertainties on the CO2 storage [72].

6.6. Inverse methods for Glaciology

6.6.1. Dating ice matrix and gas bubbles

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

Dating ice matrix and gas bubbles of ice-cores is essential to study paleoclimates. Inverse modelling implemented on 1D ice flow models is being applied for a few years to construct the ice chronology of several deep ice cores. Such a method based on a Monte Carlo algorithm was implemented under the assumption of perfect ice flow models, for one core at a time, and without including the inverse modelling of the densification models, which enables to construct the gas age scale. This approach faces three issues: 1) frequent discrepancies between core chronologies (lack of stratigraphic links between cores as data constraints), 2) frequent failure to verify relevant data constraints (perfect model assumption is too strong), and 3) frequent inconsistency between gas and ice age scales.

A new approach was proposed during the B. Lemieux-Dudon’s PhD to circumvent these restrictions. It introduces the model error in terms of correction functions on three key quantities from which one can calculate both the ice and gas age scales: a) the accumulation rate, b) the total thinning function, and c) the close-off depth in meters of ice equivalent (i.e. depth below the ice-sheet surface where the atmosphere is trapped). A variational formulation of the inverse problem is constructed. It includes several ice cores with background scenarios and paleo data as constraints, among which:

- stratigraphic links between pair of ice cores (methane, tephra, cosmogenic isotopes, etc.) to derive consistent dating between cores,
- ice and gas age markers, as well as delta-depth data (i.e., in situ depth interval between gas and ice of the same age), which enable to optimize the gas and ice age scales simultaneously.

The cost function includes covariance error matrices, and confidence intervals of the solution can be assessed. This method was applied to derive simultaneously a common age scale for the North Grip core and for the two EPICA cores (DML and DC). [76].
This method arouses some interest in the glaciological and paleo community ([68], [87]). Some further developments are however mandatory to ensure the robustness of the dating solution: (i) code optimization, (ii) diagnostics on the assimilation system, and (iii) calibration of the background error covariance matrices.

H. Toye Mahamadou Kele (a 2-year INRIA young engineer contract) joined the MOISE team to modify the code. During the first year, he implemented a shared memory parallelization of the code, and he currently works on the calibration of the covariance error matrices by implementing a posteriori diagnostics.

6.6.2. Inverse methods 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.

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 where/when/what we have to add new observations. 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.

The ANR project ADAGe started one year ago on this subject, and B. Bonan started his PhD in September 2010. During his master internship, he implemented the adjoint code of a simplified ice-sheet flow-line model, Winnie, developed by C. Ritz at LGGE. We then performed twin experiments of data assimilation. These preliminary results have been presented at two international conferences ([32], [33]).

We then implemented Ensemble Kalman Filter (EnKF) on Winnie, which we would like to compare to variational assimilation methods. Coding and testing for the EnKF are still in progress.

6.6.3. Inverse methods for full-Stokes glaciology models

Participants: Olivier Gagliardini, Maëlle Nodet, Catherine Ritz.

We are also interested in inverse modelling for another class of glaciology models, called full-Stokes models. 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 submitted for publication and presented at EGU and AGU ([36], [37]).

With O. Gagliardini, F. Gillet-Chaulet and M. Jay-Allemand (LGGE), we also implemented these methods to study a complex phenomenon: the surge of Variegated glacier (Alaska). This glacier is indeed known to surge periodically, that is to accelerate suddenly during 1-2 years, and then come back to a quiescent phase during 10-20 years. Inverse modelling allowed us to infer changes in basal conditions from surface velocities, and to come to a better understanding of the surge phenomenon, as seen in Figure 6. These results have been published ([6] and presented at AGU ([42]).
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 [81] 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.

6.7.2. Optimal transport
Participants: Maëlle Nodet, Nicolas Papadakis, Arthur Vidard.

A new activity on optimal transport has been started in collaboration with the EDP and MGMI teams of the Laboratoire Jean Kuntzmann, Grenoble and the MAP5 Laboratory, Paris, through a project funded by the ANR white program. The purpose is to define metric between images involving the so-called Wasserstein distance. Such metric would be of particular interest in order to introduce pertinent observation operators for assimilating image data 6.4. Other applications including image morphing and histogram equalization are also studied for image processing purposes. First results has been obtained by adding spatial regularity to the transport map computed by the Benamou-Brenier algorithm [66].

6.7.3. Computer vision
Participant: Nicolas Papadakis.
In collaboration with Vicent Caselles (Pompeu Fabra University, Barcelona, Spain), different image processing works have been finalized: 3D reconstruction and novel view synthesis for soccer replays [13], stereo inpainting for 3D movie post-production [29]. Other works dedicated to object tracking in image sequences have also been proposed with Aurelie Bugeau (LABRI, Bordeaux) [14]. A main attention is now given to the problem of histogram equalization of different images [15]. Our aim is now to include spatial information on color repartition during the histogram transfer for inpainting and shadow removal purposes. A journal paper dealing with this issue has been recently accepted for publication.

6.8. Multivariate risk indicators

Participant: Clémentine Prieur.

In collaboration with Véronique Maume-Deschamps, Elena Di Bernardino (ISFA, Lyon 1) 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 one other is submitted. The submitted one deals with the estimation of bivariate tails [56]. In [69] we propose to minimize multivariate risk indicators by using a Kiefer-Wolfowitz approach to the mirror stochastic algorithm. In [4] we present an estimation procedure for multivariate risk indicators making use of a plug-in estimation of level sets of bivariate cumulative distribution functions.

6.9. Stochastic Downscaling Method

Participant: Antoine Rousseau.

In collaboration with TOSCA (Inria Sophia-Antipolis), LMD (Ecole Polytechnique) and CETE (Clermont-Ferrand), we investigate a new method for the numerical simulation of the wind at small scales. In this work, we consider a new approach for the downscaling in CFD, The local model that we propose is inspired from S.B. Pope’s previous works on turbulence. We investigated a new numerical simulation method for the downscaling in CFD, with a strong orientation in applications to meteorology, particularly for the simulation of wind at small scales. The local model that we propose consists in modeling the fundamental equations of fluid motion by a stochastic Lagrangian model describing the behaviour of a fluid particle. Because of the both Lagrangian and stochastic nature of our model, it is discretized thanks to an interacting particle system, combining a time Euler scheme for stochastic differential equations and a Monte–Carlo approximation method. This model called SDM (Stochastic Downscaling Method) is adapted from previous works introduced by S.B. Pope [84] (see http://sdm.gforge.inria.fr/Accueil/index.en.php).

This year, we worked on the comparison of the SDM model (endowed with a physical geostrophic forcing and a wall log law) with simulations obtained with a LES method (Méso-NH code) for the atmospheric boundary layer (from 0 to 750 meters in the vertical direction), in the neutral case. This work allowed to deeply understand the contribution of each elements of the Lagrangian model in terms of the turbulence production and dissipation, we analyzed the returns of various closure parametrization approaches, including viscosity turbulent approach. We also investigated anisotropic effect, with the introduction of GLM model in SDM (see [84]), in particular the isotropic relaxation case. We gave our conclusions as a part of the final report for ADEME [58]. A paper is in preparation.

6.10. Mathematical modelling for CFD-environment coupled systems

Participant: Antoine Rousseau.

6.10.1. Minimal-time bioremediation of natural water resources

The objective of this work is to provide efficient strategies for the bioremediation of natural water resources. Based on a previous paper [74] that deals with an homogeneous resource in space (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 corresponding optimal discharge of a (small) bioreactor. A first mathematical model has been introduced (see [48]) and a journal paper is ready to be submitted.
6.10.2. Mathematical modelling for the confinement of lagoons

This work deals with the concept of confinement of paralic ecosystems. It is based on the recent paper [73] that presents a modelling procedure in order to compute the confinement field of a lagoon. In [59], A. Rousseau and E. Frénod (INRIA CALVI), 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.
NUMED Project-Team (section vide)
STEEP Exploratory Action

5. New Results

5.1. Definition of the research priorities, and first works

As STEEP addresses pluri-disciplinary research topics that are new both for INRIA and for the STEEP members, our very first goal of this year has been to define and specify the four research priorities described above: (a) development of numerical systemic models (economy / society / environment) at local scales; (b) calibration and convergence of integrated models; (c) consideration and management of uncertainties in integrated models; (d) environmental impacts of urban policies.

To gain better insight into these issues, we choose to start with the TRANUS model, as first experimental framework. We collaborate closely with IDDRI (Institut du Développement Durable et des Relations Internationales, Paris http://www.iddri.org/ ), that has implemented a TRANUS model applied to the city of Grenoble in the context of the AETIC project. We make use of this model as starting point to investigate the three first topics.

(a) We have started a project consisting in designing and implementing an integrated model by ourselves, in close collaboration with the EDDEN laboratory (laboratoire d’Économie du Développement Durable et de l’Énergie, Grenoble http://webu2.upmf-grenoble.fr/LEPII/spip/spip.php?article22 ) and IDDRI. The goal is to enrich the transport / land use Grenoble model with energy sectors, using the data and analysis our partners are generating in the context of the AETIC project. The goal here is to find a way to encapsulate in or combine the energy models proposed by EDDEN with the TRANUS model implemented by IDDRI for Grenoble.

(b) We participated to the still on-going calibration procedure of the Grenoble models, which is leaded by IDDRI and Modelistica. This helped us to identify the keypoints of the calibration of such a model, and to propose first semi-automatic approaches based on simple parameters optimizations.

(c) We started an uncertainty analysis of this model.

Parallel to that, we have taken an interest in the LEAM model and URBANSIM and we are preparing an ANR project on TRANUS and URBANSIM which would allow us to deeply compare the very different representations on which they are based on.

5.2. Building of Partnerships

This year, we have spent a considerable amount of time and energy to build various partnerships.

We have started a collaboration with Modelistica and Tómas de la Barra, the author of TRANUS Model. Tómas de la Barra came at INRIA for a week in June 2011. We have submitted a project to ECOS-Nord program (TRACER). Among others, we are closely working with him on TRANUS calibration.

Our collaboration with EDDEN is now fully set up. We are working with them on the design of a transport-land use-energy integrated model for the city of Grenoble using the data and analysis they are generating in the context of the AETIC project. This work is based on the TRANUS model implemented by IDDRI for Grenoble. In the same way, our collaboration with IDDRI has started to be effective. With IDDRI, we work on TRANUS calibration issues and practically on the Grenoble model they are implementing. IDDRI is the coordinator of TRACER project submitted to ECOS-Nord program.

In other respects, with AURG (Agence d’Urbanisme de la Région Grenobloise http://www.aurg.org/) and EDDEN, we are one of the driving forces behind the MUTERA project. This project aims at gathering the main actors in urban planning and transport of the Rhône-Alpes region to work on the issues regarding land-use and transport models. This group includes technicians, politicians and researchers. The kick-off meeting should take place in the beginning of 2012.
We have also started a collaboration with the LECA laboratory (Laboratoire d’Ecologie Alpine, Grenoble http://www-leca.ujf-grenoble.fr/ ). In June 2011, we have submitted a project to FRB Flagship program with them (ESNET).

In parallel of that, we got in touch with a number a potential partners and colleagues ones of which are the LET laboratory (Laboratoire d’Economie des Transport, Lyon  http://www.let.fr/ ), the department of urban and regional planning of the university of Illinois at Urbana-Champaign (visit to Brian Deal), Veolia Environnement, SOGREAT etc. A number of them have been contacted in the framework of the development of an ANR project which should be submitted to 2012 “Modèles numériques” program or in the framework of CIFRE or industrial partnerships.

5.3. First results on material flow, production and consumption analysis

This theme of research has been started this year through a 6 month internship with a student of Ecole Centrale de Lyon. A large database has been constructed, that allows to analyze material flows, production and consumption by type (and origin and destination in case of flows) at the level of the French region and departments. This database compiles data from a variety of public institutions and private organizations. Almost all the data are in the public domain. A large amount of work has gone into building the database, normalizing the different data sources at least in terms of classification, and into building tools to manipulate the data and produce usable information.

At present, this “physical accounting” has been performed on all French regions for a limited number of materials (in particular cereals, fossil fuels, and construction materials). Furthermore, it has been performed for the same materials at the departement level in the Rhone-Alpes region.

This work is now focusing on various issues:

- By correlating the relevant quantities with other variables (such as, e.g., population or land use, etc), it is possible to downscale the consumption and waste production data at the level of urban regions (where a large part of this information is unavailable). Such an information is critical for the development of decision-help tools at the level of urban areas.

- A process of quantification of the major source of information on material transports is the SitraM database (transportation database maintained by the Ministry of Ecology and Sustainable Development). These data are statistical in nature, but the statistical error is unknown. We are in the process of developping a way to estimate this error.

- The next step will then be to transform this material use in terms of environmental impact. Various types of impacts (global and local) have been identified. The quantification will rely on various known approaches (Life Cycle Analysis, regional Input-Analysis), notably drawing on part of the work performed in this area by the Global Footprint Network (even for impacts that cannot be quantified in terms of land, carbon or water footprint).
6. New Results

6.1. SouthBound results

6.1.1. L4 micro kernels

As part of our investigations about what software architectures were the best candidates to base our Ambient Middleware Stack upon, we studied different micro-kernel operating systems such as CodeZero [33], OKL4, and L4/Fiasco. The objective here is to try and quantify the development effort that would be needed before being able to execute a Java application on top of a micro-kernel. These studies included, in addition to a lot of bibliographic research, several technical experiments such as booting each of these various micro-kernel systems in QEmu, as well as on real hardware. We use a BeagleBoard as a representative example of the kind of hardware platforms typically encountered in Ambient Intelligence scenarios.

6.1.2. Virtual machines

In parallel to our study of micro-kernel architectures, we worked on virtual machines as well, in the perspective of bridging the gap between the two. The basic question here is: what does it takes to to cut down a Java virtual machine into pieces so as to run each of these pieces as a separate software component in the system. We ran two actions in order to investigate this question. First, we ported the JamVM virtual machine to run on top of the Genode operating system framework [37]. This provided us with better understanding of what are the real requirements of a Java virtual machine in terms of underlying operating systems support. Second, we focused on one particular service of the virtual machine, the garbage collector, and we precisely identified and studied the coupling between this component and all other parts of the virtual machine (bytecode interpreter, scheduler, etc). This work was done as part of a student summer internship (Yann Chevalier, INSA-Lyon 3IF). Removing a garbage collector at runtime, and “plugging in” another one dynamically proved to be vastly harder than expected. Still, this work provided us with great insights about the coupling relationships between different OS components.

6.1.3. HiKoB

Antoine Fraboulet (Amazone team), Guillaume Chelius (D-NET team) and Christophe Braillon (INRIA SED) started a new company called HiKob http://openlab.hikob.com/ in July 2011. HiKob is a development project following several successful research projects completed these last 6 years at INSA Lyon and INRIA. HiKob hardware and software products help in building complex, large-scale and distributed applications in the domains of: motion capture, biomechanical study, biologging study, building instrumentation and many more applications targeting wireless sensor network solutions for distributed and embedded measurement. HiKob business model is built on two major directions: complete solutions for industrial applications and software and hardware tools for research and innovation in the fields of sensor networking and embedded wireless measure. HiKob is supported by IT-Translation and INSAValor.

6.1.4. Service-Oriented Tainted Object Propagation

Many Java technologies allow the execution of code provided by multiple parties. Service-oriented platforms based on components such as OSGi are good examples of such a scenario. Those extensible component-based platforms are service-oriented, as components may directly interact with each other via the services they provide. However, even robust languages such as Java were not designed to handle safely code interaction between trusted and untrusted parties.

In [38], we show how basic Java interactions can break encapsulation or execution safety and why the Java security layers’ threat coverage is incomplete. We also review flaws in the Java access control design that can allow untrusted code to bypass restrictions by exploiting vulnerabilities in trusted code.
As component-based platforms become more and more integrated to our daily life, we improved our Service-Oriented Tainted Object Propagation technique to find such vulnerabilities and used it on several open-source components to further demonstrate the real exposure that those vulnerabilities bring to the fore.

6.2. NorthBound results

Another key issue in the Amazones architecture was to bring together formal methods and service oriented programming such as OSGi/Java approach. We developed the Logos framework that observes and records communications that occurs between an OSGi client and a corresponding server. This architecture is developed in the LISE ANR project, and we made various improvements to the architecture.

6.2.1. Amazones Protocol

It aims at building automata from running and observing applications. The logos framework observes a running application at builds at run-time an automata that represents the application behavior. Julien Ponge wrote the corresponding Scala code.

6.2.2. Monitored oriented programming

Another Logos extension integrates Monitored Oriented architectures such as JavaMOP and Larva. We are currently using and working with the larva people. Each time we intercept a call, it is transferred to the Larva automata manager.

6.2.3. Real time SOA

Admission control for service oriented application in real time infrastructure. This work led by Lionel Morel tries to bring together real time architecture configuration and component based architectures. The deal is to find a better way of managing the dynamicity of applications in real time context.

6.2.4. B Model Slicing to Generate Tests

In a model-based testing approach as well as for the verification of properties, B models provide an interesting modelling solution. However, for industrial applications, the size of their state space often makes them hard to handle. To reduce the amount of states, an abstraction function can be used. The abstraction is often a domain abstraction of the state variables that requires many proof obligations to be discharged, which can be very time consuming for real applications.

we propose a contribution to this problem that complements an approach based on domain abstraction for test generation, by adding a preliminary syntactic abstraction phase, based on variable elimination. We define a syntactic transformation that suppresses some variables from a B event model, in addition to three methods that choose relevant variables according to a test purpose. This way, we propose a method that computes an abstraction of a source model $M$ according to a set of selected relevant variables. Depending on the method used, the abstraction can be computed as a simulation or as a bi-simulation of $M$. With this approach, the abstraction process produces a finite state system. We apply this abstraction computation to a Model Based Testing process. We evaluate experimentally the impact of the model simplification by variables elimination on the size of the models, on the number of proof obligations to discharge, on the precision of the abstraction and on the coverage achieved by the test generation.

This work is based on a B model approach. However, in the context of AMAZONES, one of our objectives is to extend it, in order to consider models automatically generated from the usage a the tested service on a particular context.

6.2.5. Distributed Data Centric Programs Verification

Netlog is a language designed to describe distributed programs. It has a precise semantics, provides a high-level of abstraction thanks to its datalog flavor and benefits from an efficient implementation. This makes it a very interesting target language for proofs of distributed programs. In [34], with the Coq proof assistant, we formalized the distributed computation model based on message passing with either synchronous or
asynchronous behaviors; built the translation of Netlog programs; modeled the embedded machine evaluating
Netlog programs, and thus established a framework to formally verify properties of distributed programs in
Netlog. To test the framework, we proved the correctness of a concrete distributed program for constructing
spanning trees over connected graphs.

6.2.6. Managing dynamic service substitution at runtime

The service oriented approach is a paradigm allowing the introduction of dynamicity in developments. If
there are many advantages with this approach, there are also some new problems associated to service
disappearance. The particular case of service substitution is often studied and many propositions exist.
However, proposed solutions are mainly server-side in the context of web-services.

In this work, we propose a client side API-based approach to allow service substitution without any restart
of the client and without any assumption on external services. Our proposition is based on a transactional
approach, defined to authorized substitutions of services dynamically, by preserving the current run and
collected data.

We designed a framework organized by Julien Ponge [14].

6.3. Application domain results

An emerging trend into Amazones team is to apply our northBound/southBound approach to the Internet
OfThing wave. We try to apply our architectures to the IoT application domain.

6.3.1. Data Centric Applications Distribution

Peer to peer systems have been widely used to alleviate the burden of servers by transferring to peers in
a network tasks that do not require a centralization of the information. A wide range of applications are
now emerging over peer-to-peer, such as social networking, multiplayer games, mobile messaging, etc. Most
of these applications are essentially data-centric, they rely on exchange of data between peers, and can be
expressed by queries over the database.

We propose a tool that allows for such applications, programmed as a collection of queries over a database, to
be ported seamlessly without changing the initial queries from a client/server system to peer-to-peer system.
The distribution is done with overlays network defined by declarative data centric programs specified in
the Netlog language, thus resulting in a fully data centric modeling of the peer-to-peer application. The
communication between peers relies on implicit addresses which can be evaluated on the fly to ensure the
persistence of data.

We demonstrate the technique on a multiplayer online game, written in SQL, with players who connect to a
mobile ad hoc network through their portable devices. The overlay is defined by a combination of an ad hoc
routing protocol, DSDV, together with a DHT. The application runs on the QuestMonitor system, which allows
to monitor the communication between peers, the evolution of the local data stores, as well as the execution of
the declarative code.

6.3.2. Service Deployment in Disrupted Networks

OLD / REMOVE ? Ambient environments classically use wireless connections that suffers from frequent
disconnections. The hard research point is to ensure service continuity. This disconnection problem has been
widely tackled for application data with proxy and prefetching approaches. For services, disconnections are
more difficult to anticipate, since service calls are only solved at run-time.

We are currently working on service deployment and invocations in disrupted networks with a network coding
approach. This research is a joint work with the Swing team and with Aline Viana (INRIA Saclay @ TU
Berlin). The main idea is to study how social-oriented applications, that need inter-dependent services and
updates to be distributed to all or part of the mobile users community, could benefit from a network coding
approach. The project aims at assessing for the first time the performance, in terms of latency, energy efficiency
and capacity, of standard network coding techniques in presence of realistic user mobility and service demands.
Building on these results, we plan to propose original social-aware network coding techniques that take advantage of the heterogeneous nature of the opportunistic network to reduce delays and energy consumption, in presence of multiple concurrent service flows targeting either all users or specific groups of interests.

These performance issues tackle at the same time the overall network capacity optimizations, as well as the overall software stack optimizations of a device with local and autonomous network coding strategies.

An INRIA ARC project proposal, entitled SoCool, has been submitted jointly with INRIA AMAZONES, INRIA SWING, INRIA MAESTRO, INRIA Saclay, University of Nice, TU Berlin and Fordham University.
6. New Results

6.1. Exposure to diffusion in dynamic networks

In many contexts, complex networks are subject to diffusion phenomenon, like spreading of epidemics in human groups or the diffusion of information in social networks. Often, the underlying network is dynamic, that is, his links change along time. Clearly, the dynamics of links has an influence on the diffusion phenomenon taking place over the network. A first step to understand these relationship is to determine which nodes of the dynamic network are more likely to be reached by a diffusion process. We designed new notions of exposure in order to do it, based either on contacts, paths or flows in a dynamic network. In particular, the notion of dynamic flow, which we introduced, has given interesting preliminary results. We computed the exposition scores of nodes of real world dynamic networks and showed that it is correlated to the likelihood of nodes to be affected by a diffusion in the classical SI model.

6.2. Aggregation of temporal graph series

A very natural and extensively used way to represent a dynamic network, where links change along time, is to build a graph series : the series of snapshots of the network taken at different time of its evolution. The way to do so is to aggregate all the contact information on a time window into a single graph : that is, we put an edge between u and v in the graph if they are in contact at least once during the considered time window. Doing so for disjoint windows of equal length which cover the whole period of study, we obtain a series of graphs representing the dynamics of the network. A question remain : how one should choose the length of the aggregation window? The problem is critical since depending on the choice made, the properties of the dynamic network are different and the conclusion derived from its analysis may change. We design a systematic method to estimate the maximum possible aggregation length. Up to our knowledge, this is the first method addressing the problem. It is based on activity rate of dynamic paths in the dynamics. On a dynamic path, only some time steps are used to move within the network. When the aggregation time is short, the activity rate of paths is close to zero and it tends to 1 when this time grow until the whole period of experiments. Between the two behaviors, we showed that there is a phase transition that we interpret as the moment when the properties of the dynamics are distorted because of the too long aggregation time.

6.3. Characterizing changes in dynamic networks

Very often, dynamic networks are described as time series of graphs. Many works focus on analyzing or capturing into models the properties of the graphs of the series. This approach has a clear limitation : it looses the relationships between the different graphs of the series, which however contain a key information on the dynamics. In order to get more insight in the relationships between the graphs of the series, we analyzed the structure of we call the difference graphs. The difference graph of two consecutive graphs $G_1, G_2$ of the series is the graph whose edge set is the symmetric difference of the edge sets of $G_1$ and $G_2$. In other words, this is exactly the graph of the pairs whose adjacency relationship changed from $G_1$ to $G_2$. We showed that the structure of difference graphs is very particular : their edges are concentrated around a small number of vertices. This shows that the changes between two graphs of the series are not spread everywhere in the network, but are due to changes of the neighborhood of only a small number of nodes of the network. We could show this fact by computing a graph parameter called Minimum Vertex Cover (MVC), which is Np-complete to compute. Using a preprocessing step, we could compute the exact value of this parameter for all difference graphs of real world series. We obtained that the value of the MVC on difference graphs is very small compared to the expected value on a random graph with same density. We also showed that the most common models of dynamic networks do not capture this property of concentration of edges in the difference graphs of the series. Our result shed light on the way dynamic networks evolve and open the way to significant improvement of existing models.
6.4. Community detection: dynamic and overlapping

Overlapping community detection is a popular topic in complex networks. Comparing to disjoint community structure, overlapping community structure is more reasonable to describe networks at a macroscopic level. Overlaps shared by communities play an important role in combining different communities. We propose two different approaches to detect overlaps: fuzzy community detection and overlapping community detection. The former estimates membership degree of node belonging to community, and the latter allows node to be shared by communities. In this paper, a fuzzy detection and a clique optimization are introduced. Experimental studies in synthetic networks show fuzzy detection yields meaningful information in stability and hierarchy of communities. And clique optimization is efficient in capturing overlapping nodes. Applications in real networks whose community structure is not well-known find that overlapping clusters found by our fuzzy detection can provide different views than general overlapping nodes in characterize overlaps.

Although community detection has drawn tremendous amount of attention across the sciences in the past decades, no formal consensus has been reached on the very nature of what qualifies a community as such. We take an orthogonal approach by introducing a novel point of view to the problem of overlapping communities. Instead of quantifying the quality of a set of communities, we choose to focus on the intrinsic community-ness of one given set of nodes. To do so, we propose a general metric on graphs, the cohesion, based on counting triangles and inspired by well established sociological considerations. The model has been validated through a large-scale online experiment called Fellows in which users were able to compute their social groups on Facebook and rate the quality of the obtained groups. By observing those ratings in relation to the cohesion we assess that the cohesion is a strong indicator of users subjective perception of the community-ness of a set of people.

6.5. Cross-Layer Optimization for Software Layer to Physical Device layer Mapping

We develope a generic method for mapping software state machines used in protocol stacks and communication layers directly to hardware communication devices using their specifications. The proposed method can handle power modes and timing constraints imposed by hardware devices in order to optimize the software code running on top of the device. This property allows the use of the hardware device in its lowest power consumption mode while making sure that real time constraints are met. To validate the merit of the proposed method, the generated code and power consumption gain, we evaluate the optimizations that can be done on a BMAC medium access control layer used in wireless sensor networks using a large scale experimental testbed. The results show that an average energy consumption gain of up to 60% at the radio level can be achieved.
6. New Results

6.1. Scheduling Strategies and Algorithm Design for Heterogeneous Platforms


6.1.1. Virtual Machine Resource Allocation for Service Hosting on Heterogeneous Distributed Platforms

We proposed algorithms for allocating multiple resources to competing services running in virtual machines on heterogeneous distributed platforms. We developed a theoretical problem formulation, designed algorithms, and compared these algorithms via simulation experiments based in part on workload data supplied by Google. Our main finding is that vector packing approaches proposed in the homogeneous case can be extended to provide high-quality solutions in the heterogeneous case, and combined to provide a single efficient algorithm. We also considered the case when there may be errors in estimates of performance-related resource needs. We provided a resource sharing algorithm and proved that for the single-resource, single-node case, when there is no bound on the error, its performance ratio relative to an omniscient optimal algorithm is $\frac{2J}{J-1}$, where $J$ is the number of services. We also provided a heuristic approach for compensating for bounded errors in resource need estimates that performs well in simulation.

6.1.2. Dynamic Fractional Resource Scheduling vs. Batch Scheduling

We finalized this work in which we proposed a novel job scheduling approach for homogeneous cluster computing platforms. Its key feature is the use of virtual machine technology to share fractional node resources in a precise and controlled manner. Other VM-based scheduling approaches have focused primarily on technical issues or extensions to existing batch scheduling systems, while we take a more aggressive approach and seek to find heuristics that maximize an objective metric correlated with job performance. We derived absolute performance bounds and developed algorithms for the online, non-clairvoyant version of our scheduling problem. We further evaluated these algorithms in simulation against both synthetic and real-world HPC workloads and compared our algorithms to standard batch scheduling approaches. We found that our approach improves over batch scheduling by orders of magnitude in terms of job stretch, while leading to comparable or better resource utilization. Our results demonstrated that virtualization technology coupled with lightweight online scheduling strategies can afford dramatic improvements in performance for executing HPC workloads.

6.1.3. Greedy algorithms for energy minimization

This year, we have revisited the well-known greedy algorithm for scheduling independent jobs on parallel processors, with the objective of energy minimization. We have assessed the performance of the online version, as well as the performance of the offline version, which sorts the jobs by non-increasing size before execution. We have derived new approximation factors, as well as examples that show that these factors cannot be improved, thereby completely characterizing the performance of the algorithms.

6.1.4. Energy-aware mappings on chip multiprocessors

This year, in collaboration with Rami Melhem at Pittsburgh University (USA), we have studied the problem of mapping streaming applications that can be modeled by a series-parallel graph, onto a 2-dimensional tiled CMP architecture. The objective of the mapping is to minimize the energy consumption, using dynamic and voltage scaling techniques, while maintaining a given level of performance, reflected by the rate of processing the data streams. This mapping problem turned out to be NP-hard, but we identified simpler instances, whose
optimal solution can be computed by a dynamic programming algorithm in polynomial time. Several heuristics were proposed to tackle the general problem, building upon the theoretical results. Finally, we assessed the performance of the heuristics through comprehensive simulations using the StreamIt workflow suite and various CMP grid sizes.

We are pursuing this work by investigating the routing of communications in chip multiprocessors (CMPs). The goal is to find a valid routing in the sense that the amount of data routed between two neighboring cores does not exceed the maximum link bandwidth while the power dissipated by communications is minimized. Our position is at the system level: we assume that several applications, described as task graphs, are executed on a CMP, and each task is already mapped to a core. Therefore, we consider a set of communications that have to be routed between the cores of the CMP. We consider a classical model, where the power consumed by a communication link is the sum of a static part and a dynamic part, with the dynamic part depending on the frequency of the link. This frequency is scalable and it is proportional to the throughput of the link. The most natural and widely used algorithm to handle all these communications is XY routing: for each communication, data is first forwarded horizontally, and then vertically, from source to destination. However, if it is allowed to use all Manhattan paths between the source and the destination, the consumed power can be reduced dramatically. Moreover, some solutions may be found while none existed with the XY routing.

We have compared XY routing and Manhattan routing, both from a theoretical and from a practical point of view. We considered two variants of Manhattan routing: in single-path routing, only one path can be used for each communication, while multi-paths routing allows to split a communication between different routes. We established the NP-completeness of the problem of finding a Manhattan routing that minimizes the dissipated power, we exhibited the minimum upper bound of the ratio power consumed by an XY routing over power consumed by a Manhattan routing, and finally we performed simulations to assess the performance of Manhattan routing heuristics that we designed.

6.1.5. Power-aware replica placement

We have investigated optimal strategies to place replicas in tree networks, with the double objective to minimize the total cost of the servers, and/or to optimize power consumption. The client requests are known beforehand, and some servers are assumed to pre-exist in the tree. Without power consumption constraints, the total cost is an arbitrary function of the number of existing servers that are reused, and of the number of new servers. Whenever creating and operating a new server has higher cost than reusing an existing one (which is a very natural assumption), cost optimal strategies have to trade-off between reusing resources and load-balancing requests on new servers. We provide an optimal dynamic programming algorithm that returns the optimal cost, thereby extending known results without pre-existing servers. With power consumption constraints, we assume that servers operate under a set of $M$ different modes depending upon the number of requests that they have to process. In practice $M$ is a small number, typically 2 or 3, depending upon the number of allowed voltages. Power consumption includes a static part, proportional to the total number of servers, and a dynamic part, proportional to a constant exponent of the server mode, which depends upon the model for power. The cost function becomes a more complicated function that takes into account reuse and creation as before, but also upgrading or downgrading an existing server from one mode to another. We have shown that with an arbitrary number of modes, the power minimization problem is NP-complete, even without cost constraint, and without static power. Still, we have provided an optimal dynamic programming algorithm that returns the minimal power, given a threshold value on the total cost; it has exponential complexity in the number of modes $M$, and its practical usefulness is limited to small values of $M$. Still, experiments conducted with this algorithm showed that it can process large trees in reasonable time, despite its worst-case complexity.

6.1.6. Reclaiming the energy of a schedule

In this work, we consider a task graph to be executed on a set of processors. We assume that the mapping is given, say by an ordered list of tasks to execute on each processor, and we aim at optimizing the energy consumption while enforcing a prescribed bound on the execution time. While it is not possible to change the allocation of a task, it is possible to change its speed. Rather than using a local approach such as backfilling, we have considered the problem as a whole and studied the impact of several speed variation models on its
complexity. For continuous speeds, we gave a closed-form formula for trees and series-parallel graphs, and we cast the problem into a geometric programming problem for general directed acyclic graphs. We showed that the classical dynamic voltage and frequency scaling (DVFS) model with discrete modes leads to a NP-complete problem, even if the modes are regularly distributed (an important particular case in practice, which we analyzed as the incremental model). On the contrary, the VDD-hopping model leads to a polynomial solution. Finally, we provided an approximation algorithm for the incremental model, which we extended for the general DVFS model.

6.1.7. Workload balancing and throughput optimization

We have investigated the problem of optimizing the throughput of streaming applications for heterogeneous platforms subject to failures. The applications are linear graphs of tasks (pipelines), and a type is associated to each task. The challenge is to map tasks onto the machines of a target platform, but machines must be specialized to process only one task type, in order to avoid costly context or setup changes. The objective is to maximize the throughput, i.e., the rate at which jobs can be processed when accounting for failures. For identical machines, we have proved that an optimal solution can be computed in polynomial time. However, the problem becomes NP-hard when two machines can compute the same task type at different speeds. Several polynomial time heuristics have been designed, and simulation results have demonstrated their efficiency.

6.1.8. Comparing archival policies for BlueWaters

In this work, we focus on the archive system which will be used in the BlueWaters supercomputer. We have introduced two new tape archival policies that can improve tape archive performance in certain regimes, compared to the classical RAIT (Redundant Array of Independent Tapes) policy. The first policy, PARALLEL, still requires as many parallel tape drives as RAIT but pre-computes large data stripes that are written contiguously on tapes to increase write/read performance. The second policy, VERTICAL, writes contiguous data into a single tape, while updating error correcting information on the fly and delaying its archival until enough data has been archived. This second approach reduces the number of tape drives used for every user request to one. The performance of the three RAIT, PARALLEL and VERTICAL policies have been assessed through extensive simulations, using a hardware configuration and a distribution of I/O requests similar to these expected on the BlueWaters system. These simulations have shown that VERTICAL is the most suitable policy for small files, whereas PARALLEL must be used for files larger than 1 GB. We have also demonstrated that RAIT never outperforms both proposed policies, and that a heterogeneous policy mixing VERTICAL and PARALLEL performs 10 times better than any other policy.

6.1.9. Using Virtualization and Job Folding for Batch Scheduling

In this work we study the problem of batch scheduling within a homogeneous cluster. In this context, the problem is that the more processors the job requires the more difficult it is to find an idle slot to run it on. As a consequence the resources are often inefficiently used as some of them remain unallocated in the final schedule. To address this issue we propose a technique called job folding that uses virtualization to reduce the number of processors allocated to a parallel job and thus allows to execute it earlier. Our goal is to optimize the resource use. We propose several heuristics based on job folding and we compare their performance with classical on-line scheduling algorithms as FCFS or backfilling. The contributions of this work are both the design of the job folding algorithms and their performance analysis.

6.1.10. A Genetic Algorithm with Communication Costs to Schedule Workflows on a SOA-Grid

We propose in this work the problem of scheduling a collection of workflows, identical or not, on a SOA (Service Oriented Architecture) grid. A workflow (job) is represented by a directed acyclic graph (DAG) with typed tasks. All of the grid hosts are able to process a set of typed tasks with unrelated processing costs and are able to transmit files through communication links for which the communication times are not negligible. The goal of our study is to minimize the maximum completion time (makespan) of the workflows. To solve this problem we propose a genetic approach. The contributions of this paper are both the design of a Genetic Algorithm taking the communication costs into account and its performance analysis.
6.1.11. **Checkpointing policies for post-petascale supercomputers**

In this work, we provided an analysis of checkpointing strategies for minimizing expected job execution times in an environment that is subject to processor failures. In the case of both sequential and parallel jobs, we gave the optimal solution for exponentially distributed failure inter-arrival times, which, to the best of our knowledge, is the first rigorous proof that periodic checkpointing is optimal. For non-exponentially distributed failures, we developed a dynamic programming algorithm to maximize the amount of work completed before the next failure, which provides a good heuristic for minimizing the expected execution time. Our work considers various models of job parallelism and of parallel checkpointing overhead. We first performed extensive simulation experiments assuming that failures follow Exponential or Weibull distributions, the latter being more representative of real-world systems. The obtained results not only corroborate our theoretical findings, but also show that our dynamic programming algorithm significantly outperforms previously proposed solutions in the case of Weibull failures. We then performed simulation experiments that use failure logs from production clusters. These results confirmed that our dynamic programming algorithm significantly outperforms existing solutions for real-world clusters.

We have also showed an unexpected result: in some cases, when (i) the platform is sufficiently large, and (ii) the checkpointing costs are sufficiently expensive, or the failures are frequent enough, then one should limit the application parallelism and duplicate tasks, rather than fully parallelize the application on the whole platform. In other words, the expectation of the job duration is smaller with fewer processors! To establish this result we have derived and analyzed several scheduling heuristics.

6.1.12. **Scheduling parallel iterative applications on volatile resources**

In this work we study the efficient execution of iterative applications onto volatile resources. We studied a master-worker scheduling scheme that trades-off between the speed and the (expected) reliability and availability of enrolled workers. A key feature of this approach is that it uses a realistic communication model that bounds the capacity of the master to serve the workers, which requires the design of sophisticated resource selection strategies. The contribution of this work is twofold. On the theoretical side, we assess the complexity of the problem in its off-line version, i.e., when processor availability behaviors are known in advance. Even with this knowledge, the problem is NP-hard. On the pragmatic side, we proposed several on-line heuristics that were evaluated in simulation while a Markovian model of processor availabilities.

We have started this study with the simple case of iterations composed of independent tasks that can execute asynchronously. Then we have investigated a much more challenging scenario, that of a tightly-coupled application whose tasks steadily communicate throughout the iteration. In this latter scenario, if one processor computing some task fails, all the work executed for current iteration is lost, and the computation of all tasks has to be restarted. Similarly, if one processor of the current configuration is preempted, the computation of all tasks is interrupted. Changing the configuration within an iteration becomes a much riskier decision than with independent tasks.

6.1.13. **Tiled QR factorization algorithms**

In this work, we have revisited existing algorithms for the QR factorization of rectangular matrices composed of $p \times q$ tiles, where $p \geq q$. We target a shared-memory multi-core processor. Within this framework, we study the critical paths and performance of algorithms such as FIBONACCI and GREEDY, and those found within PLASMA. Although neither is optimal, both are shown to be asymptotically optimal for all matrices of size $p = q^2 f(q)$, where $f$ is any function such that $\lim_{q \to \infty} f = 0$. This novel and important complexity result applies to all matrices where $p$ and $q$ are proportional, $p = \lambda q$, with $\lambda \geq 1$, thereby encompassing many important situations in practice (least squares). We provide an extensive set of experiments that show the superiority of the new algorithms for tall matrices.

We have then extended this work to a distributed-memory environment, that corresponds to clusters of multi-core processors. These platforms make the present and the foreseeable future of high-performance computing. In the context of a cluster of multicores, in order to minimize the number of inter-processor communications (aka, “communication-avoiding” algorithm), it is natural to consider two-level hierarchical reduction trees.
composed of an “inter-node” tree which acts on top of “intra-node” trees. At the intra-node level, we propose a hierarchical tree made of three levels: (0) “TS level” for cache-friendliness, (1) “low level” for decoupled highly parallel inter-node reductions, (2) “coupling level” to efficiently resolve interactions between local reductions and global reductions. Our hierarchical algorithm and its implementation are flexible and modular, and can accommodate several kernel types, different distribution layouts, and a variety of reduction trees at all levels, both inter-cluster and intra-cluster. Numerical experiments on a cluster of multicore nodes confirm that each of the four levels of our hierarchical tree contributes to build up performance and build insights on how these levels influence performance and interact within each other. Our implementation of the new algorithm with the Dague scheduling tool significantly outperforms currently available QR factorization softwares for all matrix shapes, thereby bringing a new advance in numerical linear algebra for petascale and exascale platforms.

6.1.14. Scheduling malleable tasks and minimizing total weighted flow

Malleable tasks are jobs that can be scheduled with preemptions on a varying number of resources. In this work, we have focused on the special case of work-preserving malleable tasks, for which the area of the allocated resources does not depend on the allocation and is equal to the sequential processing time. Moreover, we have assumed that the number of resources allocated to each task at each time instant is bounded. Although this study concerns malleable task scheduling, we have shown that this is equivalent to the problem of minimizing the makespan of independent tasks distributed among processors, when the data corresponding to tasks is sent using network flows sharing the same bandwidth.

We have considered both the clairvoyant and non-clairvoyant cases, and we have focused on minimizing the weighted sum of completion times. In the weighted non-clairvoyant case, we have proposed an approximation algorithm whose ratio is the same as in the unweighted non-clairvoyant case. In the clairvoyant case, we have provided a normal form for the schedule of such malleable tasks, and proved that any valid schedule can be turned into this normal form, based only on the completion times of the tasks. We have shown that in these normal form schedules, the number of preemptions per task is bounded by 3 on average. At last, we have analyzed the performance of greedy schedules, and proved that optimal schedules are greedy for a special case of homogeneous instances. We conjecture that there exists an optimal greedy schedule for all instances, which would greatly simplify the study of this problem.

6.1.15. Parallelizing the construction of the ProDom database

ProDom is a protein domain family database automatically built from a comprehensive analysis of all known protein sequences. ProDom development is headed by Daniel Kahn (INRIA project-team BAMBOO, formerly HELIX). With the protein sequence databases increasing in size at an exponential pace, the parallelization of MkDom2, the algorithm used to build ProDom, has become mandatory (the original sequential version of MkDom2 took 15 months to build the 2006 version of ProDom).

When protein domain families and protein families are built independently, the result may be inconsistent. In order to solve this inconsistency problem, we designed a new algorithm, MPI_MkDom3, that simultaneously builds a clustering in protein domain families and one in protein families. This algorithm mixes the principles of MP_MkDom2 and that of the building of Hogenom. As a proof of concept, we successfully processed all the sequences included in the April 2010 version of the UniProt database, namely 6 118 869 sequences and 2 194 382 846 amino-acids.

6.2. Algorithms and Software Architectures for Service Oriented Platforms

Participants: Daniel Balouek, Nicolas Bard, Julien Bigot, Yves Caniou, Eddy Caron, Florent Chuffart, Simon Delamare, Frédéric Desprez, Gilles Fedak, Sylvain Gault, Haiwu He, Cristian Klein, Georges Markomanolis, Adrian Muresan, Christian Pérez, Vincent Pichon, Jonathan Rouzaud-Cornabas, Anthony Simonet, José Saray, Bing Tang.
6.2.1. Parallel constraint-based local search

Constraint Programming emerged in the late 1980’s as a successful paradigm to tackle complex combinatorial problems in a declarative manner. It is somehow at the crossroads of combinatorial optimization, constraint satisfaction problems (CSP), declarative programming language and SAT problems (boolean constraint solvers and verification tools). Up to now, the only parallel method to solve optimization problems being deployed at large scale is the classical branch and bound, because it does not require much information to be communicated between parallel processes (basically: the current bound).

Adaptive Search was proposed by [86], [87] as a generic, domain-independent constraint-based local search method. This meta-heuristic takes advantage of the structure of the problem in terms of constraints and variables and can guide the search more precisely than a single global cost function to optimize, such as for instance the number of violated constraints. A parallelization of this algorithm based on threads realized on IBM BladeCenter with 16 Cell/BE cores show nearly ideal linear speed-ups for a variety of classical CSP benchmarks (magic squares, all-interval series, perfect square packing, etc.).

We parallelized the algorithm using the multi-start approach and realized experiments on the HA8000 machine, an Hitachi supercomputer with a maximum of nearly 16000 cores installed at University of Tokyo, and on the Grid’5000 infrastructure, the French national Grid for the research, which contains 8612 cores deployed on 11 sites distributed in France. Results show that speedups may surprisingly be architecture and problem dependant. Work in progress considers communications between each computing resource, and a new problem (costa) has been tested for its capability to have an exponential distribution of its time to complete on a sequential resolution.

6.2.2. Service Discovery in Peer-to-Peer environments

In 2010 we experimentally validated the scalability of the Spades Based Middleware (SBAM). SBAM is an auto-stabilized P2P middleware designed for the service discovery. The context of this development is the ANR SPADES project (see Section 7.2.2). In 2011, we wanted to guaranty truthfulness of information exchanged between SBAM-agents. In this context, the implementation of an efficient mechanism ensuring quality of large scale service discovery became a challenge. In collaboration with LIP6 team we developed a self stabilized model called CoPIF and we implemented it in SBAM using synchronous message exchange between agents. Indeed, when a node has to read its neighbor states, it sends a message to each and wait all response. Despite the fact that this kind of implementation is expensive, especially on a large distributed data structure, experiment shown that our model implementation stay efficient, even on a huge prefix tree. We use this broadcast mechanism not only to check the truthfulness of the distributed data structure but also to propagate activation of services on the entire SPADES platform. For the end of 2011 and the beginning of 2012 we plan to work on experimental evaluation of a self-stabilization inspired fault tolerance mechanism. We do this through a collaboration with Myriads team at Rennes.

Moreover, in the occasion of demonstration session of IEEE P2P’2011, we introduced the feasibility of multisite resources aggregation, thanks to SBAM, we ran SBAM on up to 200 peers (we generated machine volatility in order to show the self-stabilization) on 50 physical nodes of Grid’5000 to demonstrate the scalability of multi sites, self-stabilization good performance of our P2P middleware SBAM.

6.2.3. Décryptyon

In 2011, The DIET WebBoard (a web interface to manage the Décryptyon Grid through the DIET middleware) only received bugfixes and a few new features: the possibility to use a totally customized command to call the DIET client, improved support for multiprocessor tasks, and a basic support for replication of tasks (possibility to launch “clones” of an important task, in order to increase the probability of having a successful result). We deployed the new versions of the DIET Webboard on the Décryptyon university grid whenever we made changes to it.

In 2011, we started to port the Rhénovia application (a neuron simulation program in Java and python) on the Décryptyon grid.
The “Help cure muscular dystrophy, phase 2” program that we submitted to the world community grid was still in progress, we received large amounts of result files every day. We had to do the sorting of these files, checking, compressing and moving them to a long term storage space on a regular basis. We also made statistics for the internet users: http://graal.ens-lyon.fr/~nbard/WCGStats/. The last update was on 2011 June 27th: 76.67%.

6.2.4. Scheduling Applications with a Complex Structure

Non-predictably evolving applications are applications that change their resource requirements during execution. These applications exist, for example, as a result of using adaptive numeric methods, such as adaptive mesh refinement and adaptive particle methods. Increasing interest is being shown to have such applications acquire resources on the fly. However, current HPC Resource Management Systems (RMSs) only allow a static allocation of resources, which cannot be changed after it started. Therefore, non-predictably evolving applications cannot make efficient use of HPC resources, being forced to make an allocation based on their maximum expected requirements.

In 2011, we have revisited CooRM, an RMS targeting moldable application, and extended it to CooRMv2, an RMS which supports efficient scheduling of non-predictably evolving applications. An application can make “pre-allocations” to specify its peak resource usage. The application can then dynamically allocate resources as long as the pre-allocation is not outgrown. Resources which are pre-allocated but not used, can be filled by other applications. Results show that the approach is feasible and leads to a more efficient resource usage while guaranteeing that resource allocations are always satisfied.

As future work, we plan to extend CooRMv2 for non-homogeneous clusters, for example, for supercomputers that feature a non-homogeneous network. Moreover, we would like to apply the concepts proposed by CooRMv2 to large scale resource managers such as XtreamOS.

6.2.5. High Level Component Model

Most software component models focus on the reuse of existing pieces of code called primitive components. There are however many other elements that can be reused in component-based applications. Partial assemblies of components, well defined interactions between components and existing composition patterns (a.k.a. software skeletons) are examples of such reusable elements. It turns out that such elements of reuse are important for parallel and distributed applications. Therefore, we have designed High Level Component Model (HLCM), a software component model that supports the reuse of these elements thanks to the concepts of hierarchy, genericity and connectors—and in particular the novel concepts of open connection.

In 2011, we have developed two specific implementations of HLCM: L2C for for C++, MPI and CORBA based applications and GLUON++ for CHARM++ based applications in collaboration with Prof. Kale’s team at the University of Illinois at Urbana-Champaign. L2C was used to study how HLCM may simplify the development of domain decomposition applications. GLUON++ was in particular used to study the performance portability of FFT library on various kind of machines. Moreover, on going work includes the study of the benefit of HLCM for MapReduce applications.

6.2.6. Simplifying Code-Coupling in the SALOME platform

The SALOME platform is a generic platform for pre- and post-processing for numerical simulations. It is made of modules which are themselves a set of components. YACS is the module responsible for coupling applications, based on spatial and temporal relationships. The coupling of domain decomposition code, such as the coupling of several instances of Code_Aster, a thermomechanical calculation code from EDF R&D, turns out to be a complex task because of the lack of abstraction of current SALOME model.

In 2011, we have proposed and implemented some extensions to the SALOME model and platform to remove this limitation. The main extension is the ability to express the cloning of a service, which generates also the cloning of connections. The actual semantic of the cloning operation has been specified in function of the nature of the service (sequential, parallel) and of the ports (data or control flow). It has greatly simplified the expression of the coupling of several instances of Code_Aster without generating any measurable overhead at runtime: no more recompilation is needed when varying the number of coupled instances.
6.2.7. Towards Data Desktop Grid

Desktop Grids use the computing, network and storage resources from idle desktop PC’s distributed over multiple-LAN’s or the Internet to compute a large variety of resource-demanding distributed applications. While these applications need to access, compute, store and circulate large volumes of data, little attention has been paid to data management in such large-scale, dynamic, heterogeneous, volatile and highly distributed Grids. In most cases, data management relies on ad-hoc solutions, and providing a general approach is still a challenging issue.

We have proposed the BITDEW framework which addresses the issue of how to design a programmable environment for automatic and transparent data management on computational 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.

Since July 2010, in collaboration with the University of Sfax, we are developing 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.

In 2011, we have surveyed P2P strategies (replication, erasure code, replica repair, hybrid storage), which provides reliable and durable storage on top of hybrid distributed infrastructures composed of volatile and stable storage. Following this 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.

6.2.8. MapReduce programming model for Desktop Grid

MapReduce is an emerging programming model for data-intense application 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 have 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 shows that we achieve linear speedup on the classical WordCount benchmark. Several scenarios involving lagger hosts and host crashes demonstrate that the prototype is able to cope with an experimental context similar to real-world Internet.

In collaboration with the Huazhong University of Science & Technology, we have developed an emulation framework to assess MapReduce on Internet Desktop Grid. We have made extensive comparison on BitDew-MapReduce and Hadoop using Grid5000 which show that our approach has all the properties desirable to cope with an Internet deployment, whereas Hadoop fails on several tests.

In collaboration with the Babes-Bolyai University of Cluj-Napoca, we have proposed a distributed result checker based on the Majority Voting approach. We evaluated the efficiency of our algorithm by computing the aggregated probability with which a MapReduce computation produces an erroneous result.

We have published two chapters in collective books around Cloud and Desktop Grid technologies. The first one, in collaboration with University of Madrid is an introduction to MapReduce and Hadoop, the second one, in collaboration with Virginia Tech is a presentation of two alternative implementations of MapReduce for Desktop Grids : Moon and Bitdew.

6.2.9. 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 EGEE, 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 EGEE applications (e.g. the fusion community’s applications) that can tolerate such a long latency most of the user communities want much smaller latencies.

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. At this stage, the prototype is fully developed and validated. It supports the XtremWeb and BOINC Desktop Grid and OpenNebula, StratusLab, OpenStack and Amazon EC2 Clouds. The first versions have been delivered to the EDGI production infrastructure. 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.

6.2.10. Performance evaluation and modeling

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 computes 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 are working on further improvements for solving some performance issues with the rest benchmarks. We plan to apply some new techniques about the instrumentation of the benchmarks which we have already discussed with people from the performance analysis community and also improve the trace replay tool in order to improve its accuracy. 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.
6.2.11. Elastic Scheduling for Functional Workflows

Non-DAG (or functional) workflows are sets of task-graph workflows with non-deterministic transitions between them, that are determined at runtime by special nodes that control the execution flow. In a current work we are focusing on formalizing and evaluating an allocation and scheduling strategy for on-line non-DAG workflows. The goal of this work is to target real-world non-DAG applications and use cloud platforms to perform elastic allocations while keeping cost and stretch fairness constraints.

To address the previous problem we consider each non-DAG workflow as a set of DAG sub-workflows with non-deterministic transitions between them. Whenever an event occurs (a sub-workflow’s execution is completed, a new workflow arrives in the system, a workflow is canceled, etc.) we need to do a rescheduling. The rescheduling strategy considers the currently-running tasks as fixed. Given that the number of events increases proportional to the number of workflows in the system, there is the risk of spending too much time on the scheduling problem and not enough on the workflows themselves. As a result, the scheduling strategy that we will adopt will be a computational inexpensive one, which will give us more room for the number of possible workflows in the system.

This work is currently in the validation step through experimentation with synthetic data. In the near future we will validate against traces of real-world applications that use non-DAG workflows.

6.2.12. Self Adaptive Middleware Deployment

A computer application can be considered as a system of components that exchange information. Each component type has its specific constraints. The application, as a whole, has also its constraints. Deploying an application on a distributed system consist, among other things, to make a mapping between application components and system resources to meet each component constraints, the application constraints, and possibly those set by the the user. Previous work exists on the deployment of middleware, including DIET (with two finished PhD). However, few take into account the issue of redeployment in the event of variation (availability, load, number) of resources. We study this problem of self adaptive deployment of middleware. It consist of achieving an initial deployment, then scrutinizing some changes in the environment, and automatically adjust the deployment (if beneficial) in case of detecting a variation that degrades the performance expected. To do this, we have surveyed the fields of autonomic computing, self adaptive systems and we have defined the different problems that must be solved to achieve this goal. From this, we first define a resource model to represent the physical system, we are to define a model of middleware-based software components, have started the implementation of the resource model to achieve a simulator.

6.2.13. Virtual Machine Placement with Security Requirements

With the number of services using virtualization and clouds growing faster and faster, it is common to mutualize thousands of virtual machines (VMs) within one distributed system. Consequently, the virtualized services, pieces of software and hardware, and infrastructures share the same physical resources. This has given rise to important challenges regarding the security of VMs and the importance of enforcing non-interference between them. Indeed, cross-VM attacks are an important and real world threat. The problem is even worse in the case of adversary users hosted on the same hardware (multi-tenance). Therefore, the isolation facility within clouds needs to be strong. Furthermore, each user has different adversaries and the placement and scheduling processes need to take these adversaries into account.

First, we have worked on resource model to describe distributed system and application model to describe the composition of virtual machine. Then we have formalize isolation requirements between users, between applications and between virtual machines. We also formalized the redundancy requirement. We have created a simulator that can load our resource model and application model. Using it, we have described the Grid’5000 infrastructure and a Virtual Cluster application. We have formalized and implemented an algorithm that takes into account the requirements and place the application. Work in progress considers using Constraint Satisfaction Problems (CSP) and SAT problems to improve the quality of placement. Moreover, we study the trade-off between performance, security requirements and infrastructure consolidation. This works is part of a project on Cloud Security with Alcatel-Lucent Bell Labs and ENSI de Bourges.
6.2.14. Scheduling for MapReduce Based Applications

After a study of the state of the art regarding scheduling, especially scheduling on grid and clouds and MapReduce application scheduling, experiments were performed over the Grid’5000 and Google/IBM Hadoop platforms. We are now working on improving a previous work by Berlinska and Drozdowski which aims at providing a good static schedule of the Map and Reduce phases. A visualization tool has been developed which draws Gantt charts resulting from Berlinska and Drozdowski’s algorithms as well as from our own scheduling heuristics.

A BlobSeer model is also developed in collaboration with the Kerdata research team that will be used for our next developments.

6.3. Parallel Sparse Direct Solvers and Combinatorial Scientific Computing

Participants: Maurice Brémond, Guillaume Joslin, Johannes Langguth, Jean-Yves L’Excellent, Mohamed Sid-Lakhdar, Bora Uçar.

6.3.1. Parallel computation of entries of the inverse of a sparse matrix

Following last year’s work on computing entries of the inverse of a sparse matrix in a serial, in-core or out-of-core environment, and that was implemented in MUMPS, we have pursued work to address this issue in a parallel environment. In such this case, it has been shown that minimizing the number of operations (or the number of accesses to the factors) and balancing the work between the processors are contradictory objectives. Several ideas have been investigated and implemented in order to deal with this issue and to reach high speed-ups. Experimental results are promising and show good speed-ups on relatively small number of processors (up to 16) when dealing with large blocks of sparse right-hand sides, while we used to experience speed-downs before.

6.3.2. Multithreaded parallelism for the MUMPS solver

Apart from using message-passing, we have in the past only exploited multicore parallelism through threaded libraries (e.g. BLAS: Basic Linear Algebra Subroutines), and a few OpenMP directives. We are currently investigating the combination of this fork-join model with threaded parallelism resulting from the task graph, which, in our context, is a tree. To do so, and in order to also target NUMA architectures, we apply ideas from distributed-memory environments to multithreaded environments. Simulations based on benchmarks followed by a first prototype implementation have validated this approach for some classes of matrices on small numbers of cores. We are currently revisiting this implementation and plan to pursue experiments on larger numbers of cores with larger classes of matrices. This starting work was done in the context of a master thesis and is the object of a starting PhD thesis. In a distributed-memory environments, it will be combined with parallelism based on message passing, where the scalability of the existing communication schemes should also be addressed. Both directions will be followed in order to face the multicore (r)evolution.

6.3.3. Low-rank approximations

Low-rank approximations are commonly used to compress the representation of data structures. The loss of information induced is often negligible and can be controlled. Although the dense internal datastructures involved in a multifrontal method, the so-called frontal matrices or fronts, are full-rank, they can be represented by a set of low-rank matrices. Applying to our context the notion of geometric clustering used by Bebendorf to define hierarchical matrices, we have shown that the efficiency of this representation to reduce the complexity of both the factorization and solve phases strongly depends on how variables are grouped. The proposed approach can be used either to accelerate the factorization and solution phases or to build a preconditioner. The ultimate goal of this work is to extend the features of the MUMPS solver to exploit low-rank properties.

This work, and the work described in the two previous paragraphs are in the context of a collaboration with ENSEEIHT-IRIT and with the partners involved in the MUMPS project (see Section 5.2 ).
6.3.4. On partitioning problems with complex objectives

Hypergraph and graph partitioning tools are used to partition work for efficient parallelization of many sparse matrix computations. Most of the time, the objective function that is reduced by these tools relates to reducing the communication requirements, and the balancing constraints satisfied by these tools relate to balancing the work or memory requirements. Sometimes, the objective sought for having balance is a complex function of a partition. We mention some important class of parallel sparse matrix computations that have such balance objectives. For these cases, the current state of the art partitioning tools fall short of being adequate. To the best of our knowledge, there is only a single algorithmic framework in the literature to address such balance objectives. We propose another algorithmic framework to tackle complex objectives and experimentally investigate the proposed framework.

6.3.5. On the Use of Cluster-Based Partial Message Logging to Improve Fault Tolerance for MPI HPC Applications

Fault tolerance is becoming a major concern in HPC systems. The two traditional approaches for message passing applications, coordinated checkpointing and message logging, have severe scalability issues. Coordinated checkpointing protocols make all processes roll back after a failure. Message logging protocols log a huge amount of data and can induce an overhead on communication performance. Hierarchical rollback-recovery protocols based on the combination of coordinated checkpointing and message logging are an alternative. These partial message logging protocols are based on process clustering: only messages between clusters are logged to limit the consequence of a failure to one cluster. These protocols would work efficiently only if one can find clusters of processes in the applications such that the ratio of logged messages is very low. We study the communication patterns of message passing HPC applications to show that partial message logging is suitable in most cases. We propose a partitioning algorithm to find suitable clusters of processes given the communication pattern of an application. Finally, we evaluate the efficiency of partial message logging using two state of the art protocols on a set of representative applications.

6.3.6. Integrated data placement and task assignment for scientific workflows in clouds

We consider the problem of optimizing the execution of data-intensive scientific workflows in the Cloud. We address the problem under the following scenario. The tasks of the workflows communicate through files; the output of a task is used by another task as an input file and if these tasks are assigned on different execution sites, a file transfer is necessary. The output files are to be stored at a site. Each execution site is to be assigned a certain percentage of the files and tasks. These percentages, called target weights, are pre-determined and reflect either user preferences or the storage capacity and computing power of the sites. The aim is to place the data files into and assign the tasks to the execution sites so as to reduce the cost associated with the file transfers, while complying with the target weights. To do this, we model the workflow as a hypergraph and with a hypergraph-partitioning-based formulation, we propose a heuristic which generates data placement and task assignment schemes simultaneously. We report simulation results on a number of real-life and synthetically generated scientific workflows. Our results show that the proposed heuristic is fast, and can find mappings and assignments which reduce file transfers, while respecting the target weights.

6.3.7. UMPa: A Multi-objective, multi-level partitioner for communication minimization

We propose a directed hypergraph model and a refinement heuristic to distribute communicating tasks among the processing units in a distributed memory setting. The aim is to achieve load balance and minimize the maximum data sent by a processing unit. We also take two other communication metrics into account with a tie-breaking scheme. With this approach, task distributions causing an excessive use of network or a bottleneck processor which participates to almost all of the communication are avoided. We show on a large number of problem instances that our model improves the maximum data sent by a processor up to 34% for parallel environments with 4, 16, 64 and 256 processing units compared to the state of the art which only minimizes the total communication volume.
6.3.8. A Divisive clustering technique for maximizing the modularity

We present a new graph clustering algorithm aimed at obtaining clusterings of high modularity. The algorithm pursues a divisive clustering approach and uses established graph partitioning algorithms and techniques to compute recursive bipartitions of the input as well as to refine clusters. Experimental evaluation shows that the modularity scores obtained compare favorably to many previous approaches. In the majority of test cases, the algorithm outperformed the best known alternatives. In particular, among 13 problem instances common in the literature, the proposed algorithm improves the best known modularity in 9 cases.

6.3.9. 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 [Eisenstat and Liu, SIAM J. Matrix Anal. Appl., 26 (2005) and 29 (2008)]. The construction algorithm has a worst case time complexity $O(mn)$ for an $n \times n$ unsymmetric matrix having $m$ nonzeros. We propose another algorithm that has a worst case time complexity of $O(m \log n)$.

6.3.10. Multithreaded clustering for multi-level hypergraph partitioning

Requirements for efficient parallelization of many complex and irregular applications can be cast as a hypergraph partitioning problem. The current-state-of-the-art software libraries that provide tool support for the hypergraph partitioning problem are designed and implemented before the game-changing advancements in multi-core computing. Hence, analyzing the structure of those tools for designing multithreaded versions of the algorithms is a crucial task. The most successful partitioning tools are based on the multi-level approach. In this approach, a given hypergraph is coarsened to a much smaller one, a partition is obtained on the the smallest hypergraph, and that partition is projected to the original hypergraph while refining it on the intermediate hypergraphs. The coarsening operation corresponds to clustering the vertices of a hypergraph and is the most time consuming task in a multi-level partitioning tool. We present three efficient multithreaded clustering algorithms which are very suited for multi-level partitioners. We compare their performance with that of the ones currently used in today’s hypergraph partitioners. We show on a large number of real life hypergraphs that our implementations, integrated into a commonly used partitioning library PaToH, achieve good speedups without reducing the clustering quality.

6.3.11. Partitioning, ordering, and load balancing in a hierarchically parallel hybrid linear solver

PDSLin is a general-purpose algebraic parallel hybrid (direct/iterative) linear solver based on the Schur complement method. The most challenging step of the solver is the computation of a preconditioner based on an approximate global Schur complement. We investigate two combinatorial problems to enhance PDSLin’s performance at this step. The first is a multi-constraint partitioning problem to balance the workload while computing the preconditioner in parallel. For this, we describe and evaluate a number of graph and hypergraph partitioning algorithms to satisfy our particular objective and constraints. The second problem is to reorder the sparse right-hand side vectors to improve the data access locality during the parallel solution of a sparse triangular system with multiple right-hand sides. This is needed to eliminate the unknowns associated with the interface in PDSLin. We study two reordering techniques: one based on a postordering of the elimination tree and the other based on a hypergraph partitioning. To demonstrate the effect of these techniques on the performance of PDSLin, we present the numerical results of solving large-scale linear systems arising from numerical simulations of modeling accelerator cavities and of modeling fusion devices.

6.3.12. Experiments on push-relabel-based maximum cardinality matching algorithms for bipartite graphs

We report on careful implementations of several push-relabel-based algorithms for solving the problem of finding a maximum cardinality matching in a bipartite graph and compare them with fast augmenting-path-based algorithms. We analyze the algorithms using a common base for all implementations and compare their relative performance and stability on a wide range of graphs. The effect of a set of known initialization heuristics on the performance of matching algorithms is also investigated. Our results identify a variant of
the push-relabel algorithm and a variant of the augmenting-path-based algorithm as the fastest with proper initialization heuristics, while the push-relabel based one having a better worst case performance.

6.3.13. Towards a scalable hybrid linear solver based on combinatorial algorithms

The availability of large-scale computing platforms comprised of tens of thousands of multicore processors motivates the need for the next generation of highly scalable sparse linear system solvers. These solvers must optimize parallel performance, processor (serial) performance, as well as memory requirements, while being robust across broad classes of applications and systems. In this study, we present a hybrid parallel solver that combines the desirable characteristics of direct methods (robustness) and effective iterative solvers (low computational cost), while alleviating their drawbacks (memory requirements, lack of robustness). We discuss several combinatorial problems that arise in the design of this hybrid solver, present algorithms to solve these combinatorial problems, and demonstrate their impact on a large-scale three-dimensional PDE-constrained optimization problem.
6. New Results

6.1. Perfect simulation

We have proposed a new approach for sampling the stationary distribution of general Markov chains that only needs to consider two trajectories. We show that this new approach is particularly effective when the state space can be partitioned into pieces where envelopes can be easily computed [26]. We further show that most Markovian queuing networks have this property and we propose efficient algorithms for some of them, in particular when the rates of events range over several orders of magnitude [45]. We also provided a novel approach for efficient sampling of queues with phase type servers [37] (this paper has received the best paper award at ASMTA 2011) and Markov chains with infinite state spaces (but with a known bounding process). Perfect sampling has been used for model checking of probabilistic models in [14].

6.2. Economic models for clouds

Recently introduced spot instances in the Amazon Elastic Compute Cloud (EC2) offer low resource costs in exchange for reduced reliability; these instances can be revoked abruptly due to price and demand fluctuations. Mechanisms and tools that deal with the cost-reliability trade-offs under this scheme are of great value for users seeking to lessen their costs while maintaining high reliability. We study how mechanisms, namely, checkpointing and migration, can be used to minimize the cost and volatility of resource provisioning. Based on the real price history of EC2 spot instances, we compare several adaptive checkpointing schemes in terms of monetary costs and improvement of job completion times. We evaluate schemes that apply predictive methods for spot prices. Furthermore, we also study how work migration can improve task completion in the midst of failures while maintaining low monetary costs. Trace-based simulations show that our schemes can reduce significantly both monetary costs and task completion times of computation on spot instance [25].

6.3. Game theory and networks

We studied the traffic routing problem in networks whose users try to minimize their latencies by employing a distributed learning rule inspired by the replicator dynamics of evolutionary game theory. The stable states of these dynamics coincide with the network’s (Wardrop) equilibrium points. Despite this abundance of stable states, we find that (almost) every solution trajectory converges to an equilibrium point at an exponential rate. When network latencies fluctuate unpredictably we show that the time-average of the traffic flows of sufficiently patient users is still concentrated in a neighborhood of evolutionarily stable equilibria and we estimate the corresponding stationary distribution and convergence times [42].

We also analyzed the distributed power allocation problem in parallel multiple access channels (MAC) by studying an associated non-cooperative game which admits an exact potential function. We show that the parallel MAC game admits a unique equilibrium almost surely. Furthermore, if the network’s users employ a distributed learning scheme based on the replicator dynamics, we show that they converge to equilibrium from almost any initial condition, even though users only have local information at their disposal [41].

Using a large deviations approach we calculate the probability distribution of the mutual information of MIMO channels in the limit of large antenna numbers. We calculate the full distribution, including its tails which strongly deviate from the Gaussian behavior near the mean. This calculation provides us with a tool to obtain outage probabilities analytically at any point in the parameter space, as long as the number of antennas is not too small [20].
6.4. Mean field analysis for networks

We have studied the deterministic limits of Markov processes made of several interacting objects. While most classical results assume that the limiting dynamics has Lipschitz properties, we show that these conditions are not necessary to prove convergence to a deterministic system.

We show that under mild assumptions, the stochastic system converges to the set of solutions of a differential inclusion and we provide simple way to compute the limiting inclusion. When this differential inclusion satisfies a one-sided Lipschitz condition, there exists a unique solution of this differential inclusion and we show convergence in probability with explicit bounds.

This extends the applicability of mean field techniques to systems exhibiting threshold dynamics such as queuing systems with boundary conditions or controlled dynamics. This is illustrated by applying our results to several types of systems: fluid limits of priority queues, best response dynamics in games, push-pull queues with a large number of sources and a large number of servers and self-adapting computing systems [65].

6.5. Idleness and failure prediction in large infrastructures

We have proposed a method to discover statistical models of availability in large distributed systems and applied it to run an enlightening study of SETI@home [19]. This was also used to make long-term availability predictions for groups of desktop grid resources [21]. We have used statistically based models of heterogeneous failures in parallel systems and assessed their tolerance [39]. A similar approach was used to design correlated resource models of Internet end hosts [38], [17].

6.6. Scheduling and Game Theory

A stochastic model of failures has been used to optimize the scheduling of checkpoints on desktop grids [28].

We have also shown that non-cooperative scheduling can be considered harmful in collaborative volunteer computing environments [33].

Optimal scheduling and route selection have been investigated using a novel approach based on Lagrangian optimization. This result is inspired from flow control in multi-path networks and was used for multiple mag-of-tasks application scheduling on grids [61].

In the similar context of broker-based networks of non-observable parallel queues, we provide lower bounds on the minimum response time. We introduce the “Price of Forgetting” (PoF), the ratio between the minimum response times achieved by a probabilistic broker and a broker with memory, that is shown to be unbounded or arbitrarily close to one depending on the coefficient of variation of the service time distributions. We also put our results in the context of game theory revisiting the “Price of Anarchy” (PoA) of parallel queues: It can be decomposed into the product of the PoA achieved by a probabilistic broker (already well understood) and the PoF [10].

6.7. Validity study of flow-based network models.

Researchers in the area of distributed computing conduct many of their experiments in simulation. While packet-level simulation is often used to study network protocols, it can be too costly to simulate network communications for large-scale systems and applications. The alternative chosen in SimGrid and a few other simulation frameworks is to simulate the network based on less costly flow-level models. Surprisingly, in the literature, validation of these flow-level models is at best a mere verification for a few simple cases. Consequently, although distributed computing simulators are widely used, their ability to produce scientifically meaningful results is in doubt. In [9], [70] we focus on the validation of state-of-the-art flow-level network models of TCP communication, via comparison to packet-level simulation. While it is straightforward to show cases in which previously proposed models lead to good results, instead we systematically seek cases that lead to invalid results. Careful analysis of these cases reveal fundamental flaws and also suggest improvements. One
contribution of this work is that these improvements lead to a new model that, while far from being perfect, improves upon all previously proposed models. A more important contribution, perhaps, is provided by the pitfalls and unexpected behaviors encountered in this work, leading to a number of enlightening lessons. In particular, this work shows that model validation cannot be achieved solely by exhibiting (possibly many) “good cases.” Confidence in the quality of a model can only be strengthened through an invalidation approach that attempts to prove the model wrong.

6.8. Vizualisation

We have proposed a methodology for detecting resource usage anomalies in large scale distributed systems. The methodology relies on four functionalities: characterized trace collection, multi-scale data aggregation, specifically tailored user interaction techniques, and visualization techniques. We have shown the efficiency of this approach through the analysis of simulations of the volunteer computing Berkeley Open Infrastructure for Network Computing architecture (BOINC). Three scenarios have been analyzed in [48], [23]: analysis of the resource sharing mechanism, resource usage considering response time instead of throughput, and the evaluation of input file size on Berkeley Open Infrastructure for Network Computing architecture. The results show that our methodology enables to easily identify resource usage anomalies, such as unfair resource sharing, contention, moving network bottlenecks, and harmful short-term resource sharing. Triva, the resulting software, has been demonstrated at the SuperComputing conference.

We also have investigated how to use trace-based visualization to understand applications I/O performance [49] and how to visually compare two traces [70] and highlight differences.

6.9. Experimental methodology

In the scientific experimentation process, an experiment result needs to be analyzed and compared with several others, potentially obtained in different conditions. Several tools are dedicated to the control of the experiment input parameters and the experiment replay. In parallel, concurrent and distributed systems, experiment conditions are not only restricted to the input parameters, but also to the software environment in which the experiment was carried out. It is therefore essential to be able to reconstruct this type of environment. This can quickly become complex for experimenters, particularly on research platforms dedicated to scientific experimentation, where both hardware and software are in constant rapid evolution. We study the concept of the reconstructability of software environments and propose a tool for dealing with this problem in [64].

We have also started investigating the systematic use of Design of Experiments to computer studies (see [61]). Nonetheless such approach provides results that are much more trustworthy than what is generally done in the parallel and distributed computing community but it also enables to shorten the experiments cycle and to use less computing resources.

6.10. Multi-core platforms

We have used memory access traces to map threads on hierarchical multi-core platforms [13]. We have also used software transactional memory to analyze and trace applications running on multi-core architectures [30].

An approach based on machine learning was used to map threads on transactional memory applications in [31].

The impact of CPU and memory affinity on multi-core platforms was investigated in [46] using numerical scientific multi-threaded applications as a typical case study. This resulted in improvement of the performance of parallel systems using a NUMA-aware load balancer [68].

We have also carried a performance evaluation of WiNoCs for parallel workloads based on collective communications [43] as well as for Infiniband networks [40].
6.11. High performance computing

We have developed a runtime system, named SGPU 2, that enable large applications to run on clusters of hybrid nodes [44].

BigDFT is a parallel simulator of the matter at the nano scale. It uses Daubechies Wavelets for High Performance Electronic Structure Calculations [16]. This tool is shown to make efficient use of massive parallel hybrid architectures [57].

6.12. Input-Output

Atmospheric models usually demand high processing power and generate large amounts of data. As the degree of parallelism grows, the I/O operations may become the major impacting factor of their performance. In [27], we evaluate the Ocean-Land-Atmosphere Model (OLAM) on the PVFS file system in order to point the I/O characteristics of the application. We show that storing the files on PVFS has lower performance than using the local disks of the cluster nodes due to file creation and network concurrency. Additionally, we study the performance of a new version of OLAM that used MPI associated with OpenMP and show that the combined strategy presents I/O times 20 times shorter than the original MPI-only version and 9 times shorter on total execution time. Finally, a survey on I/O Characterization of several applications is given in [51].
6. New Results

6.1. Kaapi

New version of Kaapi, called X-Kaapi, has been released. The kernel is written in C for hypothetical required from embedded system. On top of the kernel, several APIs co-exist: a template based C++ library called Kaapi++; a C API; a Fortran API; and a compiler that transform a source code annotated with pragma directive to a source code with calls to the runtime library function. The compiler works with C and a subset of C++. 
http://kaapi.gforge.inria.fr

6.2. Multi-criteria optimization

The main idea is the development of a methodology to generate a reasonable set of approximated Pareto’ solutions (closed to the best achievable solutions). Especially, we have applied this methodology to better take into account users’ criteria than the other existing methods offer. We have also studied the problem of selection of best algorithms in a portfolio. This research axis is currently enforced by the INRIA postdoc position of Joachim Lepping where we have started to include a learning process to select the best algorithm on a given instance.

6.3. Stochastic models for optimizing checkpoint protocol

After our past studied on design of origin checkpoint protocols, we have proposed a new stochastic performance model of the parallel execution in presence of failures. Thanks to this formulation, we are able to optimize several criteria (the time lost due to failure; the expected completion time) by making right decision of the date of each checkpoint. The model is general and it does not take into account the failure distribution law and accept variable checkpoint time estimation, which is important for dynamic parallelism applications.

6.4. Work stealing scheduling algorithm taking care of communication

On some applications, the amount of data transfers can be high. To minimize the amount of data transfers during the execution, Jean-Noel Quintin has developed an algorithm called WSCOM which uses the DAG structure of the application. For each steal request, the work-stealing algorithm tries to balance the load between the thief and the stolen processor. Thus, WSCOM tries to divide the work on the stolen processor into two parts with a small number of edges between the two parts. This cutting is done with a negligible overhead at each steal request. This algorithm has been implemented in a tool called DSMake. This tool executes the set of tasks described by a Makefile on a distributed platform. In addition, I have developed a simulator to validate algorithm performance and its behavior. We compared WSCOM and several static list-scheduling algorithms. The comparison shows that WSCOM outperforms list-scheduling algorithms, on clusters with some network congestion.

Besides, based on SIPS analysis of work stealing, Stefano Mor in his thesis compared the influence of the choice of the stolen tasks on the number of steal operations, distinguishing unsuccessful and successful steals. While standard bounds are related to unsuccessful steals, they are pessimistic with respect to the number of successful steals that define intensive data communications.
6.5. Homomorphic coding for soft error resilience

We extended our results for fault-tolerant modular computations in two directions. To improve the correction rate of Reed-Solomon codes, power-decoding techniques consist in augmenting the number of syndrome equations by raising the received word to successive powers. The correction is done by a generalization of Berlekamp-Massey algorithm acting on multiple sequences. This method is, if not equivalent, at least very close to the list-decoding proposed by Sudan in its first version, in particular, error correction rates are identical. We improve the power-decoding method by reformulation into a vector rational function reconstruction, with benefit from fast polynomial matrix arithmetic. Besides, for basic exact linear algebra computations (eg dense linear system), we designed interactive protocols between a trusted platform and a non trusted one for resilience to soft-errors.

6.6. Chimeric algorithms design

To reach provable multicriteria performance, we used the coupling of various algorithms that adapt in several contexts: recursive cascading of both sequential and parallel algorithms with work-stealing; coupling specific algorithms on heterogeneous platforms (eg CPU/GPU); interactive distributed computations; fault-tolerant computations by coupling both a trustfully platform with low computation bandwidth and an unreliable computing platform with high bandwith. A unification work is currently developed for the design of a chimeric algorithms that is composed of the parts of multiple algorithms, interactively cascaded to achieve provable multicriteria performance.
6. New Results

6.1. Towards Data-Centric Networking


- Disruption Tolerant Networking

We designed an efficient message delivery framework, called MeDeHa, which enables communication in an internet connecting heterogeneous networks that is prone to disruptions in connectivity [24]. MeDeHa is complementary to the IRTF’s Bundle Architecture: besides its ability to store messages for unavailable destinations, MeDeHa can bridge the connectivity gap between infrastructure-based and multi-hop infrastructure-less networks. It benefits from network heterogeneity (e.g., nodes supporting more than one network and nodes having diverse resources) to improve message delivery. For example, in IEEE 802.11 networks, participating nodes may use both infrastructure- and ad-hoc modes to deliver data to otherwise unavailable destinations. It also employs opportunistic routing to support nodes with episodic connectivity. One of MeDeHa’s key features is that any MeDeHa node can relay data to any destination and can act as a gateway to make two networks inter-operate or to connect to the backbone network. The network is able to store data destined to temporarily unavailable nodes till the time of their expiry. This time period depends upon current storage availability as well as quality-of-service needs (e.g., delivery delay bounds) imposed by the application. We showcase MeDeHa’s ability to operate in environments consisting of a diverse set of interconnected networks and evaluate its performance through extensive simulations using a variety of scenarios with realistic synthetic and real mobility traces. Our results show significant improvement in average delivery ratio and a significant decrease in average delivery delay in the face of episodic connectivity. We also demonstrate that MeDeHa supports different levels of quality-of-service through traffic differentiation and message prioritization.

Then, we have extended the MeDeHa framework to support multihop mobile ad-hoc networks (or MANETs). Integrating MANETs to infrastructure-based networks (wired or wireless) allows network coverage to be extended to regions where infrastructure deployment is sparse or nonexistent as well as a way to cope with intermittent connectivity. Indeed, to date there are no comprehensive solutions that integrate MANETs to infrastructure-based networks. We have proposed a message delivery framework that is able to bridge together infrastructure-based and infrastructure-less networks. Through extensive simulations, we have demonstrated the benefits of the extended MeDeHa architecture especially in terms of the extended coverage it provides as well as its ability to cope with arbitrarily long-lived connectivity disruptions. Another important contribution of this work is to deploy and evaluate our message delivery framework on a real network testbed as well as conduct experiments in “hybrid” scenarios running partly on simulation and partly on real nodes [32].

Finally, we have proposed a naming scheme for heterogeneous networks composed of infrastructure-based and infrastructure-less networks where nodes may be subject to intermittent connectivity. The proposed scheme, called Henna, aims at decoupling object identification from location and is designed to operate with status-quo Internet routing. We evaluated the proposed naming scheme using the ns-3 network simulator and demonstrated that nodes were able to receive messages in both infrastructure-based and infrastructure-less networks despite frequent disconnections and changing location identifiers (i.e., IP address), while visiting different networks [31].

Another important contribution of this work is to deploy and evaluate our message delivery framework on a real network testbed as well as conduct experiments in “hybrid” scenarios running partly on simulation and partly on real nodes. This was demonstrated at the ACM Sigcomm conference in Toronto on August 2011 [74].
These different works are the result of collaborations with Katia Obraczka and Marc Mendonca from University of California Santa Cruz (UCSC) in the context of the COMMUNITY Associated Team, see URL http://inrg.cse.ucsc.edu/community/.

Another activity in the same domain relates to efficient scheduling and drop policies in DTNs. We remind that Delay Tolerant Networks are 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. We are working on this extension within a new framework called MobiTrade, which provides a utility driven trading system for efficient content dissemination on top of a disruption tolerant network. While simple tit-for-tat (TFT) mechanisms can force nodes to give one to get one, dealing with the inherent tendency of peers to take much but give back little, they can quickly lead to deadlocks when some (or most) of interesting content must be somehow fetched across the network. To resolve this, MobiTrade proposes a trading mechanism that allows a node (merchant) to buy, store, and carry content for other nodes (its clients) so that it can later trade it for content it is personally interested in. To exploit this extra degree of freedom, MobiTrade nodes continuously profile the type of content requested and the collaboration level of encountered devices. An appropriate utility function is then used to collect an optimal inventory that maximizes the expected value of stored content for future encounters, matched to the observed mobility patterns, interest patterns, and collaboration levels of encountered nodes. Using ns-3 simulations based on synthetic and real mobility traces, we show that MobiTrade achieves up to 2 times higher query success rates compared to other content dissemination schemes. Furthermore, we show that MobiTrade successfully isolates selfish devices. For further details on MobiTrade, we refer to [41] and to the web page of the project 1 where the code can be downloaded for both the ns-3 simulator and Android devices.

- Naming and Routing in Content Centric Networks

1 http://planete.inria.fr/MobiTrade/
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 worked 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. This activity was started in October 2011 by Damien Saucez.

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

With the advent of broadcast/multicast systems (e.g., DVB-H/SH), 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 recent activities, in the context of the PhD of F. Mattoussi, included 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, while keeping backward compatibility (LDPC-Staircase "codewords" can be decoded with a GPLDPC-Staircase codec).

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. This is the first adoption of these codes in an international standard.

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 a recent Internet-Draft [66] defines 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 of 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, will be commercialized in 2012.
through an industrial partner. This codec proves that LDPC-Staircase codes can offer erasure recovery performances close to ideal codes in many circumstances while keeping decoding speeds over 1Gbps.

The fact that LDPC-Staircase codes have been preferred to a major AL-FEC competitor for the ISDB-Tmm standard, is the recognition of their intrinsic qualities and of an appropriate balance between several technical and non technical criteria.

- 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 want 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 [65] 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 [66], [65].

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 [67]. We have shown that the GOE approach, by the flexibility it offers, its simplicity, its backward compatibility and its good recovery capabilities (under finite of 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 RFC [68].

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: [59] [60] [61].

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, in [29]. 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 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. 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[56],[57],[58].

• Security of the Broadcast/Multicast Systems

We believe that sooner or later, broadcasting systems will require security services. This is all the more true as heterogeneous broadcasting technologies will be used, for instance hybrid satellite-based and terrestrial networks, some of them being by nature open, as wireless networks (e.g., wimax, wifi). Therefore, one of the key security services is the authentication of the packet origin, and the packet integrity check. A key point is the ability for the terminal to perform these checks easily (the terminal often has limited processing and energy capabilities), while being tolerant to packet losses.

The TESLA (Timed Efficient Stream Loss-tolerant Authentication) scheme fulfills these requirements. We are therefore standardizing the use of TESLA in the context of the ALC and NORM reliable multicast transport protocols, within the IETF MSEC working group. This document has been published as RFC 5776.

In parallel, 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 in [62],[63],[64]. This activity is also carried out within the IETF RMT working group.

• 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 has 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.

6.2. Network Security and Privacy

Participants: Sana Ben Hamida, Claude Castelluccia, Walid Dabbous, Mohamed Ali Kaafar, Arnaud Legout, Stevens Le Blond, Daniele Perito.

• Online users tracking and profiling techniques
Usernames are ubiquitously used for identification and authentication purposes on web services and the Internet at large, ranging from the local-part of email addresses to identifiers in social networks. Usernames are generally alphanumerical strings chosen by the users and, by design, are unique within the scope of a single organization or web service. In this work, we investigate the feasibility of using usernames to trace or link multiple profiles across services that belong to the same individual. The intuition is that the probability that two usernames refer to the same physical person strongly depends on the entropy of the username string itself. Our experiments, based on usernames gathered from real web services, show that a significant portion of the users’ profiles can be linked using their usernames. In collecting the data needed for our study, we also show that users tend to choose a small number of related usernames and use them across many services. This work is the first to consider usernames as a source of information when profiling users on the Internet. It has been published in PETS 2011 [47], one of the most prestigious conference in the area of Computer Privacy, and has been awarded the Andreas Pfitzmann award for the best contribution.

- **Online Privacy measurements and threats identification in online social networks**

In this work, 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 friends relationship 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. To the best of our knowledge, this is the first time that user interests are used for profiling, and more generally, semantics-driven inference of private data is addressed. This work has been published in the prestigious Network & Distributed System Security Symposium (NDSS) 2012 [37].

- **Privacy Enhancing Technologies**

The increasing amount of personal and sensitive information disseminated over the Internet prompts commensurately growing privacy concerns. Digital data often lingers indefinitely and users lose its control. This motivates the desire to restrict content availability to an expiration time set by the data owner. This work presents and formalizes the notion of Ephemeral Publishing (EphPub), to prevent the access to expired content. We propose an efficient and robust protocol that builds on the Domain Name System (DNS) and its caching mechanism. With EphPub, sensitive content is published encrypted and the key material is distributed, in a steganographic manner, to randomly selected and independent resolvers. The availability of content is then limited by the evanescence of DNS cache entries. The EphPub protocol is transparent to existing applications, and does not rely on trusted hardware, centralized servers, or user proactive actions. We analyze its robustness and show that it incurs a negligible overhead on the DNS infrastructure. We also perform a large-scale study of the caching behavior of 900K open DNS resolvers. Finally, we propose an Android application, Firefox and Thunderbird extensions that provide ephemeral publishing capabilities, as well as a command-line tool to create ephemeral files. This work has been published in ICNP 2011 [36].

- **Differentially private smart metering**
Several countries throughout the world are planning to deploy smart meters in households in the very near future. The main motivation, for governments and electricity suppliers, is to be able to match consumption with generation. Traditional electrical meters only measure total consumption on a given period of time (i.e., one month or one year). As such, they do not provide accurate information of when the energy was consumed. Smart meters, instead, monitor and report consumption in intervals of few minutes. They allow the utility provider to monitor, almost in realtime, consumption and possibly adjust generation and prices according to the demand. Although smart metering might help improving energy management, it creates many new privacy problems. Smart meters provide very accurate consumption data to electricity providers. As the interval of data collected by smart meters decreases, the ability to disaggregate low-resolution data increases.

We developed a new privacy-preserving smart metering system. Our scheme is private under the differential privacy model and therefore provides strong and provable guarantees. With our scheme, an (electricity) supplier can periodically collect data from smart meters and derive aggregated statistics while learning only limited information about the activities of individual households. For example, a supplier cannot tell from a user’s trace when he watched TV or turned on heating. Our scheme is simple, efficient and practical. Processing cost is very limited: smart meters only have to add noise to their data and encrypt the results with an efficient stream cipher.

This work was presented at IH’11 (the Information Hiding Conference, 2011) [34].

- **Protecting against Physical Resource Monitoring**

This work considers the problem of resource monitoring. We consider the scenario where an adversary is physically monitoring on the resource access, such as the electricity line or gas pipeline, of a user in order to learn private information about his victim. Recent works, in the context of smart metering, have shown that a motivated adversary can basically profile a user or a family solely from his electricity traces. However, these works only consider the case of a semi-honest-but-non-intrusive adversary that is only trying to learn information from the consumption reports sent by the user. This work, instead, considers the much more challenging case of a intrusive semi-honest adversary, i.e. a semi-honest adversary that is in addition physically monitoring the resource by modifying the distribution network. We aim at answering to the following question: is it possible to design a resource distribution scheme that prevents resource monitoring and provides strong protection? We propose and analyze several possible solutions. The proposed solutions provide different privacy bounds and performance results. This work was presented at WPES’11 (ACM Workshop on Privacy in the Electronic Society) [35].

- **The Failure of Noise-Based Non-Continuous Audio Captchas**

CAPTCHAs, which are automated tests intended to distinguish humans from programs, are used on many web sites to prevent bot-based account creation and spam. To avoid imposing undue user friction, CAPTCHAs must be easy for humans and difficult for machines. However, the scientific basis for successful CAPTCHA design is still emerging. This project examines the widely used class of audio CAPTCHAs based on distorting non-continuous speech with certain classes of noise and demonstrates that virtually all current schemes, including ones from Microsoft, Yahoo, and eBay, are easily broken. More generally, we describe a set of fundamental techniques, packaged together in our Decaptcha system, that effectively defeat a wide class of audio CAPTCHAs based on non-continuous speech. Decaptcha’s performance on actual observed and synthetic CAPTCHAs indicates that such speech CAPTCHAs are inherently weak and, because of the importance of audio for various classes of users, alternative audio CAPTCHAs must be developed.

This work was presented at IEEE Security and Privacy 2011 [33].
BlueBear: Privacy in P2P systems

We have started a new project called bluebear on privacy threats in the Internet. Indeed, the Internet has never been designed with privacy in mind. For instance, the Internet is based on the IP protocol that exposes the IP address of a user to any other users it is communicating with. However, we believe that current users of the Internet do not realize how much they compromise their privacy by using the Internet. Indeed, the common wisdom is that there are so many users in the Internet that it is not feasible for an attacker, apart may be for national agencies, to globally compromise the privacy of a large fraction of users. Therefore, finding a specific user is like looking for a needle in a haystack. The goal of the bluebear project is to raise attention on privacy issues when using the Internet. In particular, we want to show that without any dedicated infrastructure, it is possible to globally compromise the privacy of Internet users. BitTorrent is arguably the most efficient peer-to-peer protocol for content replication. However, BitTorrent has not been designed with privacy in mind and its popularity could threaten the privacy of millions of users.

In a first study we showed that it is possible to continuously monitor from a single machine most BitTorrent users and to identify the content providers (also called initial seeds). We performed a very large monitoring operation continuously “spying” on most BitTorrent users of the Internet from a single machine and for a long period of time. During a period of 103 days, we collected 148 million IP addresses downloading 2 billion copies of contents. We then identified the IP address of the content providers for 70% of the BitTorrent contents we spied on. We showed that a few content providers inject most contents into BitTorrent and that those content providers are located in foreign data centres. We also showed that an adversary could compromise the privacy of any peer in BitTorrent and identify the big downloaders that we define as the peers who subscribe to a large number of contents. This is a major privacy threat as it is possible for anybody in the Internet to reconstruct all the download and upload history of most BitTorrent users. This work was published in LEET 2010.

To circumvent this kind of monitoring, BitTorrent users are increasingly using anonymizing networks such as TOR to hide their IP address from the tracker and, possibly, from other peers. We explored in a second study whose goal was to Exploit P2P Applications to Trace and Profile Tor Users, to which extent a P2P protocol such as BitTorrent, when not designed to protect users information, leak information that may compromise the identity of users. We quantified such an issue with BitTorrent on top of anonymizing networks. We also designed an attack that reveals the identity of Tor users (We showed that it is possible to retrieve the IP address for more than 70% of BitTorrent users on top of TOR). Moreover, once the IP address of a peer is retrieved, it is possible to link to the IP address other applications used by this peer on top of TOR [45].

The fact that it is hard for a person to map an IP address to an identity mitigates the impact of the privacy attacks we described. However, we show that we can exploit a peer-to-peer VoIP system to associate a social identity (name, email address, etc.) to an IP address [46]. This means that anybody can now find this mapping that was only known by ISPs or big companies (like Google and Facebook), but never communicated unless in case of a legal action. The privacy threat is thus very high because this mapping enables blackmail, social attacks, targeted phishing attacks, etc.

As a proof of concept, we show that it is possible to track VoIP users mobility and BitTorrent downloads [46] using Skype, one of the most popular VoIP system with more than 500 millions registered users.

All these works received a very large media coverage (see http://www-sop.inria.fr/members/Arnaud.Legout/Projects/bluebear.html).
6.3. Network measurement, modeling and understanding


The main objective of our work in this domain is a better monitoring of the Internet and a better control of its resources. 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 ECODE FP7 project, we work on a network-wide monitoring architecture that, given a measurement task to perform, tune the monitors inside the network optimally so as to maximize the accuracy of the measurement results while limiting the overhead resulting from collected traffic. 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.

- **Internet traffic classification by means of packet level statistics**
  One of the most important challenges for network administrators is the identification of applications behind the Internet traffic. This identification serves for many purposes as in network security, traffic engineering and monitoring. The classical methods based on standard port numbers or deep packet inspection are unfortunately becoming less and less efficient because of encryption and the utilization of non standard ports. In this activity, we come up with an online iterative probabilistic method that identifies applications quickly and accurately by only using the size of packets. Our method associates a configurable confidence level to the port number carried in the transport header and is able to consider a variable number of packets at the beginning of a flow. By verification on real traces we observe that even in the case of no confidence in the port number, a very high accuracy can be obtained for well known applications after few packets were examined. In another work [39], we make a complete study about the inter-packet time to prove that it is also a valuable information for the classification of Internet traffic. We discuss how to isolate the noise due to the network conditions and extract the time generated by the application. We present a model to preprocess the inter-packet time and use the result as input to the learning process. We discuss an iterative approach for the online identification of the applications and we evaluate our method on two different real traces. The results show that the inter-packet time is an important parameter to classify Internet traffic.

  We pursued this activity further by accounting for the communication profiles of hosts for the purpose of a better traffic classification [39], [38], [40]. We use the packet size and the inter-packet time as the main features for the classification and we benefit from the traffic profile of the host (i.e. which application and how much) to refine the classification and decide in favor of this or that application. The host profile is then updated online based on the result of the classification of previous flows originated by or addressed to the same host. We evaluate our method on real traces using several applications. The results show that leveraging the traffic pattern of the host ameliorates the performance of statistical methods. They also prove the capacity of our solution to derive profiles for the traffic of Internet hosts and to identify the services they provide.

  For a more thorough study of the traffic classification problem by means of packet statistics and host profiles, we refer to the PhD dissertation of Mohamad Jaber who was the main contributor to this activity inside the EPI Planete.

- **Adaptive network-wide traffic monitoring**
  The remarkable growth of the Internet infrastructure and the increasing heterogeneity of applications and users’ behavior make more complex the manageability and monitoring of ISP networks and raises the cost of any new deployment. The main consequence of this trend is an inherent disagreement between existing monitoring solutions and the increasing needs of management applications.
In this context, we work on the design of an adaptive centralized architecture that provides visibility over the entire network through a network-wide cognitive monitoring system. Given a measurement task, the proposed system drives its own configuration, typically the packet and flow sampling rates in routers, in order to address the tradeoff between monitoring constraints (processing and memory cost, collected data) and measurement task requirements (accuracy, flexibility, scalability). We motivate our architecture with an accounting application: estimating the number of packets per flow, where the flow can be defined in different ways to satisfy different objectives (e.g., Domain-to-Domain traffic, all traffic originated from a domain, destined to a domain). The architecture and the algorithms behind it are explained in paper published in 2010 for the case of a proactive control and in [43] for the case of a reactive control. In [44] the architecture and its algorithms are specified to a flow counting application. In all these works, the performances of our architecture are being validated in typical scenarios over an experimental platform we developed for the purpose of the study. Our platform is called MonLab (Monitoring Lab) and is described with more details in the Section on produced softwares. For now, MonLab presents a new approach for the emulation of Internet traffic and for its monitoring across the different routers. It puts at the disposal of users a real traffic emulation service coupled to a set of libraries and tools capable of Cisco NetFlow data export and collection, the overall destined to run advanced applications for network-wide traffic monitoring and optimization.

The activities in this direction are funded by the ECODE FP7 STREP project (Sep. 2008 - Dec. 2011). The dissertation of Imed Lassoued [21] provides an introduction to the field in addition to details on our contributions and the MonLab emulation platform.

• Spectral analysis of packet sampled traffic

In network measurement systems, packet sampling techniques are usually adopted to reduce the overall amount of data to collect and process. Being based on a subset of packets, they hence introduce estimation errors that have to be properly counteracted by a fine tuning of the sampling strategy and sophisticated inversion methods. This problem has been deeply investigated in the literature with particular attention to the statistical properties of packet sampling and the recovery of the original network measurements. Herein, we propose a novel approach to predict the energy of the sampling error on the real time traffic volume estimation, based on a spectral analysis in the frequency domain. We start by demonstrating that errors due to packet sampling can be modeled as an aliasing effect in the frequency domain. Then, we exploit this theoretical finding to derive closed-form expressions for the Signal-to-Noise Ratio (SNR), able to predict the distortion of traffic volume estimates over time. The accuracy of the proposed SNR metric is validated by means of real packet traces. The analysis and the expressions of the SNR that stemmed from are described in [26]. In [52], we adopt such a model to design a real-time algorithm, that sets the IPFIX counter export timers in order to grant, to each flow, a target estimation accuracy. The work within this direction has been partially supported by the FP7 ECODE project.

• Monitoring the quality of the Internet access by end-to-end probes

The detection of anomalous links and traffic is important to manage the state of the network. Existing techniques focus on detecting the anomalies but little attention has been devoted to quantify to which extent network anomaly affects the end user access link experience. We refer to this aspect as the local seriousness of the anomaly. In order to quantify the local seriousness of an anomaly, we consider the percentage of affected destinations, that we call the impact factor. In order to measure it, a host should monitor all possible routes to detect any variation in performance, but this is not practical in reality. In this activity, funded by the ANR CMON project, we work on finding estimates for the impact factor and the local seriousness of network anomalies through a limited set of measurements to random nodes we call landmarks.
We initially study the user access network to understand the typical features of its connectivity tree. Then, we define an unbiased estimator for the local seriousness of the anomaly and a framework to achieve three main results: (i) the computation of the minimum number of paths to monitor, so that the estimator can achieve a given significance level, (ii) the localization of the anomaly in terms of hop distance from the local user, and (iii) the optimal selection of landmarks. We are using real data to evaluate in practice the local seriousness of the anomaly and to determine the sufficient number of landmarks to select randomly without knowing anything on the Internet topology. The localization mechanism leverages the study on the connectivity tree and the relationship between the impact factor and the minimum hop distance of an anomaly. Our first results show that the impact factor is indeed a meaningful metric to evaluate the quality of Internet access. The current work focuses on extending this solution towards a collaborative setting where different end users collaborate together by exchanging the results of their observations. The objective will be a better estimation of the impact factor by each of them and a finer localization of the origin of any network problem.

On the experimental side, we have implemented the solution in a tool called ACQUA, which stands for Application for Collaborative Estimation of QUality of Internet Access 2. We design an anomaly detection mechanism based on the histogram of delay measurements and the likelihood of observations. Then, we give to ACQUA a pipeline based software architecture, and we go deeply into experimentation inside and outside Planetlab. We show what the properties and usage of the algorithm are, focusing also on how this tool can help us to get information about the network anomalies detected. Later we extend the idea of Impact Factor Estimation (IFE) by using what we call Inverse IFE from Planetlab, where the computer of the user whose connectivity is tested has a completely passive role in the measurements procedure. We study its strong and weak points, and we show conditions under which Inverse IFE from Planetlab gives similar results to traditional IFE.

- **Applied Internet Measurements**

The performance of several Internet applications often relies on the measurability of path similarity between different participants. In particular, the performance of content distribution networks mainly relies on the awareness of content sources topology information. It is commonly admitted nowadays that, in order to ensure either path redundancy or efficient content replication, topological similarities between sources is evaluated by exchanging raw traceroute data, and by a hop by hop comparison of the IP topology observed from the sources to the several hundred or thousands of destinations. In this project, based on real data we collected, we advocate that path similarity comparisons between different Internet entities can be much simplified using lossy coding techniques, such as Bloom filters, to exchange compressed topology information. The technique we introduce to evaluate path similarity enforces both scalability and data confidentiality while maintaining a high level of accuracy. In addition, we demonstrate that our technique is scalable as it requires a small amount of active probing and is not targets dependent. This work has been published in [25].

- **Reliability of Geolocation Databases**

In this project, we question the reliability of geolocation databases, the most widely used technique for IP geolocation. It consists in building a database to keep the mapping between IP blocks and a geographic location. Several databases are available and are frequently used by many services and web sites in the Internet. Contrary to widespread belief, geolocation databases are far from being as reliable as they claim. We conduct a comparison of several current geolocation databases -both commercial and free- to have an insight of the limitations in their usability. First, the vast majority of entries in the databases refer only to a few popular countries (e.g., U.S.). This creates an imbalance.

2 http://planete.inria.fr/acqua/
in the representation of countries across the IP blocks of the databases. Second, these entries do not reflect the original allocation of IP blocks, nor BGP announcements. In addition, we quantify the accuracy of geolocation databases on a large European ISP based on ground truth information. This is the first study using a ground truth showing that the overly fine granularity of database entries makes their accuracy worse, not better. Geolocation databases can claim country-level accuracy, but certainly not city-level. This study has been published in CCR [28].

- **Impact of Live Streaming Traffic**

  Video streaming is the most popular traffic in the Internet and a strong case for content centric networks. Therefore, it is fundamental to understand the network traffic characteristics of video streaming. In this work [49], we extensively studied the network traffic characteristics of YouTube and Netflix (the most popular video streaming traffic in the USA). We have shown that the traffic characteristics vastly depends on the type of browser, mobile application, and container (Flash, Silverlight, HTML5) used.

### 6.4. Experimental Environment for Future Internet Architecture

**Participants:** Walid Dabbous, Thierry Parmentelat, Baris Metin, Frédéric Urbani, Daniel Camara, Alina Quereilhac, Shafqat Ur-Rehman, Thierry Turletti, Julien Tribino.

- **SFA Federation of experimental testbeds**

  The OneLab2 project has come to its end in spring 2010. We are now involved in the NOVI (E.U. STREP) project, the F-Lab (French A.N.R.) project, and have the lead of the “Federation” WorkPackage of OpenLab (E.U. IP) project. Within these frameworks, we are codevelopping with Princeton University a reference implementation for 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 2011 we have focused on the maturation of the SFA codebase, with several objectives in mind. Firstly we have contributed to a major overhaul of the specification as defined essentially within the GENI (N.S.F.) Project, with participations from all over the world. These changes, that affected both the core API and the schema used to expose and manage resource specifications, aimed at reaching a mature level of interoperability between the PlanetLab world and the EmuLab a.k.a. ProtoGeni world that has its own implementation, and are now available in SFA-2.0 issued late 2011.

  Secondly, the SFA codebase has been redesigned to provide a more generic shelter that other testbeds can easily leverage in order to come up with their own SFA-compliant wrapper. This is perceived as a powerful means to foster further adoption of the architecture, and the Planète team has been intramental in bringing two entirely different testbeds to the federation, namely Senslab - developed in other INRIA Project-teams - and FEDERICA, the outcome of another E.U.-funded Project. Along the same lines we are working, although more remotely, with NICTA in Australia that publishes the O.M.F. testbed for running wireless testbeds, and who are interested in adopting this federation paradigm.

  Finally, as part of the pure PlanetLab development, we have added a feature for running nodes in a 'reservable' mode, which breaks the usual best-effort PlanetLab model, but turned out very helpful both for making experiments possible, that needed a more reproducible behaviour of experiments, and also in a federation perspective, for closing the usage gap with, notably wireless testbeds, that typically have a reservable-only provisioning mechanisms.
• **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.

• **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.

• **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](http://www.nsnam.org) web site. This web site is now finalized. However, activities related to developing the consortium have slowed down during 2011 due to the leave of Mathieu Lacage. We plan to put more resources on this aspect in 2012.
Using Independent Simulators, Emulators, and Testbeds for Easy Experimentation

Evaluating new network protocols, applications, and architectures uses many kinds of experimentation environments: simulators, emulators, testbeds, and sometimes, combinations of these. As the functionality and complexity of these tools increases, mastering and efficiently using each of them is becoming increasingly difficult.

We designed the preliminary prototype of the Network Experiment Programming Interface (NEPI) whose goal is to make easier the use of different experimentation environments, and switch among them easily. NEPI intends to make it possible to write a single script to control every aspect of a potentially mixed experiment, including a hierarchical network topology description, application-level setup, deployment, monitoring, trace setup, and trace collection. We showed how a single object model which encompasses every aspect of a typical experimentation workflow can be used to completely describe experiments to be run within very different experimentation environments. 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. Last year, 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. This year we have 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. The NEPI approach has been validated by implementing support for three complementary environments: a physical testbed, a network emulator, and a network simulator. Furthermore, we showed with a concrete experiment use case, available online for reproduction, how easy it is with NEPI to integrate these environments for hybrid-experimentation [48].

Guidelines for the accurate design of empirical studies in wireless networks

Traditionally, wireless protocol proposals have been often tested and validated using only analytical and simulation models [73]. However, as the wireless environment is very complex to model accurately, and since the cost of wireless cards has decreased in an exponential way, today more and more research papers include evaluation of new proposals using experimentation on real devices. Indeed, experimentation is a mandatory step before possible deployment of new network protocols with real users. However, wireless experimentation is much more complex to set up and run than simulation, and it is important to avoid many pitfalls that can occur during experimentation. The objectives of this work are twofold. First, we described typical problems currently encountered in wireless-based experimentation, and we presented simple guidelines to avoid them [50]. Second, we proposed an experimental methodology where the detection of anomalies, calibration of the measurement setup, and clear definition of the scenario (among others) make easier the repeatability of results [55]. This work has been done in collaboration with Cristian Tala, Luciano Ahumada and Diego Dujovne from the Universidad Diego Portales of Chili, in the context of the WELCOME STIC AMSud 2011.

Multicast Video Streaming over WiFi Networks: Impact of Multipath Fading and Interference
We conducted an experimental study in order to analyse the impact of interference, multipath fading and path loss on multicast video streaming (i.e., goodput, packet loss and ultimately on the video quality) using off-the-shelf fixed WiFi equipment in a wireless (802.11 b/g) local area network (WLAN) environment. We used the ricean K-factor as a measure of multipath fading, spectrum analyzer to estimate channel interference and received signal strength indicator (RSSI) as indication of signal power and attenuation. In order to realistically measure aforementioned metrics, we conducted extensive wireless experiments against six test cases representing common real-world situations using off-the-shelf wireless equipment.

We showed that interference has more impact on performance than multipath fading. Multipath fading can result in considerable performance degradation in environments where moving objects cause perturbation. On the contrary, channel interference is more frequent and more prominent cause of performance degradation in wireless networks because ISM 2.4 GHz band is increasingly being utilized in homes and work places. Being able to quantify the impact of multipath fading and interference is crucial in planning, troubleshooting, managing as well as benchmarking and optimizing wireless networks. This study has been published in MediaWin 2011 [51].

• Making easier Experimentation

Wireless experimentations are challenging to evaluate due to the high variability of the channel characteristics and its sensitivity to interferences.

Merging traces represents a complex problem especially in wireless experimentations, due to packet redundancy in multiple probes. Merging traces solutions need to be efficient in order to process the large amount of generated traces. These solutions should provide an output data structure that allows easy and fast analysis and must be scalable in order to be used in large and various experimental settings. We have designed an algorithm that performs trace synchronization and merging in a scalable way. The algorithm output is stored in a configured MYSQL database allowing for smart packet trace storage. This solution reduces processing time by 400% and storage space by 200% with regard to raw trace file solutions. It has been implemented in an open source software called CrunchXML, available under the GNU General Public License v2 at http://twiki-sop.inria.fr/twiki/bin/view/Projets/Planete/CrunchXML.

• An Integration Framework for Network Experimentation

Many different experimentation environments address complementary aspects of network protocol evaluation, but because of their disparities and complexities it is often hard to use them to reproduce the same experiment scenario.

Simulation is often used for the evaluation of new network protocols and architectures. In order to perform more realistic simulations, modern simulators such as ns-3 integrate more detailed models and even support direct execution of real protocol code. However, such complex models require more computational and memory requirements. We have studied the feasibility of a hybrid approach based on distributing a complex simulation scenario on several nodes in a grid network. We showed that by exploiting the real time operation of the ns-3 simulator, it is possible to map such complex scenarios on grid nodes. We also proposed a basic mapping algorithm to distribute a simulation scenario in several nodes [42].
6. New Results

6.1. Optimized protocols implementation and networking equipments

6.1.1. Locating Virtual Infrastructures: Users and InP Perspectives

Participants: Paulo Gonçalves, Guilherme koslovski.

This is a joint work with Pascale Vicat-Blanc (Lyatiss) and Sébastien Soudan (Lyatiss).

The Cloud Computing wave consolidates the on-demand provisioning of configurable virtual machines. Recent projects have proposed the extension of the original IaaS paradigm to provide dynamic virtual networks to interconnect virtual IT resources, composing Virtual Infrastructures (VIs). In this new scenario, users with different objectives and expectations can rent dynamically provisioned virtual infrastructures to execute their applications during a given time slot. VIs can be allocated anywhere on top of a distributed and virtualized substrate. This decoupling from the geographical location introduces concerns such as a latency increase in network communications (user’s perspective), and the fragmentation of physical resources (Infrastructure Provider’s - InP - perspective). This context motivates efforts to investigate and deploy new models and tools which consider the geographical location of virtual infrastructures. Our work concentrates on the allocation of VIs guided by both the user’s and the InP’s constraints. We propose a formulation of the allocation problem considering the user’s expectations as well as the physical-substrate provider’s goals. Our initial experiments demonstrate that it is possible to improve the quality of the virtual-infrastructure allocation (user perspective) while simultaneously decreasing the physical substrate’s fragmentation and the substrate’s cost.

6.1.2. Energy-efficient reservation infrastructure for large-scale distributed systems

Participants: Anne-Cécile Orgerie, Laurent Lefèvre, Guérin-Lassous Isabelle.

Over the past few years, the energy consumption of Information and Communication Technologies (ICT) has become a major issue. Nowadays, ICT accounts for 2% of the global CO2 emissions, an amount similar to that produced by the aviation industry. Large-scale distributed systems (e.g. Grids, Clouds and high-performance networks) are often heavy electricity consumers because – for high-availability requirements – their resources are always powered on even when they are not in use. Reservation-based systems guarantee quality of service, allow for respect of user constraints and enable fine-grained resource management. For these reasons, in the context of Anne-Cécile Orgerie Phd (defended in September 2011), we proposed an energy-efficient reservation framework to reduce the electric consumption of distributed systems and dedicated networks. The framework, called ERIDIS, is adapted to three different systems: data centers and grids, cloud environments and dedicated wired networks. By validating each derived infrastructure, we show that significant amounts of energy can be saved using ERIDIS in current and future large-scale distributed systems [54].

6.1.3. Energy efficiency in exascale infrastructures

Participants: Mehdi Diouri, Olivier Gluck, Laurent Lefevre.

Joint work with F. Cappello (JLPC, joint laboratory between INRIA and NCSA).

In Diouri’s PhD, we address the issue of energy efficiency for exascale supercomputers. We first proposed a green architecture for exascale systems gathering some new solutions to “consume less” energy and to “consume better”. This architecture involves interactions with the different actors interfering directly or indirectly with the supercomputer: its user, its administrator, its resource manager and the energy supplier. Then we were interested into leaning on this green architecture in order to propose some green services that will be offered for applications that will run on exascale systems. Our approach consists in evaluating the power overhead induced by some existing services, and by proposing a green version of these services that takes into account the constraints imposed by the different actors involved. In 2011, we specifically aimed to apply our approach for fault tolerance protocols in their normal functioning stage and in case of failure. [38]
6.1.4. Energy profiling and green leverages for high performance computing applications

Participants: Ghislain Landry Tsafack, Jean-Patrick Gelas, Laurent Lefevre.

Ghislain Landry TSAFACK CHETSA has started his PhD in January 2011, within the framework of INRIA HEMERA project, on: “Energy profiling and green leverages for high performance computing applications” (co-advisement with Jean-Marc PIERSON and Patrícia STOLF from IRIT). During the course of this first year, we have investigated the possibility of characterizing distributed applications considering their energy / power profile. We first carried out a set of experiments for a better understanding of the application’s behavior and the impact that this behavior may have on its power consumption. Results led us to the assumption that any individual application run can be represented as a sequence of basic operations including computation, memory accesses, disk and network accesses over a given time period. We next rely on that assumption to define application’s energy profile. Application’s energy profile helps to prevent the fallout of any action that may be taken to reduce its power usage. To guarantee reasonable results, i.e., reduce energy with less performance degradation, we designed an energy prediction model capable of predicting power usage of a wide range of high performance computing (HPC) applications.

6.1.5. Towards virtualized home gateways

Participants: Jean-Patrick Gelas, Laurent Lefevre, Anne-Cécile Orgerie.

Joint work with Dino Lopez Pacheco (University of Nice) and Referi Assefa (Addis Abeba University, Ethiopia).

Virtualizing services located on end to end parts of the networks and making them available for a large number of applications and users is now becoming a real challenge. Within the scope of the GreenTouch project, we are exploring models, simulation tools (ECOFEN) and software prototypes able to demonstrate the impact of such approach in terms of energy reduction.

6.2. Quality of Service and Transport Layer for Future Networks

6.2.1. On the Impact of the Flow-Size Distribution’s Tail Index on Network Performance with TCP Connections

Participant: Paulo Gonçalves.

This is a joint work with Oana Goga (UPMC, Lip6) and Patrick Loiseau (Eurecom).

In this work, we studied the impact of the flow-size distribution on network performance in the case of a single bottleneck with finite buffer. To tackle the case where flows are transmitted with the TCP protocol, we use real experiments and ns-2 simulations. Our preliminary results show that the distribution’s tail index impacts the performance in a more complex way than what is reported in existing literature. In particular, we exhibit situations where a heavier tail gives better performance for certain metrics. We argue that a main cause of our observed results is the transient behavior at the beginning of each flow.

6.2.2. Available Bandwidth Estimation for Multihop Wireless Networks

Participants: Isabelle Guérin Lassous, Van Nam Nguyen.

Estimating the available bandwidth in IEEE 802.11-based multi-hop wireless networks is a very difficult task due to the medium sharing among contending nodes and collisions between hidden stations. Several methods have been proposed so far for these networks to compute the available bandwidth on wireless links. If some recent solutions such as ABE and IAB now take into account collisions and their impact on the mean backoff, none of them considers the packet retransmissions due to collisions, although these retransmissions have an impact on the available bandwidth. In this work, we have proposed a new available bandwidth estimation for multi-hop wireless networks called RABE (Retransmission-based Available Bandwidth). This method integrates the average number of retransmission attempts in the available bandwidth estimation, in addition to other relevant parameters like the idle periods durations and the collision probability. RABE has been evaluated
by simulation and the obtained results show that RABE can achieve a mean error ratio of 17% in comparison with the real measurement. Furthermore RABE is at least two times more accurate than ABE and ten times more accurate than IAB.

**6.2.3. On The Recovery Performance of Single- and Multipath OLSR in Wireless Multi-Hop Networks**

**Participants:** Inès Doghri, Isabelle Guérin Lassous.

In this work, we study and improve the recovery properties of single- and multipath routing strategies when facing network failure situations. In particular, we focus our study on two MANET routing protocols: OLSR and its multipath extension MP-OLSR. In various wireless multi-hop network environments, especially in multiple chain topologies, we define and evaluate the latency introduced by these protocols to find a new path after a link failure. Theoretical estimations and simulation results show that, under dual chain-topologies, this latency can be too long and incompatible with the needs of loss and delay constrained applications. As the source nodes cannot detect link failures immediately because of the delay incurred by the well-known nature of link state protocols in general, and of OLSR Topology Control (TC) messages in particular, these nodes keep sending packets along broken paths. We thus study the inconsistencies between the actual network topology and the nodes’ own representation. After analyzing the consequences of this long latency, we seek to alleviate these problems with the introduction of adapted mechanisms. We propose three new different schemes and accordingly extend the original OLSR and MP-OLSR protocols in order to decrease the expected latency and improve the protocol performance. Simulation results show a steep decrease of the latency when using these new schemes in dual chain-topologies. We also discuss these results in terms of packet loss, end-to-end delay and overhead.

**6.3. High Speed Network’s traffic metrology and statistical analysis**

**6.3.1. A long-range dependent model for network traffic with flow-scale correlations**

**Participant:** Paulo Gonçalves.

For more than a decade, it has been observed that network traffic exhibits long-range dependence and many models have been proposed relating this property to heavy-tailed flow durations. However, none of these models consider correlations at flow scale. Such correlations exist and will become more prominent in the future Internet with the emergence of flow-aware control mechanisms correlating a flow’s transmission to its characteristics (size, duration, etc.). In our present work, we study the impact of the correlation between flow rates and durations on the long-range dependence of aggregate traffic. Our results extend those of existing models by showing that two possible regimes of long-range dependence exist at different time scales. The long-range dependence in each regime can be stronger or weaker than standard predictions, depending on the conditional statistics between the flow rates and durations. In the independent case, our proposed model consistently reduces to former approaches. The pertinence of our model is validated on real web traffic traces, and its ability to accurately explain the Hurst parameter is validated on both web traces and numerical simulations.

**6.3.2. A recurrent solution of Ph/M/c/N-like and Ph/M/c-like queues**

**Participant:** Thomas Begin.

This work has been accepted for publication by the Journal of Applied Probability [50] and was performed in collaboration with Pr. Brandwajn (UCSC).
We propose an efficient semi-numerical approach to compute the steady-state probability distribution for the number of requests at arbitrary and at arrival time instants in Ph/M/c-like systems in which the inter-arrival time distribution is represented by an acyclic set of memoryless phases. Our method is based on conditional probabilities and results in a simple computationally stable recurrence. It avoids the explicit manipulation of potentially large matrices and involves no iteration. Due to the use of conditional probabilities, it delays the onset of numerical issues related to floating-point underflow as the number of servers and/or phases increases. For generalized Coxian distributions, the computational complexity of the proposed approach grows linearly with the number of phases in the distribution.

6.3.3. A Markovian model based on SIR epidemic classification to reproduce the workload dynamics of a VoD server

Participants: Shubhabrata Roy, Thomas Begin, Paulo Gonçalves.

We have devised a Markovian model, based on the SIR epidemic classification, to reproduce the workload dynamics that can be observed on a VoD (Video on Demand) server. This model basically relies on the dynamic between three distinct populations (i.e., current watchers, past watchers and potential watchers). It also embeds events with very low probability but high impact on its overall behavior corresponding to the occurrence of a flash crowd or the buzz effect on a VoD server. The steady-state solution to this model has shown that it exhibits a behavior qualitatively close to what can be expected from a real-life VoD server. We have also shown that the workload process as delivered this model satisfies a large deviation principle. Our future work aims at taking advantage of this information to devise a new scheme for allocating available resources in a VoD server.

6.3.4. A comparative study of existing MBAC using real network traces

Participants: Doreid Ammar, Thomas Begin, Isabelle Guérin-Lassous.

We have evaluated the respective performance of several MBACs (Measurement-based admission control) using a realistic framework in which the pattern of the background traffic follows experimental traces collected on real-life networks. This study has allowed to highlight the respective discrepancies between MBACs in terms of easiness to implement and attained performance. This work will now focus on the design of a new MBAC based on a iteratively learned model.

6.3.5. Graph Based Classification of Content and Users in BitTorrent

Participants: Paulo Gonçalves, Marina Sokol.

This is a joint work with Konstantin Avrachenkov (INRIA Maestro) and Arnaud Legout (INRIA Planete). 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 our present work is to provide a framework for sub-classification of P2P traffic generated by the BitTorrent protocol. Unlike previous works, we cannot rely on packet level characteristics and on the standard supervised machine learning methods. The application of the standard supervised machine learning methods is based on the availability of a large set of parameters (packet size, packet inter-arrival time, etc.). Since P2P transfers are based on the same BitTorrent protocol we cannot use this set of parameters to classify P2P content and users. Instead we can make use of the bipartite user-content graph. This is a graph formed by two sets of nodes: the set of users (peers) and the set of contents (downloaded files). From this basic bipartite graph we also construct the user graph, where two users are connected if they download the same content, and the content graph, where two files are connected if they are both downloaded by at least one same user. The general intuition is that the users with similar interests download similar contents. This intuition can be rigorously formalized with the help of graph based semi-supervised learning approach.
6.3.6. **Generalized Optimization Framework for Graph-based Semi-supervised Learning**  
**Participants:** Paulo Gonçalves, Marina Sokol.

This is a joint work with Konstantin Avrachenkov (INRIA Maestro).

We develop a generalized optimization framework for graph-based semi-supervised learning. The framework gives as particular cases the Standard Laplacian, Normalized Laplacian and PageRank based methods. We have also provided new probabilistic interpretation based on random walks and characterized the limiting behavior of the methods. The random walk based interpretation allows us to explain differences between the performances of methods with different smoothing kernels. It appears that the PageRank based method is robust with respect to the choice of the regularization parameter and the labelled data. We illustrate our theoretical results with two realistic datasets, characterizing different challenges: *Les Misérables* characters social network and Wikipedia hyper-link graph. The graph-based semi-supervised learning classifies the Wikipedia articles with very good precision and perfect recall employing only the information about the hyper-text links.

6.3.7. **On the estimation of the large deviations spectrum**  
**Participant:** Paulo Gonçalves.

This is a joint work with Julien Barral (Univ. Paris 13)

We propose an estimation algorithm for large deviations spectra of measures and functions. The algorithm converges for natural examples of multifractals.

6.3.8. **Adaptive Multiscale Complexity Analysis of Fetal Heart Rate**  
**Participant:** Paulo Gonçalves.

This is a joint work with Patrice Abry (ENS Lyon, CNRS) and Muriel Doret (Hospice civils de Lyon, Univ. Lyon 1)

*Per partum* fetal asphyxia is a major cause of neonatal morbidity and mortality. Fetal heart rate monitoring plays an important role in early detection of acidosis, an indicator for asphyxia. This problem is addressed in this paper by introducing a novel complexity analysis of fetal heart rate data, based on producing a collection of piecewise linear approximations of varying dimensions from which a measure of complexity is extracted. This procedure specifically accounts for the highly non-stationary context of labor by being adaptive and multiscale. Using a reference dataset, made of real *per partum* fetal heart rate data, collected *in situ* and carefully constituted by obstetricians, the behavior of the proposed approach is analyzed and illustrated. Its performance is evaluated in terms of the rate of correct acidosis detection versus the rate of false detection, as well as how early the detection is made. Computational cost is also discussed. The results are shown to be extremely promising and further potential uses of the tool are discussed.
SARDES Project-Team

5. New Results

5.1. Languages and Foundations: Process algebra

Participants: Damien Pous, Alan Schmitt, Jean-Bernard Stefani, Claudio Mezzina, Cinzia di Giusto.

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 project.

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 the reversible higher-order pi-calculus and obtained a new encoding of it in the higher-order pi-calculus which improves on the result we published in Concur 2010 by proving the faithfulness of the encoding with a much finer equivalence relation. We also developed a reversible variant of the higher-order pi-calculus where reversibility can be controlled by means of an explicit rollback primitive [37]. We have proved that this rollback primitive is sound and complete in that it provides a causally consistent and complete reversal of concurrent computations, and we have developed a low-level semantics for this primitive, closer to an actual distributed implementation, which we have proved equivalent to the high-level one. All these results are presented in detail in Claudio Mezzina’s forthcoming PhD thesis, and have been developed in cooperation with the INRIA Focus team at the University of Bologna.

An interesting and expressive component model for embedded systems is the BIP component model [58], developed by J. Sifakis’ team at the Verimag Laboratory, which features hierarchical software architectures, explicit constructs for specifying component compositions (glues), and multipoint synchronization under priority constraints. We have begun a process calculus analysis of BIP, with a view to combine the reactive features of BIP with the dynamic reconfiguration features of Fractal. Our first result takes the form of new process calculus, called CAB, which we have proved to be a conservative extension of BIP. CAB also enabled us to study the intrinsic expressivity of the BIP model and to prove that priority constraints are essential to BIP expressivity [34].

We have made significant progress on the formalization in the Coq proof assistant of a core higher-order π-calculus, called HOcore [20]. We have in particular adapted a canonical locally nameless representation of binding to handle alpha-conversion in our formalization. Several major theorems of HOcore, in particular the fact that IO-bisimulation is correct in relation to barbed congruence and is decidable. This work has been submitted for publication.

A longer version of our work on untyping theorems and cyclic linear logic has been accepted for publication in LMCS [24], and a book chapter on up-to techniques for bisimulations, written with Davide Sangiorgi from the INRIA Focus team in Bologna, has been published by Cambridge University Press [46].

Together with Filippo Bonchi (LIP, ENS Lyon), we have worked on a new algorithm for checking the language equivalence of non-deterministic finite automata (NFA). This algorithm improves on the standard Hopcroft and Karp’ algorithm, by using up-to techniques. The first empirical tests look really promising [47].

Together with Tom Hirschowitz (LAMA, U. de Chambéry), we have worked on a categorical model of CCS, where innocent strategies are pre-sheaves. This work has been presented at the ICE workshop [36], and a long version has been submitted to SACS.

5.2. Languages and Foundations: Proof tactics

Participants: Damien Pous, Thomas Braibant.
The goal of this work is to develop proof-assistant-based tools for verifying distributed systems and distributed abstract machines. In particular, we aim to support the derivation of fully formal proofs of correctness for abstract machines supporting the component-based languages and programming models we develop.

We have presented our work about tools for rewriting modulo AC in Coq at CPP'11 [32]. An extended version of our work on Kleene algebra (ATBR, first published at ITP'10), was accepted for publication in LMCS [19]. Also on the Coq side, we have developed a library for verifying hardware circuits, which was also presented at CPP’11 [31].

5.3. Control for adaptive systems: Discrete control for adaptive and reconfigurable systems

Participants: Eric Rutten, Noël de Palma, Olivier Gruber, Fabienne Boyer, Tayeb Bouhadiba, Xin An.

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. Both are integrated into a programming language, called BZR, and its compiler, as an extension of the Heptagon language. We thus have a complete tool-supported method from control modeling down to concrete execution, considering different execution models, and targeting either software or hardware. We explore control theory for computer science, as an original alternative to computer science for control (as more usually in embedded systems), and to classical discrete control systems (as more usually applied to manufacturing). 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.

We have obtained this year the following results:

- At the programming language level, we are continuing the development of a modelling and controller generation language called BZR, which involves DCS in its compilation. BZR is designed and developed in cooperation with the Pop Art and VerTecs (INRIA Rennes) teams [40].
- We have developed a technique for designing reconfiguration controllers in the Fractal component-based framework, where discrete control loops automatically enforce safety properties on the interactions between components, concerning, e.g., mutual exclusions, forbidden or imposed sequences [29] [48].
- We have integrated BZR with Orccad, a programming environment for real-time control systems, in cooperation with the NeCS and SED teams at INRIA Grenoble [28].
- We are investigating administration loops in virtual machine-based distributed systems [44], and the coordination of such loops, especially in relation with green computing problems. We are starting a new ANR project called Ctrl-Green on this topic in 2012.
- We work on the formal modelling and control of dynamic reconfiguration in FPGA circuits, in cooperation with the DaRT team (INRIA Lille) [43] and the Lab-STICC laboratory in Lorient [42], building upon earlier work related to the MARTE framework.
- In cooperation with GIPSA-Lab and ENSI Tunis, we have adapted the discrete controller synthesis technique to the control of decentralized systems that are composed of several subsystems spread across remote sites [17].
- In cooperation with Orange labs and GIPSA-Lab we are beginning to explore the application of discrete event systems and supervisory control to the domain of Machine to Machine and Internet of Things, with the objective to manage energy aspects; this will start with the CIFRE PhD (U. Grenoble) of Mengxuan Zhao (co-advised with H. Alla, G. Privat).

5.4. System configuration and deployment

Participants: Loris Bouzonnet, Fabienne Boyer, Willy Malvault, Noël de Palma, Vivien Quéma, Jean-Bernard Stefani.
The goal of this work is to study system configuration and software deployment issues in large distributed systems.

System configuration and software deployment in a distributed environment can be greatly aided by the use of a uniform component model to support software assembly, software configuration and deployment, as well as runtime system configuration. We have developed a specialization of the Fractal component model that provides a reference model for heterogeneous software assembly and configuration. In particular, we have shown how this reference model can be used to assemble and configure software architectures built from heterogeneous software packages (e.g. OSGI bundles for Java packages, Debian or RPM packages for Linux modules and applications). The definition of this model, a description of its implementation and its evaluation are documented in Loris Bouzonnet’s PhD thesis [13].

As an alternative to current public cloud infrastructures, which rely on large data centers, we have started the study of a cloud infrastructure based on a peer-to-peer (P2P) overlay network built on gossip-based protocols. More precisely, we have studied how to implement a distributed resource allocation service in a P2P environment maintained by a gossip-based peer-sampling protocol [81]. The resulting system, called Salute, provides for the allocation of application-specific overlays out of an underlying P2P network. By combining several P2P services (including peer-sampling, topology maintenance, and node synchronization), and by partitioning available nodes into free nodes (available for the allocation of new application overlays) and reserve nodes (nodes dedicated to the maintenance of allocated overlays), Salute provides a churn-resistant, completely decentralized cloud infrastructure. In addition, we have shown that Salute can provide its allocation service while maintaining fairness and avoiding starvation. The Salute architecture has been validated through simulations using network traces from different real-world P2P environments. The Salute architecture, algorithms and their validation are documented in Willy Malvault’s PhD thesis [14].

In a cloud computing context the complexity of deploying and configuring non-trivial software architectures is exacerbated. In line with our previous work on architecture-based distributed system management, we have proposed a novel algorithm for configuring component-based distributed applications deployed within several virtual machines in an IaaS environment. The algorithm is completely decentralized, relies on a message queuing middleware and exploits the software architecture descriptions of the applications to deploy and configure, written in an extension of the Fractal Architecture Description Language. A first version of this algorithm, that does not take into account potential failures during the configuration process, has been formally specified in collaboration with Gwen Salaün from the INRIA Vasy team in Grenoble, and presented at IEEE Cloud 2011 [35].

5.5. System support: System support for multicore machines

Participants: Vivien Quéma, Renaud Lachaize, Fabien Gaud, Baptiste Lepers, Sylvain Genevès, Fabien Mottet.

Multicore machines with Non-Uniform Memory Accesses (NUMA) are becoming commodity platforms. Efficiently exploiting their resources remains an open research problem. Most of the body of existing work focuses on increasing locality between computations and memory or I/O resources. This is achieved by allocating data items preferably in local memory nodes, by moving computations close to I/O devices or by moving already allocated memory pages close to the applications which use them most. In all these works, researchers always assume that all processors have equal memory performance. Nevertheless, this assumption is not always valid. In 2011, we have studied the performance achieved by a 16-core NUMA exhibiting an irregular connectivity between processors. Some processors are directly connected to all other processors and access memory nodes with a low latency. Other processors have a lower degree of connectivity and need more hops to access some memory nodes and access memory with a higher latency.

Current operating systems are not aware of such performance characteristics. We have shown that the completion time of applications taken from the PARSEC benchmark suite can vary by up to 15% depending on the processor they are scheduled on. We have thus proposed a new OS scheduler that takes this asymmetry into account in order to make efficient decisions. This scheduler relies on a new metric, called MAPI (number
of main Memory Accesses Per retired Instruction), to predict the impact of processor interconnect asymmetry
on the performance of applications. We have empirically evaluated the relevance of this metric on applications
taken from the PARSEC benchmark suite. We have shown that this metric helps estimating the performance
gap between running an application on a "well-interconnected" processor and on a "weakly-interconnected"
one. Using this metric, the scheduler we proposed makes efficient decisions. More precisely, we have observed
that the scheduler always performed within 3% of the best possible scheduling decision. This work is currently
under submission.

5.6. System support: Protocols for resilient systems

Participants: Vivien Quéma, Alessio Pace.

We have worked on replication protocols for P2P systems. In particular, we have worked on replication in so
called Distributed Hash-Tables (DHTs). DHTs provide a simple high-level put/get abstraction that can be used
to build efficient distributed storage systems. DHTs gained wide popularity in the last decade, fostering a large
amount of interest in the academia, and inspiring the design of key/value distributed storage systems deployed
in production.

DHTs provide a way to deterministically map objects to nodes and allow efficiently retrieving objects in
a distributed fashion. Nodes and objects are logically arranged in a large numeric key-space, according to
a given variant of consistent hashing. Typically, the node in charge of an object is the one whose position
immediately follows the object in the key-space.

To guarantee that objects are reliably stored, DHTs rely on replication. A replication protocol is in charge
of ensuring that, at any time, each object is replicated on a sufficiently large number of replicas. Several
replication strategies have been proposed in the last years. The most efficient ones use predictions about
the availability of nodes to reduce the number of object migrations that need to be performed: objects are
preferably stored on highly available nodes.

We have proposed an alternative replication strategy. Rather than exploiting highly available nodes, we have
designed a protocol that leverages nodes that exhibit regularity in their connection pattern. Roughly speaking,
the strategy consists in replicating each object on a set of nodes that is built in such a way that, with high
probability, at any time, there are always at least k nodes in the set that are available. We have evaluated this
new replication strategy using traces of two real-world systems: eDonkey and Skype. Our evaluation showed
that our regularity-based replication strategy induces a systematically lower network usage than existing state
of the art replication strategies. This work has been published at the International Symposium on Reliable
Distributed Systems, in October 2011.

5.7. System support: End-to-end caching

Participants: Sara Bouchenak, Dàmian Serrano.

Cloud Computing is a paradigm for enabling remote, on-demand access to a virtually infinite set of con-
figurable computing resources. This model aims to provide hardware and software services to customers,
while minimizing human efforts in terms of service installation, configuration and maintenance, for both cloud
provider and cloud customer. A cloud may have the form of an Infrastructure as a Service (IaaS), a Platform
as a Service (PaaS) or a Software as a Service (SaaS). Clouds pose significant challenges to the full elasticity
of clouds, their scalability and their dependability in large scale data management and large scale computing
resources. Caching is a means for high performance and scalability of distributed systems. Although caching
solutions have been successfully studied for individual systems such as database systems or web servers, if
collectively applied, these solutions violate the coherence of cached data. We precisely studied this issue in
e-Caching, a novel end-to-end caching system.

The contribution of this work is twofold: guaranteeing the coherence of cached data at multiple locations
of a distributed system, while improving the overall performance of the system. In collaboration with Marta
Patino and Ricardo Jimenez from Universidad Politecnica de Madrid, we proposed a novel distributed caching
protocol, implemented it and evaluated it with real online services. The experiments showed that e-Caching
was successfully able to improve service performance by two orders of magnitude.
5.8. System support: Performance and dependability benchmarking

Participants: Amit Sangroya, Dàmian Serrano, Sara Bouchenak [correspondant].

MapReduce has become a popular programming model and runtime environment for developing and executing distributed data-intensive and compute-intensive applications. It offers developers a means to transparently handle data partitioning, replication, task scheduling and fault tolerance on a cluster of commodity computers. MapReduce allows a wide range of applications such as log analysis, data mining, Web search engines, scientific computing, bioinformatics, decision support and business intelligence.

There has been a large amount of work on MapReduce towards improving its performance and reliability. Several efforts have explored task scheduling policies in MapReduce, cost-based optimization techniques, replication and partitioning policies. There has also been a considerable interest in extending MapReduce with other fault tolerance models, or with techniques from database systems. However, there has been very little in the way of empiric evaluation for the comparison of the different systems. Most evaluations of these systems have relied on microbenchmarks based on simple MapReduce programs. While microbenchmarks may be useful in targetting specific system features, they are not representative of full distributed applications, and they do not provide multi-user realistic workloads. Furthermore, as far as we know, no studies have investigated dependability benchmarking of MapReduce.

Thus, we provide MapReduce Benchmarking (MRB), a novel MapReduce benchmark suite to enable a thorough analysis of a wide range of features of MapReduce systems. MRB has the following features. First, it enables the empirical evaluation of the performance and dependability of MapReduce systems. This provides a means to analyze the effectiveness of scalability and fault tolerance, two key features of MapReduce. Second, it covers a variety of application domains, workload and faultload characteristics, ranging from compute-oriented to data-oriented applications, batch applications to online real-time applications. While MapReduce frameworks were originally limited to offline batch processing, recent works are exploring the extension of MapReduce beyond batch processing. Moreover, in order to stress MapReduce dependability and performance, the benchmark suite enables different fault injection rates, workloads and concurrency levels. Finally, the benchmark suite is portable and easy to use on a wide range of platforms, covering different MapReduce frameworks and cloud infrastructures. This work has been submitted for publication.

5.9. System support: Self-adaptive Internet services

Participant: Sara Bouchenak.

Although distributed services provide a means for supporting scalable Internet applications, their ad-hoc provisioning and configuration pose a difficult tradeoff between service performance and availability. This is made harder as Internet service workloads tend to be heterogeneous, and vary over time in amount of concurrent clients and in mixture of client interactions. This work proposes an approach for building self-adaptive Internet services through utility-aware capacity planning and provisioning. First, an analytic model is presented to predict Internet service performance, availability and cost. Second, a utility function is defined and a utility-aware capacity planning method is proposed to calculate the optimal service configuration which guarantees SLA performance and availability objectives while minimizing functioning costs. Third, an adaptive control method is proposed to automatically apply the optimal configuration to the Internet service. Finally, the proposed model, capacity planning and control methods are implemented and applied to an online bookstore. Experimental evaluations show that the service successfully self-adapts to both workload mix and workload amount variations, and present significant benefits in terms of performance and availability, with a saving of resources underlying the Internet service.

This work is part of the MyCloud ANR project. It has been described in a chapter of the book titled Performance and Dependability in Service Computing, 2011. There has been an industrial transfer of the MOka software prototype.
5.10. Self-Configuration of distributed system in the Cloud

Participants: Fabienne Boyer, Noël de Palma.

Cloud computing environments fall under three main kinds of offers according to the resources they provide. The Infrastructure as a Service (IaaS) level enables the access to virtualized hardware resources (processing, storage and network). The Software as a Service (SaaS) layer aims at providing the end-users with software applications. The intermediary layer, called Platform as a Service (PaaS), offers a set of tools and runtime environments that allow managing the applications life-cycle. This life-cycle includes the phases related to the design, the development, the deployment of applications, and generally speaking all their management stages (workload, fault tolerance, security). This article focuses on the deployment of distributed applications in virtualized environments such as cloud computing. Such deployments require to generate the virtual images that will be instantiated as virtual machines, thus ensuring the execution of the application on an IaaS platform. Each image embeds technical elements (operating system, middleware pieces) and functional ones (data and applicative software entities). Once it has been instantiated, each virtual machine is subjected to a stage of dynamic settings, which finalizes the global configuration of the distributed application.

On the whole, the deployment solutions currently available do not take into account these different configuration parameters, which are mostly managed by dedicated scripts. Moreover these solutions are not able to automate the images generation, their instantiation as virtual machines and their configuration independently from the kind of distributed application to be deployed. For instance, Google App Engine solution only deals with Web services organized into precisely defined tiers. In our opinion, the absence of general solutions results essentially from a lack of formalism for describing the distributed application architecture with its configuration constraints in a virtualized infrastructure such as cloud computing. Pour work focused on a general solution, for Virtual Applications Management Platform, that automates the deployment of any distributed applications in the cloud. The suggested approach is architectural, meaning that it is based upon an explicit representation of the applications’ distributed architecture. We offer, on the one hand, a formalism for describing an application as a set of interconnected virtual machines and, on the other hand, an engine for interpreting this formalism and automating the application deployment on an IaaS platform. Specifically, we study three contributions:

- A formalism that offers a global view of the application to be deployed in terms of components with the associated configuration- and interconnection constraints and with their distribution within virtual machines. This formalism extends OVF language, dedicated to virtual machines description, with an architecture description language (ADL) that allows describing a distributed application software architecture;
- A deployment engine, i.e. a runtime support able to deploy automatically an application described with this formalism. This engine is based on a decentralized protocol for self-configuring the instantiated virtual machines. In our opinion it can ease the scalability of the dynamic configuration stage;
- A performance evaluation of the proposed solution on an industrial IaaS platform.

We published in this context two journal articles (TPDS [26] and TAAS [18]) and three conference papers (Cloud11 [35], UCC11 [39] and SAC12).

5.11. Virtual Machine

Participants: Olivier Gruber, Fabienne Boyer, Damien Pous, Ludovic Demontes, Clément Deschamps.

A core aspect of the Synergy virtual machine is its ability to reconfigure component-based applications at execution time. We have focused on the reconfiguration protocol with the intent of verifying and proving its robustness.
In a first step, we have formalized and verified that any correct and complex reconfiguration through our reconfiguration protocol can be processed as a sequence of elementary reconfiguration operations and always results in a component assembly that is architecturally consistent. This aspect has been verified using model-checking techniques. This work has been done in collaboration with Gwen Salaün from the VASY team (Inria Rhône-Alpes). It lead to a publication in the Formal Method (FM’11) conference [30].

In a second step, we have considered software failures that may occur during a reconfiguration. Although the protocol is trusted code, it invokes components to reconfigure them, thereby executing unsafe code that may fail. This work with Damien Pous produced a high-level formalisation of our reconfiguration protocol and a completely certified modelisation of these algorithms in Coq [50]. This work resulted in a submitted publication.

Finally, we have also investigated the control of complex reconfiguration through using discrete synchronous control techniques with Eric Rutten and Gwenael Delaval [44].
5. New Results

5.1. Flexible Radio Node

Participants: Florin Hutu, Tanguy Risset, Jacques Verdier, Guillaume Villemaud, Cédric Levy-Bencheton.

This section summarizes the early results obtained from the research axis flexible radio nodes.

In [41], [75], a candidate architecture for LTE-Advanced receiver is proposed. Based on the combination of MIMO techniques and flexible spectrum access, LTE-Advanced terminals will require the increasing of the analog front-end complexity. To reduce the complexity of the analog front-end, an innovative architecture based on the merge between the double IQ and the code multiplexing structures is proposed. Simulation and measurement results show that, in a Gaussian case, the bit error rate is similar when using the proposed architecture and the state of the art front-end stack-up structure. A complexity evaluation study reveals significantly reduced power consumption of the proposed single front-end architecture.

The current generation of mobile terminals can communicate on multiple modes using several antennas. However, their energy consumption remains a critical parameter. In [58], [74], we explore the combination of multiple communication modes and MIMO as a possible way to reduce the energy consumption of both the terminals and the network. We propose a realistic energy model for the PHY layer of a MIMO and multi-mode terminal, taking into account the MAC layer behaviour. We show that the combination of MIMO and multi-mode provides a solution to reduce global energy consumption.

Software means programmable. Hence software defined radio means that the radio should now be programmable. We know what computer programming means, and we agree, up to a certain level, on how it should be done. But do we know what programming a radio means? Several questions are still open: what will an SDR platform look like in ten years? Will there exist software radio code? What will be the technical challenges and commercial issues behind this code? Programming is more precise than configuring or tuning, it implies a much greater level of freedom for the programmer. But it also means much cheaper implementations in many cases and in particular a re-use of the same hardware for different protocols (i.e. with different programs). This is, to our point of view [76], the main difficulty of software radio programming: reconfiguration and in particular dynamic reconfiguration. Dynamic (i.e. very fast) reconfiguration is now mandatory because some protocols, 3GPP-LTE (Third Generation Partnership Program Long Term Evolution) for instance, propose channel adapting for each frame, requiring a setting of the channel estimation parameter in a few milliseconds.

5.2. Agile radio resource sharing

Participants: Jean-Marie Gorce, Claire Goursaud, Katia Jaffrès-Runser, Nikolaï Lebedev, Guillaume Villemaud, Paul Ferrand, Philippe Mary.

This section presents our recent results concerning the realistic modeling of wireless links to develop realistic models and efficient simulations. This work include theoretical developments like symbol error outage modeling, but also some applications in the context of LTE multi-cells association, or opportunistic relaying in the context of wireless sensor networks. Other contributions about resource sharing are presented in next sections below, in the section 'network optimization' and the section 'network coding'.

In [28], we addressed the problem of finding a tractable expression for the symbol error outage (SEO) in flat Nakagami-m fading and shadowing channels. We deal with M-ary phase shift keying (M-PSK) and quadrature amplitude modulation (M-QAM) which extends our previous results on BPSK signaling. We propose a new tight approximation of the symbol error probability (SEP) holding for M-PSK and M-QAM signals which is accurate over all signal to noise ratios (SNRs) of interest. We derive a new generic expression for the inverse SEP which facilitates the derivation of a tight approximation of the SEO in a lognormal shadowing environment.
In [44], we consider on-body BAN nodes transmitting information towards a common sink, in a star topology (Body Area Networks (BAN) offer amazing perspectives to instrument and support humans in many aspects of their lives). While this setup is usual in wireless networks, the high instability of the BAN radio channel and the proximity of the body make classical communication protocols inefficient. These networks are further constrained by the low transmission power required for both battery life and health concerns. Opportunistic cooperation techniques are of great interest in such environment to ensure reliable communications. In previous works, we studied simple opportunistic relaying schemes under independent BAN links, using a packet error outage criterion. In this paper, we introduce a more realistic case where shadowing variations around the body are now assumed strongly correlated. Generally speaking, there is a lack of definitive measurements and models for the shadowing correlation in multi-hop networks, while it can play a crucial role at the higher layers. Based on the measurement and simulation results of the French BANET project, we use the BAN context as an illustrative example to exhibit how shadowing correlations have a strong impact on relaying approaches performance.

Opportunistic networking aims at exploiting sporadic radio links to improve the connectivity of a multi-hop network and to foster data transmissions. Broadcast nature of the wireless channel is an important feature that can be exploited to improve transmissions by using several potential receivers. Opportunistic relaying is thus the first brick for opportunistic networking. However, the advantage of opportunistic relaying may be balanced by energy increase related to having simultaneous active receivers. In [32], we proposed a thorough analysis of opportunistic relaying efficiency under different realistic radio channel conditions. The study aims at finding the best trade-off between two objectives: energy and latency minimizations, under a hard reliability constraint. We derive an optimal bound, namely, the Pareto front of the related optimization problem, which offers a good insight into the benefits of opportunistic routings compared with classical multi-hop routing schemes. Meanwhile, the lower bound provides a framework to optimize the parameters in physical layer, MAC layer and routing layer from the viewpoint of cross layer during the design or planning phase of a network.

This work has been extended in In [70] for relay channels. The gain induced by using relay channels in a linear network under both a capacity constraint and a realistic energy model is evaluated. We express a general model based on a convex optimization problem, allowing us to use numerical tools to obtain similar results for outer and inner bounds to the capacity of the full and half duplex relay channel. We then further the 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 for 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. This work will be published in 2012 in IEEE WCNC.

In [82] we extended this formalism derived for a linear network to a more general case: the problem of deriving fundamental trade-off bounds for wireless ad hoc networks when multiple performance criteria are of interest. It proposes a MultiObjective (MO) performance evaluation framework composed of a broadcast and interference-limited network model, a steady state performance metric derivation inspired by a discrete Markov chain formalism and formulates the associated MO optimization problem. Pareto optimal performance bounds between end-to-end delay and energy for a capacity-achieving network are given for the 1-relay and 2-relay networks and assessed through simulations.

5.3. Autonomous wireless networking

Participants: Isabelle Augé-Blum, Bernard Tourancheau, Fabrice Valois, Ibrahim Amadou, Cédric Chauvenet, Quentin Lampin, Alexandre Mouradian, Bilel Romdhani.

Designing protocols for large scale wireless sensors networks is a challenging issue, if realistic environments are considered. Finding a trade-off between energy consumption and delay, or capacity, is difficult. The most promising ideas rely on zero-protocol approaches and on virtual coordinates use. the special case of VANETs is presented in the next section.
In [64], we focus on Wireless Sensor Networks (WSNs) in a more realistic case than classical studies and previous works: we consider wireless sensor nodes having different transmission ranges according to the environment and/or to the wireless chipset. The main consequence of this heterogeneity is the existence of asymmetric links. Such links in a WSN degrade the performance of most protocols which have not been designed to support this heterogeneity and to deal with asymmetric links: so, mainly, these links are pruned. Under this assumption, we propose a routing protocol for data collection from sensors nodes to the sink node in heterogeneous WSNs. Our proposal detects and takes benefit from asymmetric links caused by this heterogeneity. Our proposal, denoted MURA, (1) provides a high delivery ratio, (2) reduces the number of duplicated packets and (3) reduces the number of hop counts by exploiting the asymmetric links.

Due to the efficiency and scalability of greedy routing in WSNs and the financial cost of GPS chips, Virtual Coordinate Systems (VCSs) for WSNs have been proposed. A category of VCSs is based on the hop-count from the sink, this scheme leads to many nodes having the same coordinate. The main advantage of this system is that the hops number of a packet from a source to the sink is known. Nevertheless, it does not allow to differentiate the nodes with the same hop-count. We propose in [87] a novel hop-count-based VCS which aims at classifying the nodes having the same hop-count depending on their connectivity and at differentiating nodes in a 2-hop neighborhood. Those properties make the coordinates, which also can be viewed as a local identifier, a very powerful metric which can be used in WSNs mechanisms.

Duty-cycled medium access protocols allow for long lasting autonomous networks by periodically putting nodes to sleep. However, this life expectancy improvement comes at the cost of a lesser network capacity and a poor adaptability to bursty traffic loads. Indeed, existing contention algorithms do not provide efficient algorithms to dynamically elect multiple senders per wake-up periods. In [84], the medium is divided in several logical channels (e.g., obtained by a time/frequency division of the communication medium) and we propose to allocate them dynamically among senders. For this purpose, we propose a joint contention/scheduling algorithm, named Extended Slot Selection (ESS), that schedules multiple sender/receiver pairs to available logical channels.

Energy-efficient communication protocol is a primary design goal for Wireless Sensor Networks (WSNs). Many efforts have been done to save energy: MAC with duty cycle, energy-aware routing protocols, data aggregation schemes, etc. Recently, beacon-less strategies have emerged as new direction to improve considerably the WSN lifetime. However, the main contributions are not suitable to real radio environments because of hole avoiding strategies based on either planarization or explicit neighbor solicitations. We propose in [34] PFMAC (Pizza-Forwarding Medium Access Control), which combines beacon-less geo-routing and energy efficient MAC protocol via a cross-layer design to save energy with higher reliability. PFMAC supports radio interferences, asymmetric radio links, etc. PFMAC supports a greedy forwarding strategy and, a reactive and optimized neighborhood discovery at 2-hop to deal with holes. Intensive simulations are proposed to highlight the behavior and the performance of PFMAC compared to BOSS over BMAC.

To provide for reliability in Wireless Sensor Networks (WSNs), Medium Access Control (MAC) protocols must be adapted by mechanisms taking cross-layer approaches into account. In [51], [52], we describe AreaCast which is designed for enhancing reliability in WSNs. AreaCast is a MAC layer mechanism independent of the routing layer, but uses only local topological and routing information to provide a communication by area instead of a traditional, node-to-node communication (i.e., unicast). In AreaCast, a source node addresses a set of nodes: an explicit relay node chosen as the next hop by a given routing protocol, and k other implicit relay nodes. The neighboring nodes select themselves as implicit relays according to their location from the explicit relay node. This mechanism uses overhearing to take advantage of the inherent broadcast nature of wireless communications. Without changing the routing protocol, AreaCast is able to dynamically avoid a byzantine node or an unstable link, allowing to benefit from the inherent topological redundancy of densely deployed sensor networks. Simulation results show that AreaCast significantly improves the packet delivery rate while having a good reliability-energy consumption trade-off.

Improving the network lifetime is an important design criterion for wireless sensor networks especially if we want to use standard solution like IPv6. In [38], we propose a novel approach which applies source-coding on addresses in heterogeneous IPv6 Cluster-based wireless sensor network. We formulate the problem
of maximizing the network lifetime when Slepian-wolf coding is applied on addresses in network composed of line-powered and battery-powered sensors. The numerical results show that a significant network lifetime improvement can be achieved (about 25% in typical scenario). In [36], we investigate the sinks mobility in IPv6-based wireless sensors networks and specially in the new IETF proposed protocol RPL (Routing Protocol for Low power and Lossy Networks). We also show that even the mobility of sinks is not an explicit design criteria, the use of mobile sinks improves the network lifetime.

5.4. Wireless networking in VANETs

Participants: Marco Fiore, Sandesh Uppoor.

VANETS (Vehicular Ad hoc Networks) represents a challenging context for designing new protocols as it offers new challenges related to the high dynamicty of the network. In cooperation with external researchers, we derived recent results on mobility modeling and data dissemination in VANETS. This work is a part of the work on ‘Autonomous wireless networking’, but dedicated specially for VANETS.

Simulation is the tool of choice for the large-scale performance evaluation of upcoming telecommunication networking paradigms that involve users aboard vehicles, such as next-generation cellular networks for vehicular access, pure vehicular ad hoc networks, and opportunistic disruption-tolerant networks. The single most distinguishing feature of vehicular networks simulation lies in the mobility of users, which is the result of the interaction of complex macroscopic and microscopic dynamics. Notwithstanding the improvements that vehicular mobility modeling has undergone during the past few years, no car traffic trace is available today that captures both macroscopic and microscopic behaviors of drivers over a large urban region, and does so with the level of detail required for networking research. In [66], we present a realistic synthetic dataset of the car traffic over a typical 24 hours in a 400-km$^2$ region around the city of Koln, in Germany. We outline how our mobility description improves today’s existing traces and show the potential impact that a comprehensive representation of vehicular mobility can have one the evaluation of networking technologies.

In [21], [30], we investigate data dissemination in vehicular networks. Content downloading in vehicular networks is a topic of increasing interest: services based upon it are expected to be hugely popular and investments are planned for wireless roadside infrastructure to support it. We focus on a content downloading system leveraging both infrastructure-to-vehicle and vehicle-to-vehicle communication. With the goal to maximize the system throughput, we formulate a max-flow problem that accounts for several practical aspects, including channel contention and the data transfer paradigm. Through our study, we identify the factors that have the largest impact on the performance and derive guidelines for the design of the vehicular network and of the roadside infrastructure supporting it.

In [45] We address cooperative caching in wireless networks, where the nodes may be mobile and exchange information in a peer-to-peer fashion. We consider both cases of nodes with large- and small-sized caches. For large-sized caches, we devise a strategy where nodes, independent of each other, decide whether to cache some content and for how long. In the case of small-sized caches, we aim to design a content replacement strategy that allows nodes to successfully store newly received information while maintaining the good performance of the content distribution system. Under both conditions, each node takes decisions according to its perception of what nearby users may store in their caches and with the aim of differentiating its own cache content from the other nodes’. The result is the creation of content diversity within the nodes neighborhood so that a requesting user likely finds the desired information nearby. We simulate our caching algorithms in different ad hoc network scenarios and compare them with other caching schemes, showing that our solution succeeds in creating the desired content diversity, thus leading to a resource-efficient information access.

Performance and reliability of content access in mobile networks is conditioned by the number and location of content replicas deployed at the network nodes. In [27], we design a practical, distributed solution to content replication that is suitable for dynamic environments and achieves load balancing. Simulation results show that our mechanism, which uses local measurements only, approximates well an optimal solution while being robust against network and demand dynamics. Also, our scheme outperforms alternative approaches in terms of both content access delay and access congestion.
5.5. Optimization in wireless networks


In the context of the common lab between Inria and Alcatel Lucent Bell Labs and the ANR Ecoscells project, we work on optimizing wireless networks performance. In one side, we work on distributed algorithms for optimal resource allocation and/or mobile-BS association. On the other side, we work on mesh wireless networks optimization.

Multi-cell processing, also called Coordinated Multiple Point (CoMP), is a promising distributed technique that uses neighbor cells’ antennas [48]. It is expected to be the part of next generation cellular standards such as LTE-A. Small cell networks in dense urban environments are limited by interferences and CoMP can strongly take advantage of this fact to improve cell-edge users’ throughput. The present study introduces a distributed criterion for mobiles to select their optimal set of Base Stations (BS) to perform CoMP, and evaluates the impact of this association on the fairness and the total cell throughput. For that, we use a known theoretical expression for the capacity outage probability of CoMP under Rayleigh fading and evaluate the goodputs of antennas associations. The proposed criterion is used in combination with fair resource allocation to perform a joint double-objective optimization of fairness and efficiency. In [48], [91], we provide the analysis of the downlink Coordinated Multiple Point (CoMP) used in conjunction with the basic MIMO. The CoMP is the joint multi-cell transmission from several BS to mobiles, coupled here to an open-loop MIMO technique that does not require the perfect channel state knowledge. We show by simulation, that even for $4 \times 4$ MIMO transmission, the CoMP can improve the spectral efficiency for some mobiles, depending on capacity outage requirements.

In [33], we considered downlink transmission in cellular networks where we target to reduce the energy consumption by switching off some base stations by such a way that the distribution of SINR remains unchanged. This is a mean of green networking in cellular networks in downlink consideration. This paper analyzes for line and plane cases, the gain in power consumption obtained after switching off base stations. By computations we observe that the more the operational cost the more the gain in power consumption.

In [47], we propose an autonomous radio resource allocation and optimization scheme that chooses the transmit power and precoding vector among codebooks for multiple antennas transmitters to improve spectral and power efficiency and provide user fairness. Network self-optimization is an essential feature for supporting the cell densification in future wireless cellular systems. The proposed self-optimization is inspired by Gibbs sampler. We show that it can be implemented in a distributed manner and nevertheless achieves system-wide optimization which improves network throughput, power utilization efficiency, and overall service fairness. In addition, we extend the work and include power pricing to parametrize and enhance energy efficiency further. Simulation results show that the proposed scheme can outperform today’s default modes of operation in network throughput, energy efficiency, and user fairness.

In [55], we focused on broadband wireless networks based on OFDMA resource management, such as LTE systems. We have investigated two optimization problems, one concerning a backhauling mesh infrastructure while the other is the allocation of modulation and coding, subcarriers and power to users in LTE. Considering a realistic SINR model of the physical layer with a fine tuned power control at each node, a linear programing model using column generation has been developed for computing power efficient schedules with high network capacity for wireless mesh backhauling networks. Correlation between capacity and energy consumption have been analyzed as well as the impact of physical layer parameters - SINR threshold and path-loss exponent. With these models, 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. Finally, simulation results show that compared to classic reuse schemes the proposed approach is able to pack more users into the same bandwidth, decreasing the probability of user outage.

In [62], we focus on broadband wireless mesh networks like 3GPP LTE-Advanced. This technology is a key enabler for next generation cellular networks which are about to increase by an order of magnitude...
the capacity provided to users. Such an objective needs a significative densification of cells which requires an efficient backhauling infrastructure. In many urban areas as well as under-developed countries, wireless mesh networking is the only available solution. Besides, economical and environmental concerns require that the energy expenditure of such infrastructure is optimized. We propose a multi-objective analysis of the correlation between capacity and energy consumption of LTE-like wireless mesh networks. We provide a linear programing modeling using column generation for an efficient computation of the Pareto front between these objectives. Based on this model, we observe that there is actually no significant capacity against energy trade-off.

In [63], broadband wireless mesh networks based on OFDMA resource management are studied considering a realistic SINR model of the physical layer with a fine tuned power control at each node. A linear programing model using column generation leads to compute power efficient schedules with high network capacity. 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.

5.6. Network coding in WSN

Participants: Jean-Marie Gorce, Cédric Lauradoux, Marco Fiore, Claire Goursaud, Marine Minier, Anya Apavatjrut, Yuanyuan Zhang, Wassim Znaidi.

Network coding associated with Fountain codes is a very efficient approach to increase the throughput of multi-hop networks. However severe outcomes are still expected, especially to develop robust and energy efficient approaches for transmitting data over a large scale networks. Network coding is also very promising for security issues as presented below.

Diversity is a powerful means to increase the transmission performance of wireless communications. For the case of fountain codes relaying, it has been shown previously that introducing diversity is also beneficial since it counteracts transmission losses on the channel. Instead of simply hop-by-hop forwarding information, each sensor node diversifies the information flow using XOR combinations of stored packets. This approach has shown to be efficient for random linear fountain codes. However, random linear codes exhibit high decoding complexity. In [19], we propose diversity increased relaying strategies for the more realistic and lower complexity Luby Transform code in a linear network. Results are provided herein for a linear network assuming uniform imperfect channel states.

In [29], the exact probability that a receiver obtains \( N \) linearly independent packets among \( K \) over \( N \) received packets is computed, when the sender/s use/s random linear network coding over a Galois Field of size \( q \). Such condition maps to the receiver's capability to decode the original information, and its mathematical characterization helps to design the coding so to guarantee the correctness of the transmission. The proposed formulation represents an improvement over the current upper bound for the decoding probability, and provides theoretical grounding to simulative results in the literature.

In [35], we focus on the proper use of fountain codes for the transmission of sporadic data in a wireless sensor network (WSN). Fountain codes offer great perspectives for the self-organization of WSNs: they self adapt to the channel error rate without any control data. When deploying fountain codes on a WSN, two problems arise. First, the size of the data transmitted by a sensor is small in comparison to the size considered traditionally with fountain codes. Second, the communications are done in an hop-by-hop fashion. It implies that the destination of the data can not acknowledge instantaneously its reception to the source. Therefore, the transmissions of useless packets for the destination can not be prevented. The flooding traffic has been evaluated as well through realistic simulations for three different relaying strategies where packets are lost due to both small scale fading and collisions for an unslotted IEEE 802.15.4 medium access layer.

Network coding has attracted the attention of many researchers in security and cryptography. We have investigated several aspects of network coding security. In [20], we propose efficient solutions to thwart pollution attacks in which an adversary injects false information into data flow. This work was further expanded
in [54] to find rational strategy to minimize the energy cost and the impact of the attack. We also came to the conclusion that dealing with pollution attacks was not enough as long as the acknowledgment messages are not also protected. The risk is to suffer from a flooding attack. This goes beyond the capabilities of cryptographic solutions and we investigate the security capabilities of multipath acknowledgment in [67].

5.7. Security


Security is an important issue for wireless networks, especially for wireless sensor networks facing an amizing increase of the number of nodes. We review in this section all contributions related to the security issue, some of them being strongly related with the PHY layer or the networking protocols. As it can be seen below, some results are strongly connected to the models and protocols derived in the other sections.

In [59], we provide the first independent analysis of the (2nd-round tweaked) 256-bit version of the SHA3 candidate SHAvite-3. By leveraging recently introduced cryptanalysis tools such as rebound attack or Super-Sbox cryptanalysis, we are able to derive chosen-related-salt distinguishing attacks on the compression function on up to 8 rounds (12 rounds in total) and free-start collisions on up to 7 rounds. In particular, our best results are obtained by carefully controlling the differences in the key schedule of the internal cipher. Most of our results have been implemented and verified experimentally.

In [50], we study a class of insider attacks called the terrorist fraud. This is a relay attack against distance bounding protocols where the prover conspires with an adversary to misrepresent the distance between himself and the verifier. In ideal situations, the adversary does not gain any knowledge about the prover’s long-term secret. This makes designing a distance bounding protocol resistant to such fraud tricky: the secrets of an honest prover must be protected, while those of a dishonest one should be disclosed as an incentive not to cheat. We demonstrate that using a secret-sharing scheme, possibly based on threshold cryptography, is well suited for thwarting the terrorist fraud. Although such an idea has been around since the work of Bussard and Bagga, this is the first time that secret-sharing and terrorist fraud have been systematically studied altogether.

In [40], we deal with the problem of radio jamming. Jamming is a major threat against wireless communications. In this paper, we evaluate the effect of jamming on an UWB link employing a PPM non-coherent receiver. We optimize the jammer parameters that are the central frequency and the bandwidth based on the metric of the signal-to-jamming ratio. The optimization depends on different system parameters such as the channel model and the integration time of the receiver.

In [23], we focus on the resiliency of wireless sensor network routing protocols against selective forwarding attacks by compromised nodes. Informally, resiliency should be understood as the capacity of the routing protocol to endure and mitigate the presence of a certain number of compromised nodes seeking to disturb the routing process. To provide for security when nodes may be compromised, cryptographic solutions must be completed by algorithmic solutions considering “beyond cryptography” approaches. After discussing the shortcomings of existing routing protocols against packet-dropping malicious nodes we describe some protocol behaviors enhancing routing resiliency under several combined routing attacks. We propose in this paper the behaviors enhancing the resiliency of routing protocols under several combined routing attacks.

5.8. Network simulation tools

Several works in 2011 have been using simulation results. Nevertheless, Swing members are strongly working on improving network simulation frameworks to provide realistic simulations. Several contributions to the simulation tools wiplan ans wsnets have been proposed.

Some contributions to WSNets concern BAN environments implementation [44] and network coding features [19], [81]. Different protocols have been also implemented for wireless sensor networks [34], [84], specifically in the context of our collaboration with Orange Labs, Grenoble.
The wiplan simulator has been developed at CITI for several years. It is based on a frequency domain ParFlow (MR-FDPF) implementation that represents a unique finite elements based method for estimating the radio propagation in complex environments. In the context of heterogeneous networks, femtocells are very promising. In order to properly simulate their behavior and their impact on the macrocell layer, it is necessary to be able to simulate the radio coverage of femtocells. Hence ParFlow is a possible deterministic model that can be used for such simulation. In \[42\], two implementations of ParFlow are presented: time domain and frequency domain. The performance are compared and the advantages/drawbacks of each model are investigated.

In \[56\] we propose to use finite difference propagation methods to evaluate the wide band properties of the fast fading. For this purpose we adapted the MR-FDPF propagation model to simulate large bandwidth by combining numerous narrow band simulations. The results are compared with a channel sounder measurement campaign covering a bandwidth of up to 70 MHz. It is verified that fading characteristics in wireless channels varies with frequency and the MR-FDPF method is capable for simulating this variation of fadings for wide band systems.

In \[56\], a new approach is proposed allowing extracting the fading statistics for indoor radio channels based on the electric field strength predicted with the MR-FDPF method. The performance of the proposed approach is verified both by simulations and measurements.

In \[65\], we propose a new hybrid modeling method for indoor-to-outdoor radio coverage prediction. The proposed method is a combination of a ray-optical channel modeling approach and the frequency domain ParFlow method. While the former is widely used for modeling outdoor propagation environments, the latter is computationally efficient and accurate for modeling indoor environments.

In \[90\], we propose to use finite difference propagation methods to evaluate the wide band properties of the fast fading. For this purpose we adapted the MR-FDPF propagation model to simulate large bandwidth by combining numerous narrow band simulations. The results are compared with a channel sounder measurement campaign covering a bandwidth of up to 70 MHz. It is verified that fading characteristics in wireless channels varies with frequency and the MR-FDPF method is capable for simulating this variation of fading for wide band systems.
ARTIS Project-Team

6. New Results

6.1. Lighting and Rendering

Participants: Mahdi Bagher, Laurent Belcour, Georges-Pierre Bonneau, Eric Bruneton, Cyril Crassin, Jean-Dominique Gascuel, Olivier Hoel, Nicolas Holzschuch, Fabrice Neyret, Cyril Soler, Fabrice Neyret, Charles de Rousiers, Cyril Soler.

6.1.1. 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. In [14] we presented a survey of these methods. We have started 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.

6.1.2. Frequency-Based Kernel Estimation for Progressive Photon Mapping

Participants: Laurent Belcour, Cyril Soler.

We have developed an extension to Hachisuka et al.’s Progressive Photon Mapping (or PPM) algorithm [32] in which we estimate the radius of the density estimation kernels using frequency analysis of light transport [29]. We predict the local radiance frequency at the surface of objects using a Gaussian approximation, and use it to drive the size of the density estimation kernels, in order to accelerate convergence (see Figure 3). The key is to add frequency information to a small proportion of photons: frequency photons. In addition to contributing to the density estimation, they will provide frequency information. This work has been published in [20].

6.1.3. Efficiently Visualizing Massive Tetrahedral Meshes with Topology Preservation

Participant: Georges-Pierre Bonneau.
Figure 3. In this figure we compare against progressive photon mapping with our algorithm for the convergence of an indirectly lit part of the scene. In the closeup, we show that our algorithm produces a lower varying estimate at an earlier stage of its execution. The images were produced using 100,000 photons per pass and 25% of frequency photons to make timing comparable.

Figure 4. Left: Variable resolution visualization of a volume mesh with multiple linear features. The topology of the substructures is guaranteed to be preserved. Right: Snapshot of the multiresolution visualization tool to explore simulation data with embedded structures on a desktop PC.
This work is the result of a collaboration with S. Hahmann from the EVASION team-project and Prof. Hans Hagen partly done during a sabbatical of G.-P. Bonneau in the University of Kaiserslautern, Germany. Interdisciplinary efforts in modeling and simulating phenomena have led to complex multi-physics models involving different physical properties and materials in the same system. Within a 3d domain, substructures of lower dimensions appear at the interface between different materials. Correspondingly, an unstructured tetrahedral mesh used for such a simulation includes 2d and 1d substructures embedded in the vertices, edges and faces of the mesh. The simplification of such tetrahedral meshes must preserve (1) the geometry and the topology of the 3d domain, (2) the simulated data and (3) the geometry and topology of the embedded substructures. This work focuses on the preservation of the topology of 1d and 2d substructures embedded in an unstructured tetrahedral mesh, during edge collapse simplification. We derive a robust algorithm, based on combinatorial topology results, in order to determine if an edge can be collapsed without changing the topology of both the mesh and all embedded substructures. Based on this algorithm we have developed a system for simplifying scientific datasets defined on irregular tetrahedral meshes with substructures, illustrated in Figure 4. We presented and demonstrated the power of our system with real world scientific datasets from electromagnetism simulations in the Springer book chapter [27].

6.1.4. Real-Time Rough Refraction

Participants: Nicolas Holzschuch, Charles de Rousiers.

![Figure 5. Left: ground truth with total internal reflection. Right: the result with our Real-time Rough Refraction technique. While images have some differences, the result remains plausible.](image)

We have developed an algorithm to render objects of transparent materials with rough surfaces in real-time, under distant illumination. 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 in I3D paper [24] the quality of our approximation by comparison against stochastic raytracing. This work is illustrated in Figure 5.
6.1.5. Interactive Indirect Illumination Using Voxel Cone Tracing  
**Participants:** Cyril Crassin, Fabrice Neyret.

![Figure 6. Real-time indirect illumination (25-70 fps on a GTX480): We rely on a voxel-based cone tracing to ensure efficient integration of 2-bounce illumination and support diffuse and glossy materials on complex scenes. (Right scene courtesy of G. M. Leal Llaguno)](image-url)

Indirect illumination is an important element for realistic image synthesis, but its computation is expensive and highly dependent on the complexity of the scene and of the BRDF of the involved surfaces. While off-line computation and pre-baking can be acceptable for some cases, many applications (games, simulators, etc.) require real-time or interactive approaches to evaluate indirect illumination. We present in the Pacific Graphics paper [16] a novel algorithm to compute indirect lighting in real-time that avoids costly precomputation steps and is not restricted to low-frequency illumination. An illustration is given in Figure 6. It is based on a hierarchical voxel octree representation generated and updated on the fly from a regular scene mesh coupled with an approximate voxel cone tracing that allows for a fast estimation of the visibility and incoming energy. Our approach can manage two light bounces for both Lambertian and glossy materials at interactive framerates (25-70fps). It exhibits an almost scene-independent performance and can handle complex scenes with dynamic content thanks to an interactive octree-voxelization scheme. In addition, we demonstrate that our voxel cone tracing can be used to efficiently estimate Ambient Occlusion. A primer of this work has been published as a poster (Best Poster Award [22]). Insights of the method were given in the Siggraph Talk 2011 [23].

The publication [22] has received the Best Poster Award at I3D’2011.

6.1.6. Fast multi-resolution shading of acquired reflectance using bandwidth prediction  
**Participants:** Mahdi Bagher, Laurent Belcour, Nicolas Holzschuch, Cyril Soler.

Shading complex materials such as acquired reflectances in multi-light environments is computationally expensive. Estimating the shading integral involves stochastic sampling of 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. Ignoring visibility, adaptively distributing computational budget across the pixels for shading is already a challenging problem. In the paper [28] we present a systematic approach to accelerate shading, by rapidly predicting the approximate spatial and angular variation in the local light field arriving at each pixel. Our estimation of variation is in the form of local bandwidth, and accounts for combinations of a variety of factors: the reflectance at the pixel, the nature of the illumination, the local geometry and the camera position relative to the geometry and lighting. An illustration is given in Figure 7.

The speed-up, using our method, is from a combination of two factors. First, rather than shade every pixel, we use this predicted variation to direct computational effort towards regions of the image with high local variation. Second, we use the predicted variance of the shading integrals, to cleverly distribute a fixed total budget of shading samples across the pixels. For example, reflection off specular objects is estimated using fewer samples than off diffuse objects.
Figure 7. The technique developed by Mahdi Bagher allows to predict in real time the local bandwidth of the image obtained by shading a measured material with all frequency distant illumination (See inset colored top-right image). This information allows a drastic economy of samples in the computation of the integrals that are needed to produce an accurate image. In particular this allows to perform multi-sampling anti-aliasing in a deferred shading pipeline with much less image-space samples than the brute-force solution.
6.2. Expressive Rendering and Visualization

Participants: Pierre Bénard, Georges-Pierre Bonneau, Alexandre Coninx, Joëlle Thollot.

6.2.1. Temporal Coherence for Stylized Animation

Participants: Pierre Bénard, Joëlle Thollot.

Figure 8. In our state-of-the-art report we review and carefully compare Temporal Coherence techniques for stylized animations.

Non-photorealistic rendering (NPR) algorithms allow the creation of images in a variety of styles, ranging from line drawing and pen-and-ink to oil painting and watercolor. These algorithms provide greater flexibility, control and automation over traditional drawing and painting. Despite significant progress over the past 15 years, the application of NPR to the generation of stylized animations remains an active area of research. The main challenge of computer generated stylized animations is to reproduce the look of traditional drawings and paintings while minimizing distracting flickering and sliding artifacts present in hand-drawn animations. These goals are inherently conflicting and any attempt to address the temporal coherence of stylized animations is a trade-off. We have published the state-of-the-art report [15] motivated by the growing number of methods proposed in recent years and the need for a comprehensive analysis of the trade-offs they propose. We formalize the problem of temporal coherence in terms of goals and compare existing methods accordingly. We propose an analysis for both line and region stylization methods and discuss initial steps toward their perceptual evaluation. The goal of our report is to help uninformed readers to choose the method that best suits their needs, as well as motivate further research to address the limitations of existing methods.

6.2.2. Visualization of data with uncertainty using perceptually guided procedural noise

Participants: Alexandre Coninx, Georges-Pierre Bonneau.

This work is the result of a collaboration with EdF R&D and Jacques Droulez, Director of Research at CNRS in Collège de France. In his PhD work, Alexandre Coninx has introduced a new method to visualize uncertain scalar data fields by combining color scale visualization techniques with animated, perceptually adapted Perlin noise. The parameters of the Perlin noise are controlled by the uncertainty information to produce animated patterns showing local data value and quality, as illustrated in Figure 9. In order to precisely control the perception of the noise patterns, we perform a psychophysical evaluation of contrast sensitivity thresholds for a set of Perlin noise stimuli. We validate and extend this evaluation using an existing computational model. This allows us to predict the perception of the uncertainty noise patterns for arbitrary choices of parameters. We demonstrate and discuss the efficiency and the benefits of our method with various settings, color maps and data sets. This work has been published at APGV’2011 [21].
Figure 9. Left: classical colormap visualization of scalar data without uncertainty. Right: in our technique, we perturb the input of the colormap using a perceptually guided procedural noise, scaled by the uncertainty of the data. The data and its uncertainty can be visualized in the same image.

6.3. Modeling and Animation

**Participants:** Georges-Pierre Bonneau, Alexandre Derouet-Jourdan, Nicolas Holzschuch, Nassim Jibai, Cyril Soler, Joëlle Thollot.

6.3.1. Multiscale Feature-Preserving Smoothing of Tomographic Data

**Participants:** Nassim Jibai, Nicolas Holzschuch, Cyril Soler.

Figure 10. Left: A contour surface extracted from noisy tomographic data contains surface noise and several topological artifacts such as small handles and holes. Right: The surface extracted from our smoothed volume is clean, and yet small features, such as the thread in the screw, and sharp edges have been preserved.

Computer tomography (CT) has wide application in medical imaging and reverse engineering. Due to the limited number of projections used in reconstructing the volume, the resulting 3D data is typically noisy. Contouring such data, for surface extraction, yields surfaces with localised artifacts of complex topology. To avoid such artifacts, we propose a method for feature-preserving smoothing of CT data, illustrated in Figure 10. The smoothing is based on anisotropic diffusion, with a diffusion tensor designed to smooth...
noise up to a given scale, while preserving features. We compute these diffusion kernels from the directional histograms of gradients around each voxel, using a fast GPU implementation. This work has been published as a Siggraph’2011 Poster [26].

6.3.2. 3D Inverse Dynamic Modeling of Strands

Participants: Alexandre Derouet-Jourdan, Joëlle Thollot.

In this work, we propose a new method to automatically and consistently convert 3D splines into dynamic rods at rest under gravity, bridging the gap between the modeling of 3D strands (such as hair, plants) and their physics-based animation. An illustration is given in Figure 11. This work is done in collaboration with F. Bertails from the BIPOP team-project. It has been published in a Siggraph’2011 poster [25].

6.3.3. Lagrangian Texture Advection: Preserving both Spectrum and Velocity Field

Participants: Eric Bruneton, Nicolas Holzschuch, Fabrice Neyret.

Texturing an animated fluid is a useful way to augment the visual complexity of pictures without increasing the simulation time. But texturing flowing fluids is a complex issue, as it creates conflicting requirements: we want to keep the key texture properties (features, spectrum) while advecting the texture with the underlying flow — which distorts it. In this context we present a new, Lagrangian, method for advecting textures: the advected texture is computed only locally and follows the velocity field at each pixel (see Figure 12). The texture retains its local properties, including its Fourier spectrum, even though it is accurately advected. Due to its Lagrangian nature, our algorithm can perform on very large, potentially infinite scenes in real time. Our experiments show that it is well suited for a wide range of input textures, including, but not limited to, noise textures. This work has been published in the IEEE Transactions on Visualization and Computer Graphics (TVCG) [18].

6.3.4. Feature-Based Vector Simulation of Water Waves

Participant: Fabrice Neyret.

We have developed a method for simulating local water waves caused by obstacles in water streams for real-time graphics applications. Given a low-resolution water surface and velocity field, our method is able to decorate the input water surface with high resolution detail for the animated waves around obstacles. We construct and animate a vector representation of the waves. It is then converted to feature-aligned meshes for capturing the surfaces of the waves (see Figure 13). Results demonstrate that our method has the benefits of real-time performance and easy controllability. The method also fits well into a state-of-the-art river animation system. This work has been published in the Journal of Computer Animation and Virtual Worlds [19].
Figure 12. Left: Our method advects open-domain textures preserving both the spectrum and the motion field, in real-time. Right: Various applications in 2D and 3D, with procedural, image, bump, displacement textures.

Figure 13. Our method permits the real-time rendering of highly detailed animated features on large river scenes.
6.3.5. Volume-preserving FFD for Programmable Graphics Hardware

**Participant:** Georges-Pierre Bonneau.

![Figure 14. Left: FFD deformation of the armadillo mesh, without volume preservation. Right: our technique: GPU-based volume preservation of the FFD deformation.](image)

This work is the result of a collaboration with S. Hahmann from the EVASION team-project, Prof. Gershon Elber from Technion and Prof. Hans Hagen from the University of Kaiserslautern.

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 work 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 (see Figure 14). 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. The results have been published in a Visual Computer journal paper [17].
5. New Results

5.1. Dynamic World Perception and Evolution Prediction

5.1.1. Environment modeling and sensor data acquisition


An overall architecture of our environment-modeling module with the inputs from heterogeneous sensors is shown in Fig. 6. The combined use of two lidars and stereo-vision helps mitigate uncertainty and allows for detection of partially occluded objects. The data processing includes the computation of probabilistic occupancy grids for each sensor and their subsequent fusion with the Bayesian Occupancy Filter (BOF). The output of the module is an estimation of the position, velocity and associated uncertainty of each observed object, which are used as input to the risk assessment module.

![Figure 6. Architecture of the environment modeling module.](image)

This architecture is implemented on our experimental platform, a Lexus LS600h car shown in Fig. 7. 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.

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. The observed region is 40 m long by 40 m wide, with a maximum height of 2 m. Cell size of the occupancy grids is 0.2x0.2 m.
The Lexus experimental platform provides to acquire sensor data in real traffic environments: eight layers of laser scans, stereo images, IMU data (accelerations), velocity, GPS position, 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.

5.1.2. Bayesian fusion of visual and telemetric information

Participants: Igor Paromtchik, Christian Laugier, Mathias Perrollaz, Amaury Nègre.

5.1.2.1. Concept of BOF and obstacle detection in occupancy grids

Obstacle detection is a widely explored domain of mobile robotics. It presents a particular interest for the intelligent vehicle community, as it is an essential building block for Advanced Driver Assistance Systems (ADAS). In the ANR project LOVE (Logiciel d’Observation de Vulnerables) and ArosDyn project, the e-Motion team proposed to perform obstacle detection within the occupancy grid framework. In order to work efficiently with occupancy grids, we have previously developed a probabilistic framework with the Bayesian Occupancy Filter (BOF) [40] [88] (patent 0552736 (2005)), which provides filtering, data fusion, and velocity estimation capabilities while allowing for parallel computation. The Fast Clustering and Tracking Algorithm (FCTA) [73] is then used to identify and track individual objects. The BOF is designed with the intent of its implementation in hardware as a system-on-chip. Like other grid based approaches, the BOF framework performs sensor fusion at the cell level [40]. The BOF evaluates probabilities of both cell occupancy and cell velocity for each cell in a four-dimensional spatio-temporal grid. The monitoring of traffic scenes includes detection and tracking of objects by the FCTA [73].

Fig. 8 shows examples of occupancy grid mapping with the proposed approach. The arrows indicate the pedestrian, the car, and the bicycle, which appear in the camera images and the occupancy grids. Because the accuracy of stereo-vision tends to become poor at large distance, the corresponding grid has been attenuated beyond 20 m and the system is tuned to give more confidence to the lidars than to the stereo-vision. One of advantages of sensor fusion is a larger viewfield so that the vehicles overtaking the ego-vehicle (they are not seen in the camera images) are correctly mapped on the resulting BOF grid. Moreover, the sensor fusion as well as the Bayesian estimation provide to filter out the laser impacts with the road surface, e.g. right lidar in Fig. 8.

Note that a large number of dynamic objects in the traffic scenes may lead to a failure of object-based fusion because of a large number of association hypotheses. The grid-based approach allows us to avoid the object association problem for sensor fusion.

5.1.2.2. Disparity space approach for a vision based occupancy grid

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 [79]. 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 [81], and on including the information from the road/obstacle segmentation of the disparity image [80]. 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 is described in [30].

Figure 9 shows an example of the occupancy grid computed by our new approach. We can observe that most objects are detected (light color), even if partially occluded (e.g. the sign on the right). Information from the road surface is also taken into consideration (dark areas). Moreover, similar to a laser scanner, it appears that regions in front of objects are seen as partially unoccupied, while less information is available behind obstacles (occupancy probability is closer to 0.5).

In 2011, we focused on including the approach into the risk estimation framework on our Lexus experimental platform. We implemented a demonstration to estimate a distance measurement to the closer object situated in the future trajectory of the vehicle. The future trajectory is estimated either by using a lane detection algorithm (in the highway) or by combining velocity and steering information of the vehicle. Figure 10 shows the HMI displayed in the car while driving.

5.1.2.3. Processing of multi-layer telemetric data in probabilistic framework

Participants: Mathias Perrollaz, Juan-David Adarve, Alexandros Makris.
The occupancy grid computation based on a laser scanner uses the classical independent beam sensor model [90]. Since our vehicle is equipped with two four-layers laser scanners, it is necessary to merge the data from the multiple layers. In the original BOF framework, the fusion was performed through the classical Bayesian Fusion methodology. As shown in figure 11, this method causes problems of misdetection when some beams go over an object. In 2011, we proposed and implemented another approach. The fusion is now obtained through a weighted sum of the occupancy grids provided by each layer. The weight of each layer is obtained by computing a confidence grid. This confidence depends both on the inclination of the layer and on the possible occlusions. The new approach provides a more precise description of the environment.

5.1.3. Sensor Fusion and parameters estimation

Participants: Agostino Martinelli, Chiara Troiani.

This is the follow up of the research activity started in 2009, when a self-calibration problem for a wheeled robot has been investigated. The main results achieved during that year were published in [69], [71] and [70]. This calibration problem allows us to introduce a general framework able to deal with any estimation problem. This framework is based on a new theoretical concept, the concept of continuous symmetry. Detecting the continuous symmetries of a given system has a very practical importance. It allows us to detect an observable state whose components are non linear functions of the original non observable state. The general theory has been developed during the last two years. Preliminary results have been published in 2010 [72] and a more complete version of these results, which include several extensions, has been published on Transaction on Robotics, in 2011 [9].

In 2011, this general framework has been extensively applied to investigate the problem of the fusion of visual and inertial data in the framework of the European project sFly. Special emphasis has been devoted to the structure from motion problem (SfM) when fusing these data. This problem has particular interest and has been investigated by many disciplines, both in the framework of computer science ([35], [54], [56], [87] and references therein) and in the framework of neuroscience and vision perception ([67], [95] and references therein). Even though prior work has answered the question of which are the observable...
Figure 10. a) segmentation of the environment with the stereo-vision algorithm. Blue areas belong to the road surface, while red areas belong to the obstacles. b) HMI shown in the car during the demonstration of risk estimation. The trajectory is estimated by considering the velocity and steering angle of the ego vehicle. Here the car in front is not considered as dangerous because it is more than 2 seconds ahead. c-d) Another example, on the highway. For this example, the trajectory is estimated by considering the road markings.
Figure 11. Occupancy grid computed after fusion of eight layers of laser data. Above: with the previous approach, some objects are not correctly represented (e.g. the barrier on the left). Below: with the new approach, the description is more precise.
modes, i.e. the states that can be determined by fusing visual and inertial measurements [35], [54], [56], the questions of how to compute these states in the absence of a prior, and of how many solutions are possible, were still unanswered. During 2011, we have derived, for the first time, a closed form solution to the SfM problem in this case, allowing the determination of the observable modes without the need for any prior knowledge. The proposed solution analytically expresses all the observable modes in terms of the visual and inertial measurements acquired during a given (short) time-interval allowing the determination of all the observable modes without the need for any prior knowledge. Additionally, we have shown that this problem can have a unique solution or two distinct solutions or infinite solutions depending on the trajectory, on the number of point-features and on the number of monocular images where the same point-features are seen. Our results are relevant in all the applications which need to solve the structure from motion problem with low-cost sensors and which do not demand any infrastructure. Typical examples are the emergent fields of space robotics [77], humanoid robotics and unmanned aerial navigation in urban-like environments [93], where the use of the GPS is often forbidden. Furthermore, our results could play an important role in neuroscience by providing a new insight on the process of vestibular and visual integration. To this regard, we remind the reader that the influence of extra retinal cues in depth perception has extensively been investigated in the last decades. In the case when this extra retinal cue is the motion parallax induced by self-motion relative to a stationary environment, the scale factor is provided by the head velocity [65], [66]. The vast majority of these studies, consider the case when the head motion is active [38], [94]. This prevents the possibility to understand the contribution of the vestibular signals because of efference copy generated by active self movement. However, a very recent study investigates this problem by performing trials with passive head movements [43]. The conclusion of this study is that the combination of retinal image with vestibular signals can provide rudimentary ability to depth perception. Our findings could provide a new insight to this problem of depth perception since by combining retinal image with vestibular signals it is possible to determine the scale factor even without any knowledge about the initial speed. New trials would be necessary in order to verify whether a mechanism reproducing our closed form solution is present in humans and/or in other animals (especially the ones without binocular vision). Our findings also show that it is possible to easily distinguish linear acceleration from gravity. Specifically, our closed form solution perform this determination by a very simple matrix inversion. This problem has also been considered in neuroscience [75], [31]. Our results could provide a new insight to this problem since they clearly characterize the conditions (type of motion, features layout) under which this determination can be performed.

Our results have been published in three conference papers [14], [11], [15] and have been accepted for publication in transactions on robotics (a version is currently available as a technical report, [29]).

In parallel to this theoretical activity an experimental activity has started in order to experimentally validate our findings in the near future and to deploy our technologies to industrial partners. To this regard, a contact with the company Delta Drone in Grenoble has been established and a valorization contract with a SME in the field of civil drone applications is currently in preparation.

5.1.4. Analysis of dynamic scenes for collision risk assessment


The grid-based environment representation is used for dynamic scene analysis in the Arosdyn project [78]. The original idea behind the risk estimation approach developed in the e-Motion team consists in considering the possible behaviors of the vehicles in the scene. Indeed, with the classical TTC(time to collision)-based approach, the risk is estimated based on the prediction of the trajectory, considering the current state of the objects. This is only valid for very short term predictions, and in some cases it can result in a over-estimation of the collision risk. Understanding the intention of the other participants of the road scene allows a longer term, more precise prediction of trajectories.

Our approach is divided into two steps: behavior recognition and behavior realization. The behavior recognition aims at estimating the probability for a vehicle to perform one of its feasible behaviors. The behaviors are semantic representations of driving maneuvers (e.g. turn left, turn right, go straight, ...). The probability
distribution over possible behaviors is obtained by inference using layered HMMs. Driving behavior realization is modeled as Gaussian Process (GP). This model allows us to obtain the probability distribution over the physical realization of the vehicle motion (i.e. trajectories) by assuming a usual driving, for a given behavior. Finally, a complete probabilistic model of the possible future motion of the vehicle is given by the probability distribution over driving behaviors, and by the realization of these behaviors. The risk calculation is performed by sampling of the paths from the corresponding GP. The fraction of the samples in collision gives the risk of collision.

In 2011, we conducted some early experiments on sensor fusion, using real data acquired with our Lexus experimental vehicle [16]. Moreover, the global framework of the Arosdyn project has been presented in [8].

5.1.5. Recognition for intelligent vehicles


We developed a generic object class recognition method. The method uses local image features and follows the part based detection approach. 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 [32] [45]. 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 [51] [44]. The approaches that perform data fusion from various sensors have proven to be the more robust in a variety of road conditions [86].

This work focuses on the development of an object class recognition system which follows the part based detection approach [64]. 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. Figure 12 presents the steps of our approach.

Figure 12. Detection procedure steps. The stereo information is used to define the regions of interest for the subsequent steps. Intensity and depth features are extracted from a dense grid within these regions. In the following the features are matched with the codebook clusters which are in turn used to estimate the posterior for the object in each position. The detections are the local maxima of the posterior.
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. Fig. 13 shows some example detections. The proposed method detects side-views of cars in various scales, in cases with partial occlusions, and under significant background clutter.

Figure 13. Car-side detection examples. True and false positive detections are represented by red and yellow bounding boxes respectively. (a) Cars in different scales with significant background clutter and significant occlusions are detected. (b) Precise detection of the un-occluded vehicle, whereas a vehicle that is heavily occluded in the left is not detected. (c) Difficult detection of a vehicle which is far and partially occluded and a false detection in the region between the road surface and the trees. (d) Detection with partial occlusion. (e) Partial detection of a taller than normal vehicle on the left. The training dataset does not contain vehicles of this type. (f) Successful detection of a partially occluded car and a false positive arising from a bus and a van. Training separate detectors for these type of vehicles as well will help to avoid these false alarms.

5.1.6. Context-aware Bayesian estimation of risk at road intersections for cooperative vehicles

Participants: Stéphanie Lefèvre, Christian Laugier.

The work developed in this PhD is done in collaboration with Renault (CIFRE thesis) and concerns safety applications for cooperative vehicles.

In a few years, car manufacturers will start equipping vehicles with V2X communication devices, which will allow vehicles to share information with other vehicles and with roadside units using a dedicated communication channel. This new sensor on the car opens a whole new world of possibilities for Advanced Driver Assistance Systems (ADAS). In particular, the fact that the vehicle is able to “see” a car before it even enters the field-of-view of the driver allows for a better assistance in the tasks of perceiving, analyzing, predicting, and estimating the risk of a situation.

Early in the PhD we identified safety applications at road intersections as a relevant application domain for V2X technologies. The variety and complexity of scenes at road intersections makes reasoning and interpretation particularly difficult. On the other hand, intersections are a location of many accidents (they represent up to 50% of accidents in some countries), therefore reducing the accident rates in these areas would
have a considerable impact of global traffic safety. We also identified the key issues (and challenges) to be 1) situation understanding and 2) risk assessment, to be carried out from incomplete models and uncertain data.

The focus of the year 2010 was on the first of these two problems. We developed a Bayesian Network that could estimate a driver’s intended exit lane at an intersection 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 the position and turn signal of a vehicle and the driver’s intended exit lane. Initial results of this work were published in IEEE CIVTS’11 [12], then in IEEE IV’11 [13] with a more thorough evaluation.

The objective of the work conducted in 2011 was twofold:

1. Extend the initial system: add some filtering and take into account the priority rules.
2. Estimate the risk of a situation, based on the estimated behavior/intention of the drivers in the scene.

We proposed a probabilistic motion model for vehicles approaching and traversing an intersection that incorporates some knowledge about how the context (i.e. the traffic rules, the presence of other vehicles, the geometry and topology of the intersection) influences vehicle behavior. The distinctive features of our algorithm are:

- The explicit use of priority rules
  Priority rules are explicitly taken into account in the motion model: the necessity for a driver to stop and/or yield to another vehicle at an intersection is estimated, jointly with the driver’s intention to comply. This allows for a flexible and computationally inexpensive computation of risk. Flexible because depending on the final application one can decide to compute different types of risk, e.g. the probability that a specific vehicle is a violator, or the probability that a crash will occur between two vehicles, or the risk of a specific maneuver for a vehicle. Inexpensive because these can be computed without performing trajectory prediction for the vehicles in the scene.

- The assumption that drivers generally respect traffic rules
  Instead of making the classical assumption that vehicles’ trajectories are independent, we model their mutual influences by introducing a prior knowledge that drivers generally respect priority rules. The motion model therefore takes into account the priority rules and the presence of other vehicles to better interpret correctly a vehicle’s behavior. The advantages are twofold. Firstly, we are able to better estimate the maneuver intention of the drivers, which means our situation assessment capabilities are improved. Secondly, risk is estimated with a higher sensitivity. We avoid risk overestimation while still being able to detect dangerous situations as well as the conventional, more conservative, methods.

This reasoning is implemented using a Bayesian filter which estimates the hidden variables M (maneuver intention), D (distance to intersection), H (intention to stop) and H’ (necessity to halt) jointly for all the vehicles in the scene, using the position, speed and heading information shared between the vehicles via V2X communication. Inference on the hidden variables is carried out by a particle filter. The algorithm was described in an INRIA research report [27]. In this report we showed by reasoning on theoretical scenarios that our assumption that drivers tend to respect priority rules should lead to improved situation assessment and risk assessment (see Fig. 14).

Recently, data has been collected at an intersection using the Renault demonstrator vehicles, so that our algorithm can be tested on real data. Preliminary results seem to confirm that the intuitions described in the research report were correct. A Graphical User Interface is in the process of being developed so that demonstrations of the system can be carried out live in the Renault demonstrator vehicles (see Fig. 15).

5.2. Human Centered Navigation in the physical world
Figure 14. Illustration of a scenario where the advantage of taking into account the interactions between vehicles for maneuver prediction is obvious for ADAS applications. The behavior of the red vehicle is interpreted differently depending on whether or not the interactions with the green vehicle are considered.

Figure 15. Graphical User Interface for warning a driver of a violation of priority rules at an intersection (the violator vehicle is displayed in red).
5.2.1. Goal oriented risk based navigation in dynamic uncertain environment  

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 [47]. 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 Risk-RRT, is a sampling-based partial planner guided by the risk of collision. Results concerning this work were published in [48] [49] [50].

In 2011, our algorithms were integrated into an embedded software architecture for social aware navigation (see fig. 16 ). For this purpose we started to migrate our algorithms to a new experimental plateform. Moreover, we adapted the code to the open source software called ROS (Robot Operating systems ²) which offers tools to develop robot applications based in state of the art algorithms. Particularly, localization and visualization tools have been used. We have linked the control of our robotic wheelchair, the Risk-RRT planning and the social filter modules described in 5.2.2 into the framework ROS as shown in figure 16. The main objective was to increase the visibility of our approach and develop common libraries with research groups in robotics. In 2011, in the scope of the AEN PAL project, we started a collaboration with the EPI Arobas and complementary developments have been put on the INRIA forge.

Next two sections are conducted under the french project PAL “Personally Assisted Living” with a goal to enhance the quality of living by providing more autonomy in the daily activities of the disabled.

5.2.2. Social conventions based navigation  
Participants: Jorge Rios-Martinez, Anne Spalanzani, Christian Laugier.

---

²Willow Garage Inc., http://www.ros.org
The objectives of this work are to integrate the notion of comfort in the classical safe navigation methods. If one considers that the navigation system transports a person, the integration of social conventions in the navigation strategy starts to be crucial. In this work, we propose to integrate the notions of personal space and interaction between people. We propose to enrich the knowledge the robot has, with a representation of the social conventions. The robot must take into consideration interactions to avoid groups of people (even if passing through the group is the “best” path for a conventional planning algorithm), or to join a group with a behavior close to the one of a human. To understand the behaviors of interaction between humans and the management of space, the works developed in the area of sociology to define some concepts as **Personal space**, **o-space** and **F-formations** are used.

- **Personal Space**
  
  In [53], Hall describes the use of space between humans, he observed the existence of some rules that conducted people to keep distances from others. He proposed a classification of the space around a person (its **Personal Space**) in social interaction in four zones:
  
  - the public zone > 3.6m,
  - the social zone > 1.2m
  - the personal zone > 0.45m
  - the intimate zone < 0.45m
  
  This is a useful tool for a robot to understand the intentions of the humans. It is well known that these measures are not strict and that they change depending on age, culture and type of relationship but the categories proposed explain very well reactions like the uncomfortable sense of a stranger invading your intimate zone or the perception of somebody looking social interaction because he is entering to your social zone.

- **F-formation**

![Figure 17. Examples of F-formations: (a) Vis-a-vis, (b) L-Shape, (c) C-Shape.](image)

In [57], Kendon observed that people interacting in groups follow some spatial patterns of arrangement. When people are executing some activity they claim an amount of space related to that activity, this space is respected by other people and Kendon referred it as individual’s **transactional segment**. This transactional segment can vary depending on body size, posture, position and orientation during the activity. Moreover the groups can establish a joint or shared transactional segment and only the intervenants have permitted access to it, they protect it and others tend to respect it. The **o-space** is that shared transactional segment. A F-formation system is the spatial-orientation arrangement that people create, share and maintain around their o-space. We can see in fig. (17) three examples of F-formations.

The first stage in order to achieve an integration of social concepts with robot navigation was to include estimations of the risk of disturbing personal space and interaction space in the general risk estimation. A strategy to detect interactions in the environment based in the velocity, position and orientation of humans was implemented.
In fig. 18 we observe the results of the proposed integration, the robot (green rectangle) can use the detections of conversations (light ellipses) between humans (blue circles) for add more risk to paths that invade the space of conversations. When a conversation is detected, a bi-dimensional Gaussian $G$ is created to represent the interaction space, also called o-space, the center of this space is approximated by taking into account the the participants’ poses. Then, $G$ is used to obtain an estimation of risk of disturbing by passing around the conversation. The navigation strategy is based on the Risk-RRT algorithm. Details of this approach were published in [18].

5.2.3. Autonomous Wheelchair for the Elderly People’s Assistance

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

The elderly and the disabled are expected to benefit from the new technologies in the field of autonomous navigation robotics. Normal users of electric wheelchairs will also benefit from the development of more automatic functionalities bringing an extra driving comfort, especially during delicate maneuvers such as narrow door passages. This contribution is similar to the installation of driving assistance on a car. A simple improvement of the classical powered wheelchair can often diminish several difficulties of control. Comfort defined as a state of ease and satisfaction of bodily wants, with freedom from pain and anxiety, has recently emerged as a design goal in autonomous navigation systems. Designers are becoming more aware of the importance of the user when scheming solution algorithms. The idea of comfort is especially important in the case of wheelchairs where the occupants are weak as result of their age or disease.

For any robot that is designed to transport people, the trajectory should be smooth and correspond to the user’s understanding as much as possible. Since human interpretation of the environment often differs from a robot’s interpretation, the decisions taken by the system might seem incomprehensible to a human observer. For example an autonomous vehicle could refuse to move forward due to some obstacle, while a human user would easily be able to move its way through. This undesirable behaviors may prove irritating and with time may lead to users stopping from using the system.

In 2011 we setup a robotic wheelchair as a trial platform. The wheelchair is a differential drive robot equiped with a SICK LMS-200 lidar to get 2D range information from the environment, odometry sensors, and a velocity controller we have also added a kinect sensor in order to perform some in the field of social
interactions. Some basic functions can be executed including the mapping of the environment using a Rao-Blackwellized Particle Filter \[ 52 \], localization using an Adaptive Monte Carlo Localization approach (AMCL) \[ 91 \], global planning using an A* algorithm \[ 60 \] and local reactive planning using the Dynamic Window Algorithm \[ 46 \].

Alongside we started working with the Kinect sensor to detect and track people. Using the given tracking information, the wheelchair is able to follow a human located in front of it. This behavior is aimed to bring assistance not only to the user but also to the caregiver by allowing him to move without pushing the wheelchair. The technical implementation of the related approaches has been done on the basis of the ROS middleware due to easy integration with other opensource robotics software which benefit sharing and testing developed software.

In 2012 we shall focus on the estimation of the user intentions by learning models of behavior. We’ll then use these models to propose an adaptive autonomous navigation method that best answer the user needs.

5.2.4. Multi-Robot Distributed Control under Environmental Constraints

**Participants:** Agostino Martinelli, Alessandro Renzaglia.

This research has been carried out in the framework of the European project sFly. In recent years it is revealed more and more the importance of using multi-robot systems for security application, otherwise impossible to be performed by a single robot.

The main problem approached is the optimal surveillance coverage of an unknown and complex environment, i.e. finding the optimal deployment for the robots and the way to safely reach such configuration. The solution for the 2D case without obstacles is already known in literature \[ 39 \]. On the other hand, for the non-convex case, it is still a difficult problem. In \[ 84 \] we firstly proposed a possible strategy based on a combination of the repulsive potential field method and the Voronoi partition. Then, in the last two years we have mainly approached the coverage problem by using a new stochastic optimization method. This work is in collaboration with professor Elias Kosmatopoulos, from CERTH (Thessaloniki), and professor Lefteris Doitsidis, from TUC (Crete), partners in the sFly project.

The Kosmatopoulos’s group has proposed a new adaptive stochastic optimization algorithm for a general class of multi-robot passive and active sensing applications \[ 59 \], \[ 58 \]. This method possesses the capability of being 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 measurement at each iteration of the algorithm employed to optimize it. As a result, it perfectly suits for multi-robot optimal coverage in non-convex environments, where the analytical form of the function to be optimized is unknown but the function is available for measurement (through the robots’ sensors) for each multi-robot configuration.

The main results obtained for the 2D case by using this method has been published in \[ 85 \], \[ 83 \]. We assume the robots are equipped with global positioning capabilities and visual sensors able to monitor the surrounding environment. The goal is to maximize the area monitored by the team, by identifying the best configuration of the team members. Moreover, in 2011, a distributed version of the algorithm was presented in \[ 17 \]. In multi-robot systems, a distributed approach is desirable for several fundamental reasons. The most important are failure of the central station and limited communication capabilities. The proposed approach has the following key advantages with respect to previous works:

- it can solve the problem in a distributed way;
- it does not require any a priori knowledge on the environment;
- it works in any given environment, without the necessity to make any kind of assumption about its topology;
- it can incorporate any kind of constraints, for instance regarding a possible existing threshold on the maximum distance on the monitored region, or a limited visibility angle;
- it does not require a knowledge about these constraints since they are learnt during the task execution;
- its complexity is low allowing real time implementations.
Figure 19. (a) Wheelchair used in the emotion team, (b) Two people being tracked using the kinect and the map of the environment done by the wheelchair.
The previous approach has been also extended for the more important and realistic 3D case. Working in collaboration with the ETHZ (Zurich), some simulations using real data, which were collected with the use of a miniature quadrotor helicopter specially designed for the needs of the European project sFly, have been performed (see fig. 20). This work has lead to two joint publications with CERTH and TUC: one conference paper to present (CDC2011) and one journal papers under review, and two joint publications with CERTH, TUC and ETHZ: one conference paper ([10]) and one journal papers under review.

In 2011, this approach has been combined with human aware navigation technics presented in section 5.2.5.

In the next months, the algorithm will be implemented on real MAVs for the final demo of the project. This demo will include experimentation both in indoor and outdoor complex environments.

Finally, a new collaboration with professor Kosmatopoulos has recently begun. The objective of this work is to develop a new efficient and scalable algorithm for multi-robot active control to perform cooperative simultaneous localization and mapping (CSLAM) and target tracking. The main idea is to use a convex optimization algorithm based on Semi-Definite Programming and Sum-of-Squares polynomials. Preliminary simulation results are very promising and a journal paper is under preparation.

5.2.5. Exploring stochastic optimization method to navigate between humans


Suppose that we have a robot navigating in an unknown and complex environment where people are moving and interacting. In such scenario the respect of the humans’ comfort becomes an important goal to achieve. The discomfort concept could be very general but we focus on the one mentioned before, i.e., the discomfort caused by disturbing one interaction or a personal space of humans. The approach here is to minimize the discomfort while the robot is navigating. As we cannot measure directly the value of discomfort, we can...
infer it by modeling the concepts presented before using simple equations and after by applying a method of optimization. We propose to exploit a new stochastic and adaptive optimization algorithm (CAO) [59]. This method is very useful in particular when the analytical expression of the optimization function is unknown but numerical values are available for any state configuration. Furthermore, the proposed method can easily incorporate any dynamical and environmental constraints. To validate the performance of the proposed solution, several simulation results are provided.

In fig. 21 the model for discomfort function is shown together with robot navigation. At each step the robot randomly generate configurations in the environment and selects the one that takes it closer to the goal while minimizing values for the discomfort function of humans in the environment, this is repeated until goal is reached. Several executions of proposed approach in different scenarios can be observed in fig. 22.

The details of this approach have been submitted to ICRA2012.

5.3. Bayesian Modelling of Sensorimotor Systems and Behaviors

Results proposed in this section were done in collaboration with the LPPA collège de France.

5.3.1. Bayesian programming applied to a 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 [61]. 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 plan recognition from noisy and incomplete observations. Previous plan recognition works in multiplayer games were mainly based on planning and case-based reasoning (CBR) [92], [68], [55], [76] or HMMs [41]. CBR allows for taking domain knowledge into account easily while not dealing efficiently with uncertainty/incompleteness of information, HMMs deal with uncertainty quite well but domain knowledge is harder to structure. We found different ways to decompose the joint $P(Observation_{1:N}, Plan_{1:M})$ which allows for tractable and robust inference. For instance with the help of intermediate variables which can be derived from domain knowledge (as we did) or found automatically (e.g. cross-validation on a HMM). Particularly, we were able to structure dependencies between domain knowledge extracted variables using coherence variables. We then learn the parameters of such joint distributions from data. Supervised (labeled), and semi-supervised learning (when we label automatically from clustering) have led to a publication at CIG (IEEE) 2011 [19] and unsupervised learning (using only raw game data) led to a publication at AIIDE (AAAI) 2011 [20].

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 9th (out of 18) and 4th (out of 10). We also published last year’s result on multiple units control in real-time engagements (see [23]) at CIG (IEEE) 2011 [21]. As optimal micro-management is almost always intractable (P-space) in real situations, we considered each unit as a Bayesian sensory motor robot which makes a fusion of its sensory inputs about the world, the enemy units, but also its allies (without explicit communication for less complexity) and higher level directions. So the units only take short term decision on where to go and who to attack, higher level planning is done at a squad (and then army) level and given as a sensory input. Results in micro-management tournaments are state of the art. In the more general case, they could be improved by reinforcement learning of the models parameters.
We are now working on concurrent goals resources attribution, still in the context of incomplete knowledge about the opponent. We are also working on correlating low-level observations (effects) and high-level inferences (causes) about the enemy strategy to be able to predict its future behavior.

5.3.2. 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, which have been widely argued for and against. The debate appears to be stagnating. This work is based on the belief that mathematical modeling of these theories could provide breakthroughs. More precisely, it is proposed that casting these theories into a single, unified mathematical
framework would be the most efficient way of comparing the theories and their properties in a systematic manner.

Bayesian modeling provides a mathematical framework that precisely allows such comparisons. The same tool, namely probabilities, can be used both for defining the models and for comparing them. Moreover, the use of a unified framework implies that common hypotheses would have common mathematical translations. This helps toward more principled studies of the competing theories.

Following this integrative approach, the motor, auditory and perceptuo-motor theories are thus cast into one unifying Bayesian framework in which they all appear as instances of various questions asked to one probabilistic communication model. This allows to compare these theories through quantitative testing in various paradigms. The work is aimed at understanding the differences in the predictions given by the different theories, and from these predictions to suggest experiments involving human subjects.

The model was used first to work on purely theoretical simulations aimed at studying with diverse paradigms the decrease in the performances predicted by the different theories due to communication noise. It was then used to work on plosive syllables production and perception, thanks to VLAM, a vocal tract simulation tool, which allows to map articulatory parameters to acoustic signals.
6. New Results

6.1. High level model for shapes

6.1.1. Constructive implicit 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.

Within Cédric Zanni’s PhD, we are collaborating with a researcher in formal computation, Evelyne Hubert, to improve and extend the analytical methods for computing closed form solutions for convolution surfaces. We introduced a warping method for enabling the modeling of complex helical shapes from a single implicit primitive (fig. 3), which greatly enhances efficiency [14]. We also proposed a method based on anisotropic, surface Gabor noise, for generating procedural details on skeleton-based implicit surfaces. The surfaces enhanced with details can still be smoothly blended, with a natural transition between the details they carry. A paper has been submitted for publication. We are currently developing normalized convolution surfaces, invariant through homothetic transformations, and which will provide an intuitive blending sharpness parameter, usable with simple additive blending.

![Figure 3. Top: Helical primitives. Bottom: Implicit modeling of a squid. Each tentacle is made of two helical primitives.](image)

Lastly, 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 is currently submitted for publication.
6.1.2. Ontology-based mesh segmentation  
**Participant:** Olivier Palombi.

The smart use of data by automated systems is now a problem having implications as various as the optimization of the functioning of web search engines or medico-surgical simulation. The program MyCF is an attractive innovation in this field, gathering an organization of the biomedical knowledge in the form of ontologies, 3D acquisition of anatomical structures and also the possibility to export these 3D structures to biomechanics simulation programs, like SOFA. Our work consisted in creating, thanks to the program Protégé, an ontology of the functions of the human body (which didn’t existed then) and to couple it with the anatomical ontology FMA (ontology we significantly reworked). The objective was to build connections introducing a link between anatomical structures and functions they bear. Once this goal reached, it became possible for a computer to link the elimination of an anatomical structure and the loss of a function, this real-time. Thus we established the foundations of an ontology which currently gather 84484 entities (4330 of which are functional entities) and 4159 relations. So, our contributions are threefold: first of all, the creation of an ontology of functions of the human body which is an original one; then the redrafting of the FMA ontology which allowed us to complete some lacks and above all to make of it a tool more oriented towards the practical applications we waited about it, and finally, the contribution which seems to us the more significant is, without any doubt, the institution of a link between these two ontologies. This work is at is moment unfinished and should be pursued in order to approach always more the reality in the field of medico-surgical simulation.

6.1.3. French translation of the Foundational Model of Anatomy (FMA) ontology  
**Participant:** Olivier Palombi.

The goal of this study, performed in collaboration with the LITIS ¹ was to facilitate the translation of FMA vocabulary into French. We compare two types of approaches to translate the FMA terms into French. The first one is UMLS-based on the conceptual information of the UMLS metathesaurus. The second method is lexically-based on several Natural Language Processing (NLP) tools. The two approaches permitted us to semi-automatically translate 3,776 FMA terms from English into French, this was to added to the existing 10,844 French FMA terms in the HMTP (4,436 FMA French terms and 6,408 FMA terms manually translated) [9].

6.1.4. Homology computation  
**Participants:** Dobrina Boltcheva, Jean-Claude Léon.

This work is a part of the BQR project IDEAL (see Section 8.1.2) which is performed in collaboration with Leila de Floriani from the University of Genova in Italy. The main goal of this project is to study non-manifold geometrical models and to find out features allowing to classify these models and criteria for determining their shape. We are interested in non-manifold models such as idealized industrial CAD models, since they are still ill-understood even if they are frequently used in computer graphics and many engineering applications. This year, we have worked on the computation of topological invariants on non-manifold simplicial complexes, such as the homology groups, since they play a crucial role in the field of shape description and analysis. The goal was also to acquire a better understanding of the behaviour of the homological groups on non-manifold models. A first step towards this goal has been achieved this year and we have developed an efficient method for computing the homology of a large simplicial complex from the homologies of its sub-complexes. This work has been already published in a research report in 2010 and a journal paper in the context of the international conference on Solid and Physical Modeling [2].

6.1.5. Creased paper modeling  
**Participants:** Marie-Paule Cani, Stefanie Hahmann, Damien Rohmer.

Although very common in real life, 3D creased paper models are rarely seen in virtual scene due to the lack of available modeling tools.

¹Laboratoire d’Informatique, de Traitement de l’Information et des Systèmes (LITIS) Institut National des Sciences Appliquées de Rouen – Université du Havre – Université de Rouen
We developed a new approach to efficiently generate a 3D model of creased paper from a boundary curve (see fig. 4). The generated surface lies on the given boundary curve while preserving the lengths with respect to the original pattern. Contrary to other approaches, this method can seamlessly handle sharp creases while automatically generating the optimal mesh to this shape. The generation is fast enough to be used interactively, and the physical properties such as developability are approximately preserved. This generation of static surface generation has been published in EG short paper [22].

### 6.1.6. Spline surface models for arbitrary topologies

**Participant:** Stefanie Hahmann.

In geometric modeling, quad meshes have always been popular, in the sense that NURBS surfaces which are composed of a tensor product network of quadrilateral patches, are the inevitable standard for describing free-form shapes. They are defined on a chessboard-like assembly of quadrilateral parameter domains. But when modeling shapes of arbitrary topological type with tensor product NURBS, it is necessary to overcome the restriction of the tensor product configuration where always 4 patches meet at a common corner by using singular parameterizations.

We are developing new smooth parametric surface models defined on irregular quad meshes, which 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. Herein, tensor product polynomial patches are assembled with tangent plane continuity, in one-to-one correspondence to the mesh faces. They are thus capable to represent manifold shapes of arbitrary topological type since no restriction on the number of patches assembled around a mesh vertex exists.

While subdivision surfaces can also produce a smooth shape with only a few subdivision steps from a coarse mesh, our parametric surfaces have the advantage to provide an explicit parameterization. Moreover, all classical modeling operations such as trimming, intersection, blending and boolean operations can be performed with parametric patches. Tensor product patches can furthermore make profit from the powerful tools of existing modeling systems for purposes of evaluation, display, interrogation and all operations cited above.

In collaboration with G.-P. Bonneau (Artis 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 on-going work.
6.1.7. **Point sampled surfaces**  
**Participant:** Stefanie Hahmann.

Point sampled geometry from scanned data exhibits very characteristic shapes, due to the presence of sharp features in most manufactured and designed objects. Therefore, reconstruction of surfaces from unorganized point sets using MLS fitting requires additional attention. In fact, it is an inherent property of MLS fitting to produce smooth surfaces, thus all sharp features in the point cloud may also be smoothed out. Instead of searching for appropriate new fitting functions our approach was to introduce a new method for selecting an appropriate local point neighborhood for the projection operator so that a standard MLS fitting automatically reproduces sharp features.

This work was part of Christopher Weber’s Ph.D. thesis, which has been co-advised by S. Hahmann and H. Hagen from TU Kaiserslautern, Germany. First part of the work on Gauss map clustering for feature point detection has been published in the SMI 2010 proceedings. The second part has been submitted to a Computer Graphics journal. The thesis has been defended in August 2011.

6.1.8. **Volume preserving Free-Form 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 G. Elber from TECHNION, H. Hagen from TU Kaiserslautern, G.-P. Bonneau and S. Barbier from Artis INRIA. It has been accepted for publication in the journal The Visual Computer [6].

6.2. **High level models for animation**

6.2.1. **Geometrical methods for skinning character animations**  
**Participants:** Marie-Paule Cani, Stefanie Hahmann, Damien Rohmer.
Skeletal animation is an efficient and widely used technique in video games or movie industry due to its flexibility and simplicity. Still, the skinning method do not take into account informations about the physical properties of the underlying material. Therefore effects such as muscle bulging or fat tissue compression cannot be modeled without the addition of a tedious manual correction. Within Damien Rohmers PhD [1], an active geometry framework was proposed in order to enhance geometry information with a priori knowledge about how the underlying material can deform. For instance, bending the belly of an animal will be constraint to generate bulges that will preserve locally the volume (see fig. 6). The process can be either purely automatically generated, or it can be artistically controled.

6.2.2. 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 [12]. 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.

6.2.3. Frame-based simulation of deformable solids

Participants: Guillaume Bousquet, François Faure.

We present a new type of deformable model which combines the realism of physically based continuum mechanics models and the usability of frame-based skinning methods [4]. The degrees of freedom are coordinate frames (see Figure 15). In contrast with traditional skinning, frame positions are not scripted but move in reaction to internal body forces. The displacement field is smoothly interpolated using blending techniques such as dual quaternions. The deformation gradient and its derivatives are computed at each sample point of a deformed object and used in the equations of Lagrangian mechanics to achieve physical realism. This allows easy and very intuitive definition of the degrees of freedom of the deformable object. The meshless discretization allows on-the-fly insertion of frames to create local deformations where needed. We formulate the dynamics of these models in detail and describe some pre-computations that can be used for speed. We show that our method is effective for behaviors ranging from simple unimodal deformations to complex realistic deformations comparable with Finite Element simulations.
Figure 7. Physical realism is obtained with very sparse sampling: two frames are sufficient to model a dynamically deformable bunny [3].

We extend the approach (see fig. 7) to the simulation of complex, intricated material distributions using material-aware shape functions [3]. Given a volumetric map of the material properties of an object and a number of control nodes, a distribution of the nodes is computed automatically, as well as the associated shape functions. Reference frames are attached to the nodes, and deformations are applied to the object using linear blend skinning. A continuum mechanics formulation is derived from the displacements and the material properties. We introduce novel material-aware shape functions in place of the traditional radial basis functions used in meshless frameworks. These allow coarse deformation functions to very efficiently resolve non-uniform stiffnesses. Complex models can thus be simulated at high frame rates using a small number of control nodes.

To encourage its use, the software is freely available in the simulation platform SOFA.

6.3. Towards interactive digital creation media

6.3.1. Sketch-based modeling and shape editing

Participants: Adrien Bernhardt, Rémi Brouet, Marie-Paule Cani, Jean-Claude Léon, Olivier Palombi.

Figure 8. User-sketched high-field modeling a terrain.
3D modeling from a sketch is a fast and intuitive way of creating digital content.

We developed a method based on convolution surfaces for inferring free-form shapes in 3D from arbitrary progressive sketches, without any a priori knowledge on the objects being represented (see the section describing the Aestem Studio software. We recently investigated whether 2D deformation could be a better approach than sketching for defining the 2D sketch given as input [18]. Results are very promising: We are planning both to allow such intuitive deformations, combined with sketching, within the Aestem software, and to extend them to the editing of the 3D deformed shape. This will be done in the context of Rémi Brouet’s PhD thesis, co-advised by Renaud Blanch from the IIHM/LIG team.

We also develop methods for interpreting complex sketches (contours with T-junctions) based on some a priori knowledge. Our first work on this topic used the conventions of anatomical drawing to infer the 3D geometry of vascular systems, with branching and occlusions, from a single sketch [10]. We are also investigating the design of realistic terrains from a single sketch, within the PhD thesis of Adrien Bernhardt (see fig. 8). Our first advances include a new representation and new methods for generating a high-field from user-sketched constraint curves [30], [17].

6.3.2. Free-form sculpture

Participants: Marie-Paule Cani, Lucian Stanculescu.

In the context of Lucian Stanculescu’s PhD thesis, co-adviced by Raphaëlle Chaine from LIRIS (Lyon), we developed an interactive sculpting system enabling both arbitrary deformation and topological changes of a free-form shape [11]. Our method is based on a semi-regular mesh which adaptively refines and changes its topology according to the need. See Figure 9. We are currently extending the method for handing the sculpting of composite objects made of many different components.

6.3.3. Hand Navigator

Participants: Jean-Rémy Chardonnet, 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, our recent work focused on developing new devices for interacting with the model to deform. For the past two 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 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. An ongoing extension of the patent and a partnership with HAPTION company are new.
step toward the industrialization of this device. Dissemination to general public has been performed at the “Fête de la Science” and another exhibition. Publications will take place after setting up the patent extension. 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 has been set up with GIPSA-Lab to study the muscular activity during interactions.

6.3.4. Procedural modeling

Participants: Marie-Paule Cani, Arnaud Emilien.

We developed a method for procedurally generating villages with the appropriate roads and streets on arbitrary terrains, in collaboration with Eric Galin from LIRIS, Lyon. This work will be continued within Arnaud Emilien’s PhD thesis towards more general models for populating terrains with houses, vegetation, and animals. We will focus on the development of intuitive ways to edit procedural models, to overcome the main drawback of these approaches.

6.3.5. Computational model of film editing

Participant: Remi Ronfard.

![Figure 10. Interface developed for automatic editing of animated movies [31], [21].](image)

Building on Remi Ronfard’s experience leading the virtual cinematography research team at Xtranormal Technology, Montreal, we designed a novel computational model for automatic editing of animated movies (see fig. 10). A prototype has been implemented in a collaboration with the Bunraku/mimetic team, and demonstrated in poster sessions at the Symposium on Computer Animation (SCA) [31] and International Conference on Interactive Digital Storytelling (ICIDS) [21]. This early work opens new directions that will be further explored by the IMAGINE team, including corpus-based learning of cinematography and editing styles.
EXMO Project-Team

6. New Results

In the continuation of our previous work, in 2011 we developed our work on evaluation of ontology matching and especially in running new experiments and generating new tests (§ 6.1.1). We also continued our work on trust in semantic peer-to-peer systems (§ 6.2.2), the use of the $\mu$-calculus for evaluating RDF path queries (§ 6.2.1) and ontology matching for linking data (§ 6.1.2).

6.1. Ontology matching and alignment

We pursue our work on ontology matching and alignment support with contributions to evaluation, data interlinking and multilingual matching.

6.1.1. Evaluation

Participants: Cássia Trojahn dos Santos [Contact], Jérôme Euzenat, Jérôme David.

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 [4]. This year, the evaluation campaign had 16 different teams entered the evaluation which consisted of 5 different sets of tests. The participating systems and evaluation results were presented in the 6th Ontology Matching workshop, that was held in Bonn, DE [17][9].

The main activities carried out in 2011 were related to the automation and execution of the OAEI 2011 campaign, in the framework of the SEALS project (see § 8.2.1). This involved the following main tasks:

- describe evaluation processes within the early version of the SEALS platform [11];
- develop a client allowing participants to validate their wrapped tools and evaluate (offline and locally) their tools;
- develop a test generator for automatic generation of systematic benchmarks [12];
- providing participants with a better way to bundle their tools so that they can be evaluated within the SEALS platform; and
- analysis and report of the evaluation campaign results [9].

This work has been used in the OAEI 2011 evaluation campaign. More information on OAEI can be found at http://oaei.ontologymatching.org/.

6.1.2. Ontology matching for linked data

Participants: Zhengjie Fan, Jérôme Euzenat [Contact], Jérôme David.

The web of data consists of using 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.

In the context of the Datalift project (see § 8.1.1), we are developing a data interlinking module. Based on our analysis of the relationships between ontology matching and data interlinking [13], our goal is to generate data interlinking scripts on 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 Frei Universität Berlin, for processing linking scripts. So far we have demonstrated the ability to process simple scripts.

This work is part of the PhD of Zhengjie Fan, co-supervised with François Scharffe (LIRMM), within the Datalift project.
6.1.3. Multilingual ontology matching

Participants: Cássia Trojahn dos Santos [Contact], Jérôme David, Jérôme Euzenat, Giuseppe Pirrò.

We have participated in the creation of a benchmark for multilingual ontology matching, the MultiFarm dataset. This dataset is composed of a set of ontologies translated in different languages and the corresponding alignments between these ontologies. It is based on the OntoFarm dataset, which has been used successfully for several years in the Ontology Alignment Evaluation Initiative. By translating the ontologies of the OntoFarm set into eight different languages – Chinese, Czech, Dutch, French, German, Portuguese, Russian, and Spanish – we created a comprehensive set of realistic test cases. We plan to include this new dataset in the OAEI 2012 campaign.

Finally, in the context of the Cameleon project (see § 8.3.1) we have been working on the creation of a multilingual comparable corpora using as seed a set of multilingual aligned ontologies. These resources will be exploited in the process of populating and enriching ontologies as well as in the process of cross-lingual ontology alignment.

6.2. 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.

6.2.1. Path queries and $\mu$-calculus

Participants: Melisachew Wudage Chekol [Contact], Jérôme Euzenat, Pierre Genevès, Nabil Layaida.

Querying the semantic web is mainly done through SPARQL [15]. One of its extensions, PSPARQL (Path SPARQL) provides queries with paths of arbitrary length. We study the static analysis of queries written in this language with techniques based on $\mu$-calculus interpretation that have been used for XPATH. We have more specifically considered PSPARQL query containment: determining whether, for any graph, the answers to a query are contained in those of another query [18][14]. To that extent, we proposed an encoding of RDF graphs as transition systems and PSPARQL queries as $\mu$-calculus formulas. We then reduce the containment problem to testing satisfiability in the logic.

This work is part of the PhD of Melisachew Wudage Chekol, co-supervised with Nabil Layaida (WAM).

6.2.2. Trust in peer-to-peer semantic systems

Participants: Manuel Atencia [Contact], Jérôme Euzenat, Marie-Christine Rousset.

In a semantic peer-to-peer network, peers use separate ontologies and rely on alignments between their ontologies for translating queries. Nonetheless, alignments may be incorrect –unsound or incomplete– and generate flawed translations, thus leading to unsatisfactory answers. We have put forward a trust mechanism that can assist peers to select those peers in the network that are better suited to answer their queries [8]. The trust that a peer has towards another peer depends on a specific query and represents the probability that the latter peer will provide a satisfactory answer. In order to compute trust, we exploit both alignments and peers’ direct experience, and perform Bayesian inference. We have implemented our technique and conducted an evaluation. Experimental results showed that trust values converge as more queries are sent and answers received. Furthermore, the use of trust is shown to improve both precision and recall of query answers.

This work has been developed in collaboration with Marie-Christine Rousset (LIG) in the context of the DataRing project (see §8.1.2).
6. New Results

6.1. Large-scale image search

6.1.1. Aggregating local image descriptors into compact codes

Participants: Matthijs Douze, Hervé Jégou [INRIA Rennes], Patrick Pérez [Technicolor], Florent Perronnin [Xerox RCE], Jorge Sánchez [Xerox RCE], Cordelia Schmid.

In [5] we consolidate and extend earlier results for large-scale image search. Different ways of aggregating local image descriptors into a vector are compared. The Fisher vector, see Figure 1, is shown to achieve better performance than the reference bag-of-visual words approach for any given vector dimension. Furthermore, we jointly optimize dimensionality reduction and indexing in order to obtain a precise vector comparison as well as a compact representation. The evaluation shows that the image representation can be reduced to a few dozen bytes with good search accuracy. Given such small codes, searching a 100 million image dataset takes about 250 ms on one processor core.

![Figure 1. Illustration of the similarity of the Fisher vectors of local image regions despite viewpoint changes.](image)

6.1.2. Searching in one billion vectors: re-rank with source coding

Participants: Laurent Amsaleg [CNRS, IRISA], Matthijs Douze, Hervé Jégou [INRIA Rennes], Romain Tavenard [University Rennes I].

In this work [13] we extend our earlier work [4]. An additional level of processing is added to the product quantizer to refine the estimated distances. It consists in quantizing the difference vector between a point and the corresponding centroid. When combined with an inverted file, this gives three levels of quantization. Experiments performed on SIFT and GIST image descriptors show excellent search accuracy outperforming three state-of-the-art approaches.

6.1.3. Combining attributes and Fisher vectors for efficient image retrieval

Participants: Matthijs Douze, Arnau Ramisa, Cordelia Schmid.

Attributes were recently shown to give excellent results for category recognition. In [9] we demonstrate their performance in the context of image retrieval. We show that combining attributes with Fisher vectors improves performance for retrieval of particular objects as well as categories. Furthermore, we implement an efficient coding technique for compressing the combined descriptor to very small codes. Experimental results show that our approach significantly outperforms the state of the art, even for a very compact representation of 16 bytes per image. We show that attribute features combined with Fisher vectors improve the retrieval of image categories and that those features can supplement text features.

6.1.4. Bag-of-colors for improved image search

Participants: Matthijs Douze, Hervé Jégou [INRIA Rennes], Christian Wengert [Kooaba].
In [19] we investigate the use of color information when used within a state-of-the-art large scale image search system. We introduce a simple color signature generation procedure, used either to produce global or local descriptors. As a global descriptor, it outperforms several state-of-the-art color description methods, in particular the bag-of-words method based on color SIFT. As a local descriptor, our signature is used jointly with SIFT descriptors (no color) to provide complementary information.

6.2. Learning and structuring of visual models

6.2.1. Learning to rank and quadratic assignment

Participants: Thomas Mensink, Jakob Verbeek, Tiberio Caetano [NICTA Canberra].

In [16] we show that the optimization of several ranking-based performance measures, such as precision-at-k and average-precision, is intimately related to the solution of quadratic assignment problems, especially when the score function allows for pairwise label dependencies. Both the task of test-time prediction of the best ranking and the task of constraint generation in estimators based on structured support vector machines can all be seen as special cases of quadratic assignment problems. Although such problems are in general NP-hard, we identify a polynomially-solvable subclass (for both inference and learning) that still enables the modeling of a substantial number of pairwise rank interactions. We show preliminary results on a public benchmark image annotation data set, which indicates that this model can deliver higher performance over ranking models without pairwise rank dependencies. This work was performed during a visit to NICTA Canberra by T. Mensink (March – June, ’11) and J. Verbeek (May ’11).

6.2.2. Learning structured prediction models for interactive image labeling

Participants: Thomas Mensink, Jakob Verbeek, Gabriela Csurka [Xerox RCE].

In [25] we propose structured models for image labeling that take into account the dependencies among the image labels explicitly. These models are more expressive than independent label predictors, and lead to more accurate predictions. While the improvement is modest for fully-automatic image labeling, the gain is significant in an interactive scenario where a user provides the value of some of the image labels. Such an interactive scenario offers an interesting trade-off between 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. 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. See Figure 2.

6.2.3. Modeling spatial layout with Fisher vectors for image categorization

Participants: Frédéric Jurie [University of Caen], Josip Krapac, Jakob Verbeek.

In [15] we introduce an extension of bag-of-words image representations to encode spatial layout. Using the Fisher kernel framework we derive a representation that encodes the spatial mean and the variance of image regions associated with visual words. We extend this representation by using a Gaussian mixture model to encode spatial layout, and show that this model is related to a soft-assign version of the spatial pyramid representation. We also combine our representation of spatial layout with the use of Fisher kernels to encode the appearance of local features. Through an extensive experimental evaluation, we show that our representation yields state-of-the-art image categorization results, while being more compact than spatial pyramid representations. In particular, using Fisher kernels to encode both appearance and spatial layout results in an image representation that is computationally efficient, compact, and yields excellent performance while using linear classifiers.

6.2.4. Unsupervised metric learning for face identification in TV video

Participants: Ramazan Cinbis, Jakob Verbeek, Cordelia Schmid.

The goal of face identification is to decide whether two faces depict the same person or not. In [8] we address the identification problem for face-tracks that are automatically collected from uncontrolled TV video data. Face-track identification is an important component in systems that automatically label characters in TV series or movies based on subtitles and/or scripts: it enables effective transfer of the sparse text-based supervision to other faces. We show that, without manually labeling any examples, metric learning can be effectively used to address this problem. This is possible by using pairs of faces within a track as positive examples, while negative training examples can be generated from pairs of face tracks of different people that appear together in a video frame. In this manner we can learn a cast-specific metric, adapted to the people appearing in a particular video, without using any supervision. Identification performance can be further improved using semi-supervised learning where we also include labels for some of the face tracks. We show that our cast-specific metrics not only improve identification, but also recognition and clustering. See Figure 3.

Figure 3. Projections of face signatures projected to two dimensions, using (a) a metric trained on faces detected in still images, (b) using hand labeled faces detected in videos, (c) a metric trained from face tracking results (no manual labeling). Face signatures of different people are color coded. A good face metric can be learned directly from face tracking results, without using any hand labeled examples.
6.2.5. Large-scale image classification

Participants: Miro Dudik [Yahoo! Research], Zaid Harchaoui, Jerome Malick [INRIA Grenoble, BIPOP Team].

We introduced in [10] 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 regular $\ell_1$-regularization penalty, we propose a simple and provably efficient coordinate descent algorithm. Furthermore, we showed 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 were obtained, where our approach performed significantly better than state-of-the-art approaches for Fisher vectors with 16 Gaussians.

6.3. Human action recognition

6.3.1. Action recognition by dense trajectories


Feature trajectories have shown to be efficient for representing videos. Typically, they are extracted using the KLT tracker or matching SIFT descriptors between frames. However, the quality as well as quantity of these trajectories is often not sufficient. Inspired by the recent success of dense sampling in image classification, in [18] we propose an approach to describe videos by dense trajectories. An overview of our framework is shown in Figure 4. We sample dense points from each frame and track them based on dense optical flow. Our trajectories are robust to fast irregular motions as well as shot boundaries. Additionally, dense trajectories cover the motion information in videos well. We also investigate how to design descriptors to encode the trajectory information. We introduce a novel descriptor based on motion boundary histograms, which is robust to camera motion. This descriptor consistently outperforms other state-of-the-art descriptors, in particular in uncontrolled realistic videos. We evaluate our video description in the context of action classification with a bag-of-features approach. Experimental results show a significant improvement over the state of the art on four datasets of varying difficulty, e.g., KTH, YouTube, Hollywood2 and UCF sports.

Figure 4. Illustration of dense trajectories extraction and description. Left: dense sampling of feature points at multiple scales; middle: tracking feature points with a dense optical flow field; right: descriptors are computed along the trajectory.
6.3.2. Weakly supervised learning of interactions between humans and objects  
Participants: Vittorio Ferrari [ETH Zürich], Alessandro Prest, Cordelia Schmid.

In [7] we introduced a weakly supervised approach for learning human actions modeled as interactions between humans and objects. Our approach is human-centric: we first localize a human in the image and then determine the object relevant for the action and its spatial relation with the human. The model is learned automatically from a set of still images annotated only with the action label. Our approach relies on a human detector to initialize the model learning. For robustness to various degrees of visibility, we build a detector that learns to combine a set of existing part detectors. Starting from humans detected in a set of images depicting the action, our approach determines the action object and its spatial relation to the human. Its final output is a probabilistic model of the human-object interaction, i.e. the spatial relation between the human and the object. We present an extensive experimental evaluation on the sports action dataset from Gupta et al., the PASCAL 2010 action dataset, and a new human-object interaction dataset. In the PASCAL visual object classes challenge 2011 our approach achieved best results on three out of ten action classes and the best result on average over all classes.

6.3.3. Explicit modeling of human-object interactions in realistic videos  
Participants: Vittorio Ferrari [ETH Zürich], Alessandro Prest, Cordelia Schmid.

In [26] we introduced an approach for learning human actions as interactions between persons and objects in realistic videos. Previous work typically represents actions with low-level features such as image gradients or optical flow. In contrast, we explicitly localize in space and track over time both the object and the person, and represent an action as the trajectory of the object wrt to the person position. Our approach relies on state-of-the-art approaches for human and object detection as well as tracking. We show that this results in human and object tracks of sufficient quality to model and localize human-object interactions in realistic videos. Our human-object interaction features capture relative trajectory of the object wrt the human. Experimental results on the Coffee & Cigarettes dataset show that (i) our explicit human-object model is an informative cue for action recognition; (ii) it is complementary to traditional low-level descriptors such as 3D-HOG extracted over human tracks. When combining our human-object interaction features with 3D-HOG features, we show to improve over their separate performance as well as over the state of the art. See Figure 5.

6.3.4. Actom sequence models for efficient action detection  

In [12] we address the problem of detecting actions, such as drinking or opening a door, in hours of challenging video data. We propose a model based on a sequence of atomic action units, termed "actoms", that are characteristic for the action. Our model represents the temporal structure of actions as a sequence of histograms of actom-anchored visual features. Our representation, which can be seen as a temporally structured extension of the bag-of-features, is flexible, sparse and discriminative. We refer to our model as Actom Sequence Model (ASM). Training requires the annotation of actoms for action clips. 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. We show that our ASM method outperforms the current state of the art in temporal action detection.

6.3.5. A time series kernel for action recognition  

In [11] we address the problem of action recognition by describing actions as time series of frames and introduce a new kernel to compare their dynamic aspects. Action recognition in realistic videos has been successfully addressed using kernel methods like SVMs. Most existing approaches average local features over video volumes and compare the resulting vectors using kernels on bags of features. In contrast, we model actions as time series of per-frame representations and propose a kernel specifically tailored for the purpose of action recognition. Our main contributions are the following: (i) we provide a new principled way to compare
the dynamics and temporal structure of actions by computing the distance between their auto-correlations, (ii) we derive a practical formulation to compute this distance in any feature space deriving from a base kernel between frames, and (iii) we report experimental results on recent action recognition datasets showing that it provides useful complementary information to the average distribution of frames, as used in state-of-the-art models based on bag-of-features.

Figure 5. Example results showing the automatically detected human (green) and related object (blue).
Morpheo Team

6. New Results

6.1. Mesh repair with topology control

Participant: Franck Hétéroy.

This work is done in collaboration with Carlos Andújar, Pere Brunet and Álvar Vinacua from Universitat Politecnica de Barcelona, Spain, and has been published in the CAD journal [7]. The purpose is to propose an efficient method to create 2-manifold meshes from real data, obtained as soups of polygons with combinatorial, geometrical and topological noise (see Figure 3). We propose to use a voxel structure called a discrete membrane and morphological operators to compute possible topologies, between which the user chooses.

![Figure 3. Two topologically different 2-manifold mesh repairs, from the same polygon soup.](image)

6.2. Topology computation on simplicial shapes

Participants: Dobrina Boltcheva, Franck Hétéroy.

This work is a part of the BQR project IDEAL (see Section 8.1.1) which is performed in collaboration with Leila de Floriani from the University of Genova in Italy. The main goal of this project is to study non-manifold geometrical models and to find out features allowing to classify these models and criteria for determining their shape. We are interested in non-manifold models such as idealized industrial CAD models, since they are still ill-understood even if they are frequently used in computer graphics and many engineering applications.

We have developed an efficient method to compute the homology of a large (non-manifold) simplicial complex, from the homologies of its sub-complexes. Computed topological invariants play a crucial role in the field of shape description and analysis. This work has been published in the CAD journal [5] and presented at the SIAM conference on geometric and physical modeling (GD/SPM’11).

6.3. Scale Space Representations on Manifolds

Participant: Edmond Boyer.
In collaboration with Radu Horaud and Andrei Zaharescu, we developed a novel approach for the scale-space representations of scalar functions defined over Riemannian manifolds. One of the main interest in such representations stems from the task of 3D modelling where 2D surfaces, endowed with various physical properties, are recovered from images. Multi-scale analysis allows to structure the information with respect to its intrinsic scale, hence enabling a wide range of low-level computations, similar to what is usually used for representing images. In contrast to the Euclidean image domain, where scale spaces can be easily obtained through convolutions with Gaussian kernels, surfaces require a more general approach that must handle non-Euclidean spaces. Such a generalized scale-space framework is the main contribution of this work, which builds on the spectral decomposition available with the heat-diffusion framework to derive a computational approach for representing scalar functions on 2D Riemannian manifolds using an intrinsic scale parameter. In addition, we proposed a feature detector and a region descriptor, based on these representations, extending the widely used DOG detector and HOG descriptor to manifolds. Experiments on real datasets with various physical properties, i.e., scalar functions, demonstrated the validity and the interest of this approach[16].

**Figure 4. Fine-to-coarse representations of scalar functions (color) defined over 2D Riemannian manifolds.**

### 6.4. Topologically-Robust 3D Shape Matching
**Participant:** Edmond Boyer.

3D Shape matching is an important problem in computer vision. One of the major difficulties in finding dense correspondences between 3D shapes is related to the topological discrepancies that often arise due to complex kinematic motions. In this work done in collaboration with Jan Cech, Radu Horaud and Avinash Sharma a shape matching method is proposed that is robust to such changes in topology. The algorithm starts from a sparse set of seed matches and outputs dense matching. We use a shape descriptor based on properties of the heat-kernel and which provides an intrinsic scale-space representation. This descriptor incorporates (i) heat-flow from already matched points and (ii) self diffusion. At small scales the descriptor behaves locally and hence it is robust to global changes in topology. Therefore, it can be used to build a vertex-to-vertex matching score conditioned by an initial correspondence set. This score is then used to iteratively add new correspondences based on a novel seed-growing method that iteratively propagates the seed correspondences to nearby vertices. The matching is farther densified via an EM-like method that explores the congruency between the two shape embeddings. The method is compared with two recently proposed algorithms and we show that we can deal with substantial topological differences between the two shapes[15].

### 6.5. Motion-based segmentation of mesh sequences
**Participants:** Romain Arcila, Franck Hétroy.
Mesh animations, or sequences of meshes, represent a huge amount of data, especially when acquired from scans or videos. In collaboration with the university of Lyon (LIRIS lab), we address the problem of partitioning these sequences, in order to both recover motion information and be able to compress them. Following last year’s method, we proposed this year a second and third motion-based segmentation algorithm, which clusters mesh vertices into static or rigidly moving components (see Figure 5). These methods are based on spectral clustering of the vertex transformations and are more robust and general than the previous one. This work has been submitted for publication to a journal, and is part of the PhD thesis of Romain Arcila [1].

Figure 5. Temporally varying segmentation of a mesh sequence, into rigidly moving components. Right hand and arm are merged at some time since they start to move accordingly. They are later split, when they start to follow different motions.

6.6. Surface Flow

Participants: Antoine Letouzey, Benjamin Petit, Jean-Sébastien Franco, Edmond Boyer.

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[14], [18].

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[13].
6.7. Learning Temporally Consistent Rigidities

Participants: Jean-Sébastien Franco, Edmond Boyer.

We present a novel probabilistic framework for rigid tracking and segmentation of shapes observed from multiple cameras. Most existing methods have focused on solving each of these problems individually, segmenting the shape assuming surface registration is solved, or conversely performing surface registration assuming shape segmentation or kinematic structure is known. We assume no prior kinematic or registration knowledge except for an over-estimate $k$ of the number of rigidities in the scene, instead proposing to simultaneously discover, adapt, and track its rigid structure on the fly. We simultaneously segment and infer poses of rigid subcomponents of a single chosen reference mesh acquired in the sequence. We show that this problem can be rigorously cast as a likelihood maximization over rigid component parameters. We solve this problem using an Expectation Maximization algorithm, with latent observation assignments to reference vertices and rigid parts. Our experiments on synthetic and real data show the validity of the method, robustness to noise, and its promising applicability to complex sequences. This work was presented at the CVPR 2011 conference [11].
6.8. Ontology-based mesh segmentation

**Participants:** Sahar Hassan, Franck Hétroy.

Patient-specific 3D virtual models of anatomical organs are becoming more and more useful in medicine, for instance for diagnosis or follow-up care purposes. These models are usually created from 2D scan or MRI images. However, small or thin geometrical features, such as ligaments, are sometimes not visible on these images. We propose to use an anatomical ontology, called MyCorporisFabrica [http://www.mycorporisfabrica.org/](http://www.mycorporisfabrica.org/), to add missing parts to reconstructed virtual organs. This ontology describes definitions of and relationships between organs: e.g., femur is part of the leg. The first step towards the full achievement of this process is to segment virtual models, often represented by 2D meshes, into meaningful parts. In our case, “meaningful” means “related to the ontology”: each part should refer to an organ defined in the ontology. An algorithm to decompose a given organ into sub-organs according to the ontology has been proposed in the PhD thesis of Sahar Hassan [2]: first, we approximate organ shapes by geometric primitives, then we segment a given organ mesh by optimizing objective functions which are related to these primitives.

6.9. Detection and quantification of brain aneurysms

**Participants:** Sahar Hassan, Franck Hétroy.

Aneurysms are excrescences on blood vessels. They can break, letting the blood propagate outside the vessel, which often leads to death. In some cases, the blood clots sufficiently fast so that people survive. However, a neurosurgeon or a neuroradiologist should intervene very quickly in order to repair the vessel before the aneurysm breaks once more.

The purpose of our research is to help neurosurgeons and neuroradiologists to plan surgery, by giving them quantitative information about the size, shape and geometry position of aneurysms. This work was part of the PhD of Sahar Hassan [2], and has presented at the International Conference on Computer Analysis of Images and Patterns (CAIP) [12]. The method we propose first extracts a centered skeleton from the input voxel set of the vascular tree, then detects aneurysms by studying variations of vessel diameters along the skeleton. The name of an aneurysm-carrying vessel is also given thanks to a partial graph matching technique, and accurate measures to decide the treatment are provided.

6.10. Dimensionality reduction for character animation

**Participants:** Maxime Tournier, Lionel Reveret.

This work investigates and proposes a mathematical framework to perform statistical analysis and dimensionality reduction on rotational trajectories derived from motion capture data. Motion capture data consists in a set of trajectories in the space of 3D rotations (SO(3)) and as such do not present properties of an Euclidian space. Consequently there is no easy to way to apply standard dimensionality reduction techniques on these data. Using the formalism of exponential maps and Principle Geodesics Analysis (PGA), it has been shown that it is possible to rigorously derive a dimensionality reduction analysis on such data. This reduction can be typically applied for compression of motion capture data and probabilistic implementation of the Inverse Kinematics problem. This approach has shown good properties in the context of physically-based animation with a Lagrangian formulation of rigid body dynamics coupled with geometric integrators. These integrators allow a good preservation of momentum using only first order equations, achieving both real-time and high level of realism. These works were developed through the PhD thesis of Maxime Tournier [4]. Early development of PGA on motion capture data had been published at Eurographics in 2009. Its integration into a GPLVM framework has been published this year in the IEEE CG&A journal [9]. Its extension into the context of physically-based animation is currently under preparation for publication.

6.11. Animation of quadrupeds locomotion

**Participant:** Lionel Reveret.
Following a study on locomotion of quadrupeds by a team in the National Museum of Natural History (MNHN), a new theory on motion planning has been proposed. This theory, the Antero-Posterior Sequences (APS), allows a characterization of the sequence of foot placement for quadrupeds for all regular gaits with very few parameters, as well as transition between gaits, starting from stop to full gallop. In collaboration with the MNHN and the robotics department of the University of Versailles-Saint Quentin en Yvelines (UVSQ), a rigorous software implementation has been specified and developed. This software allows automatically generating foot planning of quadrupeds locomotion according to a desired speed transition. Co-workers for this project were Ludovic Maes and Anick Abourachid at the MNHN and Vincent Hugel at the UVSQ. A patent has been written and finalized for this project.

In parallel, collaboration on physical simulation of quadrupeds locomotion has been carried on with Stelian Coros (previously at University of British Columbia (UBC), now at Disney Research) and Michiel van de Panne (UBC). Automatic video analysis of dog walking, trotting and running has been used to optimize parameters of physical controllers. This work has been published at SIGGRAPH 2011 [6].
6. New Results

6.1. Calibration of a mixed camera system

An approximately Euclidean representation of the visible scene can be obtained directly from a range, or
time-of-flight, camera. An uncalibrated binocular system, in contrast, gives only a projective reconstruction of
the scene. This paper analyzes the geometric mapping between the two representations, without requiring an
intermediate calibration of the binocular system. The mapping can be found by either of two new methods, one
of which requires point correspondences between the range and colour cameras, and one of which does not. It
is shown that these methods can be used to reproject the range data into the binocular images, which makes it
possible to associate high resolution colour and texture with each point in the Euclidean representation.

6.2. Computation of scene flow

A simple seed growing algorithm for estimating scene flow in a stereo setup is presented. Two calibrated and
synchronized cameras observe a scene and output a sequence of image pairs. The algorithm simultaneously
computes a disparity map between the image pairs and optical flow maps between consecutive images. This,
together with calibration data, is an equivalent representation of the 3D scene flow, i.e. a 3D velocity vector
is associated with each reconstructed point. The proposed method starts from correspondence seeds and
propagates these correspondences to their neighborhood. It is accurate for complex scenes with large motions
and produces temporally-coherent stereo disparity and optical flow results. The algorithm is fast due to inherent
search space reduction. An explicit comparison with recent methods of spatiotemporal stereo and variational
optical and scene flow is provided.

6.3. 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.
6.4. A differential model for the complex cell

The receptive fields of simple cells in the visual cortex can be understood as linear filters. These filters can be modelled by Gabor functions, or by Gaussian derivatives. Gabor functions can also be combined in an energy model of the complex cell response. This work proposes an alternative model of the complex cell, based on Gaussian derivatives. It is most important to account for the insensitivity of the complex response to small shifts of the image. The new model uses a linear combination of the first few derivative filters, at a single position, to approximate the first derivative filter, at a series of adjacent positions. The maximum response, over all positions, gives a signal that is insensitive to small shifts of the image. This model, unlike previous approaches, is based on the scale-space theory of visual processing. In particular, the complex cell is built from filters that respond to the 2-D differential structure of the image. The computational aspects of the new model are studied in one and two dimensions, using the steerability of the Gaussian derivatives. The response of the model to basic images, such as edges and gratings, is derived formally. The response to natural images is also evaluated, using statistical measures of shift insensitivity. The relevance of the new model to the cortical image-representation is discussed.

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 that 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 conjugate expectation-maximization algorithm. The convergence properties of the proposed algorithm are thoroughly investigated. Several local/global optimization techniques are proposed in order to increase its convergence speed. Two initialization strategies are proposed and compared. A consistent model-selection criterion is proposed. The algorithm and its variants are tested and evaluated within the task of 3D localization of several speakers using both auditory and visual data.
6. New Results

6.1. A Lightweight Augmented Virtuality System for Providing a Faithful and Spatially Manipulable Visual Hand Representation

**Participants:** Sabine Coquillart, Olivier Martin, Andreas Pusch.

We introduced the technical foundations of a system designed to embed a lightweight, faithful and spatially manipulable representation of the user’s hand into an otherwise virtual world - Augmented Virtuality (AV). A highly intuitive control during pointing-like near space interaction can be provided to the user, as well as a very flexible means to experimenters, in a variety of contexts. Our approach essentially relies on stereoscopic video see-through Augmented Reality (AR) technology and a generic, extendible framework for managing 3-D visual hand displacements. Research from human-computer interaction, perception and motor control has contributed to the elaboration of our proposal which combines a) acting in co-location, b) avoiding occlusion violations by assuring a correct scene depth ordering and c) providing a convincing visual feedback of the user’s hand. This system has already successfully been used in one case and further promising applications are studied [17], [18].

6.2. Effects of Hand Feedback Fidelity on Near Space Pointing Performance and User Acceptance

**Participants:** Sabine Coquillart, Olivier Martin, Andreas Pusch.

We conducted an experiment to test the effects of different hand representations on near space pointing performance and user preference. Subjects were presented with varying levels of hand realism, including real hand video, a high and a low level 3D hand model and an ordinary 3D pointer arrow. Behavioural data revealed that an abstract hand substitute like a 3D pointer arrow leads to significantly larger position estimation errors in terms of lateral target overshooting when touching virtual surfaces with only visual hand movement constraints. Further, questionnaire results show that a higher fidelity hand is preferred over lower fidelity representations for different aspects of the task [18].
6. New Results

6.1. Multimedia Models and Formats

6.1.1. SMIL timesheets

With the advent of HTML5 and its support in most popular browsers, HTML is becoming an important multimedia language. Video and audio can now be embedded in HTML pages without worrying about the availability of plugins. However, a major issue is to specify the dynamic behavior of documents (user interactions, timing and synchronization with continuous contents). This is done usually by writing (often complex) scripts, which require programming skills from the authors.

To address this issue, we have created the timesheets.js library, a scheduler that allows HTML documents to be animated and synchronized in a purely declarative way. This work is based on the SMIL Timing and Synchronization module and the SMIL Timesheets specification, with a few extensions.

The library is implemented in JavaScript, which makes it usable in any browser. Authors can specify the dynamic behavior of HTML5 (+CSS3) documents. They can thus develop multimedia applications without writing a single line of JavaScript. Timesheets can also be used with other XML document languages, such as SVG for instance. This approach was validated in a class with students learning web multimedia.

6.1.2. Multimedia content adaptation

Multimedia documents may have to be played on multiple devices such as mobile phones, tablets, desktop computers, set-top boxes, etc. Usage and platform diversity requires documents to be adapted according to execution contexts, sometimes unpredictable at design time. In a joint work with project-team Exmo, we have designed a semantic adaptation framework for multimedia documents. This framework captures the semantics of document composition and transforms the relations between media objects according to adaptation constraints.

6.2. XML Processing

Mature results about XML processing were obtained along three main directions: the formalization and implementation for checking the impact of schema evolution on validation and queries; logical extensions supporting a notion of counting and the shuffle operator in trees; and the decision of a subtyping relation for a very expressive type algebra supporting a notion of polymorphism.

In addition, preliminary results were obtained on the definition of a rigorous logical framework for the static analysis of semantic web languages, on the static analysis of cascading style sheets, and on the equipment of an IDE with new static analysis features for XQuery.

6.2.1. Impact of XML schema evolution

In the ever-changing context of the web, XML schemas continuously change in order to cope with the natural evolution of entities they describe. Schema changes have important consequences. First, existing documents valid with respect to the original schema are no longer guaranteed to fulfill the constraints described by the evolved schema. Second, the evolution also impacts programs manipulating documents whose structure is described by the original schema.

---

1 http://wam.inrialpes.fr/timesheets/
We have proposed a unifying framework for determining the effects of XML Schema evolution both on the validity of documents and on queries [2]. The system is very powerful in analyzing various scenarios in which forward/backward compatibility of schemas is broken, and in which the result of a query may not be anymore what was expected. Specifically, the system offers a predicate language which allows one to formulate properties related to schema evolution. The system then relies on exact reasoning techniques to perform a fine-grained analysis. This yields either a formal proof of the property or a counter-example that can be used for debugging purposes. The system has been fully implemented and tested with real-world use cases, in particular with the main standard document formats used on the web, as defined by W3C. The system identifies precisely compatibility relations between document formats. In case these relations do not hold, the system can identify queries that must be reformulated in order to produce the expected results across successive schema versions.

6.2.2. Counting in trees

A major challenge of query language design is the combination of expressivity with effective static analyses such as query containment. In the setting of XML, documents are seen as finite trees, whose structure may additionally be constrained by type constraints such as those described by an XML schema. We have considered the problem of query containment in the presence of type constraints for a class of regular path queries extended with counting and interleaving operators [1]. The counting operator restricts the number of occurrences of children nodes satisfying a given logical property. The interleaving operator provides a succinct notation for describing the absence of order between nodes satisfying a logical property. We have proposed a logic supporting these operators, which can be used to solve common query reasoning problems such as satisfiability and containment of queries in exponential time [4].

6.2.3. Typing higher-order programs

We have considered a type algebra equipped with recursive, product, function, intersection, union, and complement types together with type variables and universal quantification over them. We have defined the subtyping relation between such type expressions, and have proved its decidability [9].

This has solved an open problem that was attracting a considerable research effort. The novelty, originality and strength of our solution reside in introducing a logical modeling for the semantic subtyping framework. We have modeled semantic subtyping in a tree logic and use a satisfiability-testing algorithm in order to decide subtyping. We have shown how the subtyping relation can be decided in EXPTIME. We have reported on practical experiments made with a full implementation of the system. This has provided a powerful polymorphic type system aiming at maintaining full static type-safety of functional programs that manipulate trees, even with higher-order functions, which is particularly useful in the context of XML.

6.2.4. Detection of inconsistent paths and dead code in XML IDEs

One of the challenges in web software development is to help achieving a good level of quality in terms of code size and runtime performance, for increasingly popular domain specific languages such as XQuery. We have presented an IDE equipped with static analysis features for assisting the programmer [8]. These features are capable of identifying and eliminating dead code automatically. The tool is based on newly developed formal programming language verification techniques, which are now mature enough to be introduced in the process of software development.

6.2.5. Static analysis of semantic web languages

We work with the Exmo project-team on the static analysis of semantic web languages such as RDF, OWL and SPARQL by investigating modal logics over graphs. We seek to build a rigorous logical reasoning framework based on $\mu$-calculus adapted for the web semantic languages [7] [11]. In particular, we studied the containment problem for SPARQL queries: determining whether, for any graph, the answers to a query are contained in those of another query. Our approach consists in encoding RDF graphs as transition systems and queries as $\mu$-calculus formulas and then reducing the containment problem to testing satisfiability in the logic.
6.2.6. Static analysis of style sheets

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 proposed a novel approach to fill this lack [14] by analyzing CSS style sheets using the same logic and compile-time verification technique we use for other XML problems. We have developed an original tool based on our XML Reasoning Solver (see section 5.2). 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.

6.3. Multimedia Authoring

6.3.1. C2M project

The C2M project (see section 7.2.2) aims at developing industrial solutions that allow multimedia developers to achieve mass production with high quality results. It uses the SCENARI platform for document production and we have proposed a solution for dealing with multimedia content in such a framework [16]. Indeed, automatic tools are not always sufficient for generating high quality documents; manual editing of documents in their publishing format is often necessary to tune a number of details.

Our approach consists in providing a post-editing service to allow authors to adjust their multimedia presentations directly on the final form of documents. The first step is to provide a web rendering engine based on the latest advances in web standards, as described in section 6.1.1. The second step consists in designing web-aware authoring tools based on this library, thus providing authors with direct editing services for producing high quality multimedia documents while preserving the advantages of using an XML production workflow. We have developed a prototype of this authoring tool in which all editing templates are described with XUL (XML User interface Language) and XBL (XML Binding Language) elements that we have defined for handling time-based content and widgets (timeContainer, timeNode, timeLine, etc.).

With such a solution, we can combine two worlds: a semantic-oriented authoring approach, as provided by an XML document workflow, and a direct web-based editing system. The first guarantees homogeneous rendering while the latter enables direct adjustments on the final form of the document. Bridging these two worlds is made possible by using declarative web languages (namely HTML5, SMIL and CSS) and implementing their timing part in the browser (with the Timesheets.js library, see section 6.1.1). The authoring components are directly mapped to the document structures.

6.3.2. On-line editing of multimedia web content

In cooperation with EPFL (Lausanne) we have continued to explore the concept of template-driven editing for XML multimedia contents (see section 3.3.2). This year, we have carried out more experiments with very different types of contents, including structured documents, factual data, and multimedia objects [17].

These experiments have been done with the AXEL library developed by EPFL, based on our joint work on template languages. AXEL is an innovative client-side authoring tool that runs in the browser for editing XML documents, driven by an XTiger template. It allows average web users to easily edit XML content on web servers with no specific knowledge of XML. Our experiments have shown that the template-based approach significantly enhances the ability for web users to directly feed various applications with structured content.
6.4. Augmented Environments

A large part of the research on augmented environments specializes in the use of visual media. In WAM, we focus on the use of audio media and we put a strong emphasis on mobility.

We have developed the first indoor augmented reality audio navigation application running on personal AR devices such as mobile phones. The main idea behind the development of this application was a joint use of three concepts:

- Continuous localization by using embedded sensors together with physiological models of walking and assumptions about walking in structured indoor environments.
- OpenStreetMap Indoor Mapping used for map-aided positioning, assistive routing for visually impaired people, and environmental queries through audio panoramics.
- Guidance and navigation through AR audio, both 3D and environmental, with mixing of synthetic and natural sounds and support for timely audio information presentation.

We have demonstrated that these concepts are inter-dependent [12], and that bringing them together is a way to find new solutions to problems which are difficult to tackle when looking at them separately. These three concepts are implemented with web technologies we use XML languages and XML tools for interactive audio, building modeling, and personal navigation module configuration. This enables easy authoring of sound objects or audio icons used for building sonification [15], indoor navigation maps and panoramics, and walking models. Adaptability of navigation to preferences of people is based on the concept of audio stylesheets for OpenStreetMap data rendering, XML configuration of the Pedestrian Dead Reckoning module, and assistive routing specification.

We have developed two mobile browsers and a framework for generic navigation:

1. The Mixed Reality Browser [10] that we have developed can display PoI content either remotely through panoramics with spatialized audio, or on-site by walking to the corresponding place. MRB is the only browser of geolocalized data to use a declarative XML format for PoIs, panoramics, 3D audio and to be based on HTML5 both for the iconic and full information content of PoIs. MRB can be used for any kind of augmented reality visits. A cultural heritage visit of Grenoble (see section 7.1.2) has been realized with the tourist office of Grenoble and the CCSTI (Centre de Culture Scientifique Technique et Industrielle de Grenoble).

2. The Pedestrian Way Browser that we have developed can be used for indoor-outdoor navigation with assistive audio technology for visually impaired people on pedestrian ways with precise geospatial description. Its main characteristic is to be based entirely on the OpenStreetMap XML format for the representation of the route. We anticipate that in the context of the european project Venturi (see section 7.3.1), we will have a convergence of the MRB and PWB, allowing visually impaired people to undertake cultural heritage visit. An demonstration showing the use of the PWB in a structured outdoor environment is available online: http://www.youtube.com/watch?v=h2b8yfCauZ8

3. We have created an extensible client-server framework named TARA which allows navigation on an OpenStreetMap XML graph (indoor and outdoor) by computing routes in real-time. User preferences, like stairs versus lift, are supported through a ponderation of paths in the routing algorithm. The client is an HTML5 running in the browser on mobiles and desktop computers. It can therefore be used for simulation, to test or learn a route before the navigation in the real world. The user interface is based on three modalities, touch, audio and visual and can be operated by visually impaired people through VoiceOver using only touch and audio. Localization through embedded or external sensors is not mandatory as step by step instructions can be accessed through touch modality. The server is a full REST server (Sling-Apache) giving priority to the representation of geospatial resources and allowing environmental queries through the use of XQuery.