Activity Report 2011

Section New Results
ALGORITHMS, PROGRAMMING, SOFTWARE AND ARCHITECTURE

1. ALF Project-Team ................................................................. 4
2. CAIRN Project-Team ............................................................. 13
3. CELTIQUE Project-Team .................................................... 23
4. ESPRESSO Project-Team .................................................... 26
5. S4 Project-Team ................................................................. 35
6. TASC Project-Team ............................................................ 39
7. VERTECS Project-Team ..................................................... 43

APPLIED MATHEMATICS, COMPUTATION AND SIMULATION

8. ASPI Project-Team ............................................................... 48
9. I4S Team ............................................................................ 50
10. IPSO Project-Team ............................................................ 52

COMPUTATIONAL SCIENCES FOR BIOLOGY, MEDICINE AND THE ENVIRONMENT

11. FLUMINANCE Project-Team ................................................ 60
12. SAGE Project-Team ............................................................ 69
13. SERPICO Team ................................................................. 77
14. SYMBIOSE Project-Team .................................................. 81
15. VISAGES Project-Team ...................................................... 85

NETWORKS, SYSTEMS AND SERVICES, DISTRIBUTED COMPUTING

16. ACES Project-Team ........................................................... 95
17. ASAP Project-Team .......................................................... 99
18. ASCOLA Project-Team ...................................................... 109
19. ATLANMOD Team .......................................................... 116
20. CIDRE Project-Team ......................................................... 118
21. DIONYSOS Project-Team .................................................. 123
22. DISTRIBUTCOM Project-Team .................................... 130
23. KERDATA Team ............................................................... 139
24. MYRIADS Team .............................................................. 146
25. TRISKELL Project-Team .................................................... 151

PERCEPTION, COGNITION, INTERACTION

26. DREAM Project-Team ........................................................ 154
27. LAGADIC Project-Team ..................................................... 158
28. METISS Project-Team ....................................................... 161
29. MIMETIC Team ............................................................... 170
30. TEMICS Project-Team ...................................................... 178
31. TEXMEX Project-Team ..................................................... 190
32. VR4I Team .................................................................. 200
6. New Results

6.1. Processor Architecture

Participants: Damien Hardy, Pierre Michaud, Nathanaël Prémillieu, Ricardo Andrés Velasquéz, Luis-Germán García Morales, Bharath Narasimha Swamy, André Seznec.

Our research in computer architecture covers memory hierarchy, branch prediction, superscalar implementation, as well as SMT and multicore issues.

This year, we have also initiated new research directions within the context of the ERC DAL project.

6.1.1. Null block management on the memory hierarchy

Participant: André Seznec.

It has been observed that some applications manipulate large amounts of null data. Moreover, these zero data often exhibit high spatial locality. On some applications more than 20% of the data accesses concern null data blocks. To reduce the pressure on main memory, we have proposed a hardware compressed memory that only targets null data blocks, the decoupled zero-compressed memory [27]. Borrowing some ideas from the decoupled sectored cache [20], the decoupled zero-compressed memory, or DZC memory, manages the main memory as a decoupled sectored set-associative cache where null blocks are only represented by a validity bit. Our experiments show that for many applications, the DZC memory allows to artificially enlarge the main memory, i.e. it reduces the effective physical memory size needed to accommodate the working set of an application without excessive page swapping. Moreover, the DZC memory can be associated with a ZCA cache [5] to manage null blocks across the whole memory hierarchy. For some applications, such a management significantly decreases the memory traffic and therefore can significantly improve performance.

This work corresponds to the PhD of Julien Dusser defended in December 2010.

6.1.2. Emerging memory technologies

Participant: André Seznec.

Phase change memory (PCM) technology appears more scalable than DRAM technology. As PCM exhibits access time slightly longer but in the same range as DRAMs, several recent studies have proposed to use PCMs for designing main memory systems. Unfortunately, PCM technology suffers from a limited write endurance; typically each memory cell can only be written a large but still limited number of times (10 millions to 1 billion writes are reported for current technology). Research proposals have essentially focused their attention on designing memory systems that will survive the average behavior of conventional applications. However, PCM memory systems should be designed to survive worst-case applications, i.e., malicious attacks targeting the physical destruction of the memory through overwriting a limited number of memory cells.

In 2010, we have proposed the first design of a secure PCM-based main memory that would by construction survive overwrite attacks [19]. This secure PCM-based main memory requires a significant read and write extra memory traffic (an extra memory write per 8 demand memory writes) on all applications. Concurrent proposals require even higher extra read and write memory traffic. In collaboration with a research group from IBM, we have proposed a hardware method to detect malicious overwrite attacks on the main memory, thus limiting the memory traffic overhead on non-malicious applications [32].

6.1.3. Microarchitecture exploration of control flow reconvergence

Participants: Nathanaël Prémillieu, André Seznec.
After continuous progress over the past 15 years [18], [17], the accuracy of branch predictors seems to be reaching a plateau. Other techniques to limit control dependency impact are needed. Control flow reconvergence is an interesting property of programs. After a multi-option control-flow instruction (i.e., either a conditional branch or an indirect jump including returns), all the possible paths merge at a given program point: the reconvergence point.

Superscalar processors rely on aggressive branch prediction, out-of-order execution and instruction level parallelism for achieving high performance. Therefore, on a superscalar core, the overall speculative execution after the mispredicted branch is cancelled, leading to a substantial waste of potential performance. However, deep pipelines and out-of-order execution induce that, when a branch misprediction is resolved, instructions following the reconvergence point have already been fetched, decoded and sometimes executed. While some of this executed work has to be cancelled since data dependencies exist, cancelling the control independent work is a waste of resources and performance. We have proposed a new hardware mechanism called SYRANT, SYmmetric Resource Allocation on Not-taken and Taken paths, addressing control flow reconvergence at a reasonable cost. Moreover, as a side contribution of this research we have shown that, for a modest hardware cost, the outcomes of the branches executed on the wrong paths can be used to guide branch prediction on the correct path.

6.1.4. Confidence estimation for the TAGE predictor

Participant: André Seznec.

For the past 15 years, it has been shown that confidence estimation of branch prediction (i.e., estimating the probability of correct or incorrect prediction) can be used for various usages such as fetch gating or throttling for power saving or for controlling resource allocation policies in an SMT processor. In many proposals, using extra hardware and particularly storage tables for branch confidence estimators has been considered as a worthwhile silicon investment.

The TAGE predictor, presented in 2006 [18], is so far considered as the state-of-the-art conditional branch predictor. We have shown that very accurate confidence estimations can be done for the branch predictions realized by the TAGE predictor by simply observing the outputs of the predictor tables. Many confidence estimators proposed in the literature only discriminate between high confidence predictions and low confidence estimations. It has been recently pointed out that a more selective confidence discrimination could be useful. The observation of the outputs of the predictor tables is sufficient to grade the confidence in the branch predictions with a very good granularity. Moreover a slight modification of the predictor automaton allows to discriminate the prediction in three classes, low-confidence (with a misprediction rate in the 30% range), medium confidence (with a misprediction rate in 8-12% range) and high confidence (with a misprediction rate lower than 1%) [37].

6.1.5. Improving branch prediction accuracy

Participant: André Seznec.

The TAGE predictor [18] is often considered as state-of-the-art in conditional branch predictors proposed by academia. For the 3rd championship branch prediction, we have further improved its accuracy by augmenting it with small side predictors (Loop predictor, Statical Corrector Predictor, Immediate Update Mimicker) [34]. This predictor won the conditional branch track of the 3rd championship branch prediction. In order to further argue for real hardware implementation of the TAGE predictor, we have presented several propositions to reduce the complexity of its hardware design, to reduce its energy consumption [36] and further improve branch accuracy. On a hardware implementation of a conditional branch predictor, the predictor tables are updated at retire time. A retired branch normally induces three accesses to the branch predictor tables: read at prediction time, read at retire time and write for the update. We show that in practice, the TAGE predictor accuracy would not be significantly impaired by avoiding a systematic second read of the prediction tables at retire time for correct prediction. Combined with the elimination of silent updates, this significantly reduces the number of accesses to the predictor. Furthermore, we present a technique allowing to implement the TAGE predictor tables as bank-interleaved structures using single-port memory components. This significantly
reduces the silicon footprint of the predictor as well as its energy consumption without significantly impairing its accuracy.

Correctly predicting the indirect branches has become critical with the introduction of object oriented programming, java programming as well as with the renewed importance of interpreters. The ITTAGE indirect branch predictor was introduced in [15]. Three versions of the ITTAGE predictor were presented at the indirect branch track at the championship branch prediction by three different teams, and secured the three first places. Our proposition [35] won the championship.

6.1.6. Hardware acceleration of sequential loops

Participant: Pierre Michaud.

In a decade it will be possible to put on a single chip several hundreds of superscalar cores. A simple application of Amdahl's law shows that it will make sense to dedicate to sequential performance the silicon area and power budget corresponding to that of several tens, or perhaps several hundreds of conventional superscalar cores. This will lead to a sequential accelerator which will be used to accelerate sequential programs and sequential code sections in parallel programs. The question is, what will this sequential accelerator look like? In a previous work, we have proposed a possible solution for implementing a sequential accelerator, which is to implement a superscalar core with a very "aggressive" microarchitecture and design, and to replicate this core and migrate the execution periodically on the replicas to keep the temperature resulting from the high power density under control [11]. However, future sequential accelerators will probably rely on a combination of several techniques, some already known, some yet to be invented.

We have started exploring a new solution for sequential acceleration, the hardware acceleration of dynamic loops, which are periodic sequences of dynamic instructions. A loop accelerator sits beside a conventional superscalar core and is specialized in executing dynamic loops [40]. Dynamic loops are detected and accelerated automatically, without help from the programmer or the compiler. The execution is migrated from the superscalar core to the loop accelerator when a dynamic loop is detected, and back to the superscalar core when a loop exit condition is encountered. Our simulations show that about one third of all the instructions executed by the SPEC CPU2006 benchmark suite belong to dynamic loops with a length of several thousands dynamic instructions, or more. The loop body size is quite diverse, ranging from a few instructions to several hundreds.

We have described a possible loop accelerator microarchitecture that exploits loop properties and avoids the main bottlenecks of conventional superscalar microarchitectures. Our preliminary study demonstrates significant global speedup on some benchmarks, with a local acceleration for loops typically around 2. Our future research on loop acceleration will explore the solution space for obtaining greater performance speedups.

6.1.7. Exploiting confidence in SMT processors

Participants: Pierre Michaud, André Seznec.

Simultaneous multithreading (SMT) [57] processors dynamically share processor resources between multiple threads. The hardware allocates resources to different threads. The resources are either managed explicitly through setting resource limits to each thread or implicitly through placing the desired instruction mix in the resources. In this case, the main resource management tool is the instruction fetch policy which must predict the behavior of each thread (branch mispredictions, long-latency loads, etc.) as it fetches instructions.

We propose the use of Speculative Instruction Window Weighting (SIWW) [25] to bridge the gap between implicit and explicit SMT fetch policies. SIWW estimates for each thread the amount of outstanding work in the processor pipeline. Fetch proceeds for the thread with the least amount of work left. SIWW policies are implicit as fetch proceeds for the thread with the least amount of work left. They are also explicit as maximum resource allocation can also be set. SIWW can use and combine virtually any of the indicators that were previously proposed for guiding the instruction fetch policy (number of in-flight instructions, number of low confidence branches, number of predicted cache misses, etc.). Therefore, SIWW is an approach to design SMT fetch policies, rather than a particular fetch policy.
Targeting fairness and throughput is often contradictory and a SMT scheduling policy often optimizes only one performance metric with the sacrifice of the other metric. Our simulations show that the SIWW fetch policy can achieve at the same time state-of-the-art throughput, state-of-the-art fairness and state-of-the-art harmonic performance mean.

As a side contribution of this study, we have published a study on fairness metrics for SMT processors and multicore systems [24]. This study was done in collaboration with Hans Vandierendonck from the University of Ghent.

6.1.8. Analytical model to estimate the interaction between hardware faults on caches and predictors

Participant: Damien Hardy.

This research was undertaken during Damien Hardy’s stay in the Computer Architecture group of the University of Cyprus (June-August 2011).

Technology trends suggest that in tomorrow’s computing systems, failures will become a commonplace due to many factors, and the expected probability of failure will increase with scaling. Faults can result in execution errors (e.g., on caches) or simply in performance loss (e.g., predictors). Although faults can occur anywhere in the processor, the performance implications of a faulty cell vary depending on how the array is used in a processor.

A direction to determine the impact on performance due to permanent faulty cells is to predict the performance vulnerability by using analytical models. Such models, studied at the University of Cyprus, are representative for the average performance and its probability distribution. So far, analytical models have been defined to determine the impact on performance of faulty mechanisms such as caches and predictors in isolation without any interactions between them.

On the other side, in the real-time systems community, caches and predictors have been intensively studied to estimate the worst-case execution time of application by using static analysis. The ongoing research aims at defining an analytical model of performance that captures the effects of faults on both caches and predictors. This analytical model will be useful to predict future processors performance vulnerability to faults and to determine the benefits in terms of performance of reliability mechanisms.

6.1.9. Hardware support for transactional memory

Participants: Mridha-Mohammad Waliullah, André Seznec.

Parallel programming has become immensely important to harness the power of today’s many core CPU. Over several years, a lot of efforts has been laid out to make parallel programming easier. Transactional memory (TM) has come out as an infrastructure that promises to simplify parallel programming. Implementation of TM in hardware is carried out to get higher performance, which is referred to as hardware transactional memory (HTM). We have focused mainly into two issues related to HTM: (1) exploring TM benchmarks to better understand the performance bottlenecks and (2) exploring innovative techniques that can streamline common case transactional execution to achieve higher performance [38].

This work was done in the framework of the Ercim postdoc stay (01/04/11 to 30/11/11) of Mridha-Mohammad Waliullah.

6.1.10. Microarchitecture research initiated in the DAL project

Participants: Pierre Michaud, Luis-Germán García Morales, Bharath Narasimha Swamy, André Seznec.
Multicore processors have now become mainstream for both general-purpose and embedded computing. Instead of working on improving the architecture of the next generation multicore, with the DAL project, we deliberately anticipate the next few generations of multicores. While multicores featuring 1000s of cores might become feasible around 2020, there are strong indications that sequential programming style will continue to be dominant. Even future mainstream parallel applications will exhibit large sequential sections. Amdahl’s law indicates that high performance on these sequential sections is needed to enable overall high performance on the whole application. On many (most) applications, the effective performance of future computer systems using a 1000-core processor chip will significantly depend on their performance on both sequential code sections and single threads.

We envision that, around 2020, the processor chips will feature a few complex cores and many (may be 1000’s) simpler, more silicon and power effective cores.

In the DAL research project, we will explore the microarchitecture techniques that will be needed to enable high performance on such heterogeneous processor chips. Very high performance will be required on both sequential sections, -legacy sequential codes, sequential sections of parallel applications-, and critical threads on parallel applications, -e.g. the main thread controlling the application. Our research will focus on enhancing single processes performance.

On the microarchitecture side, we will explore both a radically new approach, the sequential accelerator [11], and more conventional processor architectures. We will also study how to exploit heterogeneous multicore architectures to enhance sequential thread performance. Two PhD thesis have been initiated on these topics at fall 2011.

6.2. Compiler, vectorization, interpretation

Participants: Erven Rohou, David Yuste, André Seznec.

The usage of the bytecote-based languages such as Java has been generalized in the past few years. Applications are now very large and are deployed on many different platforms, since they are highly portable. With the new diversity of multicore platforms, functional, but also performance portability will become the major issue in the next 10 years. Therefore our research effort focuses on efficiently compiling towards bytecodes and on efficiently executing the bytecodes through JIT compilation or through direct interpretations.

6.2.1. Iterative and JIT compilation

Participants: Erven Rohou, David Yuste.

Over the last decade, iterative compilation has been an attempt to overcome the difficulty to generate extremely optimized code by letting the compilers explore many alternatives to select the best one. In this research, we extend previous work in the direction of portability. Future processors will be increasingly diverse and heterogenous, and portability is likely to be attained thanks to a bytecode format and JIT compilers. We explore how iterative compilation performed offline can generate useful information to allow the online JIT compiler to generate efficient code at very limited cost.

Part of this research is done in collaboration with STMicroelectronics, in the context of the Nano2012 Mediacom project.

6.2.2. Split vectorization

Participants: Erven Rohou, David Yuste, André Seznec.

We attempt to reconcile two apparently contradictory trends of computing systems. On the one hand, hardware heterogeneity favors the adoption of bytecode format and late, just-in-time code generation. On the other hand, exploitation of hardware features, in particular SIMD extensions through vectorization, is key to obtaining the required performance.
We showed in [33] that speculatively vectorized bytecode is: (1) robust — the approach is general enough to allow execution, both when using SIMD capabilities and also in the absence of SIMD extensions, or when using an unmodified, non-vectorizing JIT compiler; (2) risk-free — the penalty of running vectorized bytecode without SIMD support is kept at a minimum; (3) efficient — the improvement of running vectorized bytecode with SIMD support is maximized.

In [31], we focused on providing an infrastructure capable of supporting diverse SIMD targets (SSE, AltiVec, NEON), across a wide range of vectorizable kernels, with performance comparable to monolithic compiler vectorization.

This research is done within the framework of the HIPEAC2 network in collaboration with Albert Cohen (INRIA Alchemy), Ayal Zaks and Dorit Nuzman (IBM Research Labs, Haifa, Israel).

6.2.3. Vectorization Technology To Improve Interpreter Performance

Participants: Erven Rohou, David Yuste.

Recent trends in consumer electronics have created a new category of portable, lightweight software applications. Typically, these applications have fast development cycles and short life spans. They run on a wide range of systems and are deployed in a target independent bytecode format over Internet and cellular networks. Their authors are untrusted third-party vendors, and they are executed in secure managed runtimes or virtual machines. Furthermore, due to security policies, these virtual machines are often lacking just-in-time compilers and are reliant on interpreter execution.

The main performance penalty in interpreters arises from instruction dispatch. Each bytecode requires a minimum number of machine instructions to be executed. In this work we introduce a powerful and portable representation that reduces instruction dispatch thanks to vectorization technology. It takes advantage of the vast research in vectorization and its presence in modern compilers. Thanks to a split compilation strategy, our approach exhibits almost no overhead. Complex compiler analyses are performed ahead of time. Their results are encoded on top of the bytecode language, becoming new SIMD IR (i.e., intermediate representation) instructions. The bytecode language remains unmodified, thus this representation is compatible with legacy interpreters.

This approach drastically reduces the number of instructions to interpret and improves execution time. SIMD IR instructions are mapped to hardware SIMD instructions when available, with a substantial improvement.

6.2.4. Tiptop

Participant: Erven Rohou.

Hardware performance monitoring counters have recently received a lot of attention. They have been used by diverse communities to understand and improve the quality of computing systems: for example, architects use them to extract application characteristics and propose new hardware mechanisms; compiler writers study how generated code behaves on particular hardware; software developers identify critical regions of their applications and evaluate design choices to select the best performing implementation.

We propose [41] that counters be used by all categories of users, in particular non-experts, and we advocate that a few simple metrics derived from these counters are relevant and useful. For example, a low IPC (number of executed instructions per cycle) indicates that the hardware is not performing at its best; a high cache miss ratio can suggest several causes, such as conflicts between processes in a multicore environment.

We propose tiptop: a new tool, similar to the UNIX top utility, that requires no special privilege and no modification of applications. Tiptop provides more informative estimates of the actual performance than existing UNIX utilities, and better ease of use than current tools based on performance monitoring counters. With several use cases, we have illustrated possible usages of such a tool.

6.3. Understanding performance issues

Participants: Junjie Lai, Ricardo Andrés Velasquéz, Pierre Michaud, André Seznec.
6.3.1. Behavioral application-dependent superscalar core modeling

**Participants:** Ricardo Andrés Velasquéz, Pierre Michaud, André Seznec.

In recent years, research in microarchitecture has shifted from single-core to multi-core processors. Cycle-accurate models for many-core processors featuring hundreds or even thousands of cores are out of reach for realistic workloads. Approximate simulation methodologies which trade accuracy for simulation speed are necessary for conducting certain research, in particular for studying the impact of resource sharing between cores, where the shared resource can be caches, on-chip network, memory bus, power, temperature, etc.

Behavioral superscalar core modeling is a possible way to trade accuracy for simulation speed in situations where the focus of the study is not the core itself but what is outside the core, i.e., the **uncore**. In this modeling approach, a superscalar core is viewed as a black box emitting requests to the uncore at certain times. A behavioral core model can be connected to a cycle-accurate uncore model. Behavioral core models are built from detailed simulations. Once the time to build the model is amortized, significant simulation speedups are achieved.

We have proposed a new method for defining behavioral models for modern superscalar cores. Our method, behavioral application-dependent superscalar core (BADCO) modeling, requires two traces generated with cycle-accurate simulations. For the first trace, all the requests from the core (which includes the level-1 caches) to the uncore are forced with a null latency, i.e., we simulate a perfect uncore. For the second trace, all the requests are forced with a fixed and very long latency. Then we build a BADCO model from the timing information recorded in these two traces. A BADCO model is basically a directed graph where each node represents a group of micro-ops that may carry some requests to the uncore. Edges in this graph represent dependencies between requests. After the model is built, it can be used for simulations. During simulation, the BADCO model emulates the processor’s reorder buffer, the level-1 miss status holding registers, and honors dependencies between nodes. We have compared BADCO with Lee et al.’s PDCM behavioral core model. BADCO is more accurate than PDCM on average and is more reliable [42]. BADCO predicts the execution time of a thread running on a modern superscalar core with an error typically under 5%. From our experiments, we found that BADCO is qualitatively accurate, being able to predict how performance changes when we change the uncore. The simulation speedups obtained with BADCO are typically greater than 10.

6.3.2. Architecture for Lattice QCD

**Participants:** Junjie Lai, André Seznec.

Simulation of Lattice QCD is a challenging computational problem that requires very high performance exceeding sustained Petaflops/s. In the framework of the ANR Cosinus PetaQCD project, we are modeling the demands of this application on the memory system and synchronization mechanisms.

In particular, GPUs have become popular to execute computing intensive scientific applications. In [39], we have introduced a GPU timing model to provide more insights into the applications’ performance on GPU. A GPU CUDA program timing estimation tool (TEG) is developed based on the GPU timing model. Especially, TEG illustrates how performance scales from one warp (CUDA thread group) to multiple concurrent warps on SM (Streaming Multiprocessor). Because TEG takes the native GPU assembly code as input, it allows to estimate the execution time with only a small error. TEG can help programmers to better understand the performance results. It also allows to identify and quantify performance bottlenecks.

6.4. WCET estimation

**Participants:** Damien Hardy, Benjamin Lesage, Isabelle Puaut, Erven Rohou, André Seznec.

Predicting the amount of resources required by embedded software is of prime importance for verifying that the system will fulfill its real-time and resource constraints. A particularly important point in hard real-time embedded systems is to predict the Worst-Case Execution Times (WCETs) of tasks, so that it can be proven that tasks temporal constraints (typically, deadlines) will be met. Our research concerns methods for obtaining automatically upper bounds of the execution times of applications on a given hardware. Our focus this year is on (i) multicore architectures, (ii) applications compiled using just-in-time (JIT) compilation, and (iii) definition of both predictable and efficient hardware.
6.4.1. Timing analysis for multicore platforms with shared caches

Participants: Benjamin Lesage, Damien Hardy, Isabelle Puaut.

WCET estimation for multicore platforms is challenging due to the possible interferences between cores due to shared hardware resources such as shared caches, memory bus, etc. Moreover, multi-core platforms use a hierarchy of caches, whose worst-case behavior has to be predicted safely and as tightly as possible.

6.4.1.1. Scalable fixed-point free instruction cache analysis

Estimating worst-case execution times (WCETs) for architectures with caches requires the worst-case number of cache misses to be upper bounded. Most existing static cache analysis methods, a large majority of those applying to unicores and a single cache level, use fixed-point computation and do not scale well with large code sizes.

Estimating WCETs for multicores requires the base cache analysis used to analyze every cache level in the memory hierarchy to be fast. To address this issue, we have proposed in [28] a new fast and scalable instruction cache analysis technique. In contrast to existing work, neither fixed-point computation nor heavyweight interprocedural analysis are required. Thus, code sizes that are too long to analyze with existing techniques become analyzable with lower analysis time and memory consumption, and with only a slight degradation of the analysis precision. Experimental results show a reduction of the analysis execution time of a factor 5 in average (with a peak near 30 for the largest and most complex code) with a limited waste of tightness of the analysis.

6.4.1.2. Static analysis of cache hierarchies

Participants: Benjamin Lesage, Damien Hardy, Isabelle Puaut.

Determining the worst-case number of cache misses in multicore architectures requires the definition of analysis techniques applicable to cache hierarchies. In our previous work [6] we have defined the first technique that safely analyzes such hierarchies, first considering non-inclusive caches and LRU cache replacement.

We have recently generalized our previous work to support different cache hierarchy management policies between cache levels: non-inclusive, inclusive and exclusive cache hierarchies. Moreover, our analysis now supports cache hierarchies with different replacement policies: Least Recently Used (LRU), Pseudo-LRU, First-In First-Out (FIFO), Most Recently Used (MRU) and Random. Experimental results, detailed in [21] show that the method is precise in many cases (non-inclusive and exclusive cache hierarchies with the LRU replacement policy) and has a reasonable computation time. Nevertheless, considering inclusion enforcement mechanisms and non-LRU replacement policies leads to an increase of the analysis pessimism. Moreover, these two sources of pessimism are cumulative, thus resulting, in some cases, in a significant overestimation. Although inclusive cache hierarchies with non-LRU replacement policies can be analyzed statically, the cache hierarchies to be favored to obtain the tighter WCET estimates are hierarchies of non-inclusive or exclusive caches with the LRU replacement policy.

6.4.1.3. Reconciling Predictability and Just-In-Time Compilation

Participants: Isabelle Puaut, Erven Rohou.

Virtualization and just-in-time (JIT) compilation have become important tools to address application portability issues without deteriorating average-case performance. Unfortunately, JIT compilation raises predictability issues, which currently hinders its dissemination in real-time applications. Our work aims at reconciling the two domains, i.e. taking advantage of the portability and performance provided by JIT compilation, while providing predictability guarantees.
As a first step towards this ambitious goal, we have proposed two structures of code caches and have demonstrated their predictability [26]. On the one hand, the binary code caches we propose avoid too frequent function recombinations, providing good average-case performance. On the other hand, and more importantly for the system determinism, we show that the behavior of the code cache is predictable: a safe upper bound of the number of function recombinations can be computed, enabling the verification of timing constraints. Experimental results show that fixing function addresses in the binary cache ahead of time results in tighter Worst Case Execution Times (WCETs) than organizing the binary code cache in fixed-size blocks replaced using a Least Recently Used (LRU) policy.

6.4.1.4. Predictable shared caches for mixed-criticality real-time systems

**Participants:** Benjamin Lesage, Isabelle Puaut, André Seznec.

The general adoption of multi-core architectures has raised new opportunities as well as new issues in all application domains. In the context of real-time applications, it has created one major opportunity and one major difficulty. On the one hand, the availability of multiple high performance cores has created the opportunity to mix on the same hardware platform the execution of a complex critical real-time workload and the execution of non-critical applications. On the other hand, for real-time tasks timing deadlines must be met and enforced. Hardware resource sharing inherent to multicores hinders the timing analysis of concurrent tasks. Two different objectives are then pursued: enforcing timing deadlines for real-time tasks and achieving highest possible performance for the non-critical workload.

In this work, we suggest a hybrid hardware-based cache partitioning scheme that aims at achieving these two objectives at the same time. Plainly considering inter-task conflicts on shared cache for real-time tasks yields very pessimistic timing estimates. We remove this pessimism by reserving private cache space for real-time tasks. Upon the creation of a real-time task, our scheme reserves a fixed number of cache lines per set for the task. Therefore uniprocessor worst case execution time (WCET) estimation techniques can be used, resulting in tight WCET estimates. Upon the termination of the real-time task, this private cache space is released and made available for all the executed threads including non-critical ones. That is, apart the private spaces reserved for the real-time tasks currently running, the cache space is shared by all tasks running on the processor, i.e. non-critical tasks but also the real-time tasks for their least recently used blocks. Experiments show that the proposed cache scheme allows to both guarantee the schedulability of a set of real-time tasks with tight timing constraints and enable high performance on the non-critical tasks.

6.4.2. Preemption delay analysis for floating non-preemptive region scheduling

**Participant:** Isabelle Puaut.

This is joint work with Stefan M. Petters, Vincent Nélis and José Marinho, ISEP Porto, Portugal.

In real-time systems, there are two distinct trends for scheduling task sets on unicore systems: non-preemptive and preemptive scheduling. Non-preemptive scheduling is obviously not subject to any preemption delays but its schedulability may be quite poor, whereas fully preemptive scheduling is subject to preemption delays, but benefits from a higher flexibility in the scheduling decisions.

The time-delay involved by task preemptions is a major source of pessimism in the analysis of the task Worst-Case Execution Time (WCET) in real-time systems. Cache related preemption delays (CRPD) are the most important ones, and are caused by the preemting tasks that modify the cache; the preempted task then suffers an indirect delay after the preemption to reload the cache with useful information.

Preemptive scheduling policies including non-preemptive regions are a hybrid solution between non-preemptive and fully preemptive scheduling paradigms, which enables to conjugate both worlds benefits. In this work [29], we exploit the connection between the progression of a task in its operations, and the knowledge of the preemption delays as a function of its progression. Thus the pessimism in the preemption delay estimation is reduced, in comparison to state of the art methods, due to the increase in information available in the analysis.
6. New Results

6.1. Dynamically and Heterogeneous Reconfigurable Platforms

6.1.1. New Reconfigurable Architectures

6.1.1.1. Power models of reconfigurable architectures

Participants: Robin Bonamy, Daniel Chillet, Olivier Sentieys.

Including a reconfigurable area in complex systems-on-chip is now considered as an interesting solution to reduce the area of the global system and to support high performances. But the key challenge in the context of embedded systems is currently the power budget of the system, and the designer needs some early estimations of the power consumption of its system. Power estimation for reconfigurable systems is a difficult problem because several parameters need to be taken into account to define an accurate model.

Hardware implementation of an algorithm provides different choices to the designer compared to software implementation. It is possible to vary the parallelism level or loop unrolling index, which has a direct impact on area and execution time and therefore on power and energy consumption. First we evaluated delay, area, power and energy impacts of loop transformations using High Level Synthesis tools. We have made several power measurements on a real FPGA platform and for different task implementations in order to build a model of energy consumption versus execution time. Work is in progress to also characterize energy consumption of tasks through extracting the number of elementary operators used in the hardware implemented task.

Furthermore, we also consider the opportunity of the dynamic reconfiguration, which makes possible to partially reconfigure a specific part of the circuit while the rest of the system is running. This opportunity has two main effects on power consumption. First, thanks to the area sharing ability, the global size of the device can be reduced and the static (leakage) power consumption can thus be reduced. Secondly, it is possible to delete the configuration of a part of the device which reduces the dynamic power consumption when a task is no longer used. Although the cost of the reconfiguration is still important, in some cases this technique can be interesting to reduce the power of the system. To evaluate the potential gain of the dynamic reconfiguration, we have made some measurements on a Virtex 5 board. We have defined a first model of the power consumption of the reconfiguration. This model shows that the power consumption not only depends on the bitstream file size but also on the content of the reconfiguration region [41], [42].

These experiments allow us to define energy and delay models that will be used by the operating system including a power management strategy to decide on-line which task instances must be executed to efficiently manage the available power using dynamic partial reconfiguration [82].

6.1.1.2. High-level modeling of reconfigurable architectures

Participants: Robin Bonamy, Daniel Chillet, Sébastien Pillement.

To help System-on-Chip designers to explore the large design space, high-level methodologies and tools are more and more often required. The exploration phase is particularly difficult when the system must satisfy a large number of constraints, like performance, real time and power consumption. If the classical multiprocessor system-on-chips can be modeled without any difficulty, dynamically reconfigurable embedded accelerators are not correctly covered by the usual modeling languages.

In this context, we have extended the AADL (Architecture Analysis and Design Language) language to include the reconfiguration aspect included in nowadays’ MPSoC [19], [40]. This work is part of a more general project, Open-People, which proposes complete methodology for power and energy consumption analysis. The proposal is based on AADL property extensions which are applied on component models. A three-level model has been defined for every targeted FPGA. The first level defines a generic FPGA which allows to model every possible FPGA. The second level allows the specialization of the FPGA for a specific family. Finally, the third level provides the support to describe the deployment of an application on a specific FPGA circuit.
To complete these levels of description, we started the development of techniques for constraint verifications. These developments are based on the OCL language, which allows to extract characteristics on the AADL model, compute mathematical expressions and finally verify mathematical constraints. These verifications have been developed for power and energy consumption, they include static and dynamic power estimation and soon the power consumption during the dynamic reconfiguration process.

6.1.1.3. Reconfiguration controller

Participants: Manh Pham, Daniel Chillet, Sébastien Pillement.

Dynamically reconfigurable architectures, which can offer high performance, are increasingly used in different domains. Unfortunately, lots of applications cannot benefit from this new paradigm due to large timing overhead. Even for partial reconfiguration, modifying a small region of an FPGA takes few ms. To cope with this problem we have developed an ultra-fast power-aware reconfiguration controller (UPaRC) to boost the reconfiguration throughput up to 1.433 GB/s. UPaRC can not only enhance the system performance, but also auto-adapt to various performance and consumption conditions. This could enlarge the range of supported applications and can optimize power-timing trade-off of reconfiguration phase for each selected application during run-time. The energy-efficiency of UPaRC over state-of-the-art reconfiguration controllers is up to 45 times more efficient. This work has been accepted for publication in DATE’2012 [56].

6.1.2. Management of Dynamically Reconfigurable Systems

6.1.2.1. Spatio-Temporal Scheduling based on Artificial Neural Networks

Participants: Antoine Eiche, Daniel Chillet, Sébastien Pillement, Olivier Sentieys.

Management of task execution on dynamically reconfigurable accelerators is known to be a difficult problem due to the large number of possibilities of task instantiations. The problem to solve can be defined as the “spatio-temporal task scheduling”. The problem becomes even more difficult to solve when the solutions must be produced during the execution of the application, i.e. on-line. In this context, new algorithms must be defined and, to solve this problem, we propose to define a neural network based on the Hopfield model.

We are therefore able to address heterogeneous multiprocessor systems and to manage the reconfigurable resources embedded within MPSoC [23]. Our latest works on this topic focused on two different issues. First we demonstrated that neural network structures used for task scheduling can continue to produce valid solutions even if one or several neurons are in fault [70]. This characteristic is very important for present and future technologies for which the fabrication process variability can lead to increase the number of defaults in the circuit. The second focus concerned the optimization of the neural network convergence by using parallel evaluation of neurons. We have shown how to define several neuron packets (from the neural network) that can be evaluated in parallel without modifying the convergence property [44], [76].

6.1.2.2. Flexible Communication OS Service

Participants: Daniel Chillet, Sébastien Pillement, Ludovic Devaux.

In a multiprocessor system, to gain the advantages of parallelism, efficient communication and memory management are highly required. Recent developments in the partial and dynamic reconfigurable computing domain demand better ways to manage simultaneous task execution. But, the requirements are slightly different from the traditional software based systems. In this context, Operating System (OS) services like scheduling, placement, inter-task communication have been developed to make such platforms flexible and self-sufficient. For task communications within flexible architectures, we defined a specific network-on-chip adapted to dynamically and partially reconfigurable resources included into modern SoC. The characterization of the DRAFT network was completed and its integration inside reconfigurable systems on chip was realized [14]. We then focused on the run-time communication service [50] and dynamic memory management [49] in reconfigurable System-on-Chips (RSoCs). We first developed a hardware communication block and the communication schemes supported by this new OS service. The originality relies on the implementation of this services directly inside the FPGA. We then demonstrated the requirements and advantages of having a local memory task or a dynamically configurable memory task, in order to improve effectiveness and efficiency of the proposed schemes.
6.1.3. Fault-Tolerant Reconfigurable Architectures

**Participants:** Sébastien Pillement, Manh Pham, Stanislaw Piestrak.

In terms of complex systems implementation, reconfigurable FPGA circuits are now part of the mainstream thanks to their flexibility, performances and high number of integrated resources. FPGAs enter new fields of applications such as aeronautics, military, automotive or confined control thanks to their ability to be remotely updated. However, these fields of applications correspond to harsh environments (cosmic radiation, ionizing, electromagnetic noise) and with high fault-tolerance requirements. We then propose a complete framework to design reconfigurable architecture supporting fault-tolerance mitigation scheme [58]. The proposed framework allows simulation, validation of mitigation operations, but also to size architecture resources. The physical implementation of the fault-tolerant reconfigurable platform permits to validate the proposed model and the effectiveness of the framework. This implementation shows the potential of dynamically reconfigurable architectures for supporting fault-tolerance in embedded systems. We also worked on new approach in order to include dependability in the DRAFT coarse-grained reconfigurable architecture [37].

6.1.4. Low-Power Architectures

6.1.4.1. Ultra Low-Power Architecture for Control-Oriented Applications in Wireless Sensor Nodes

**Participants:** Olivier Sentieys, Steven Derrien, Vivek D. Tovinakere, Philippe Quémerais, Romain Fontaine.

This research work aims at developing ultra low-power SoC for wireless sensor nodes, as an alternative to existing approaches based low-power micro-controllers. The proposed approach reduces the power consumption by using a combination of hardware specialization and power gating techniques. In particular, we use the fact that typical WSN applications are generally modeled as a set of small to medium grain tasks that are implemented on low power microcontroller using light weight thread-like OS constructs.

Rather than implementing these tasks in software, we instead propose to map each of these tasks to their own specialized hardware structures that we call a hardware micro-task. Such hardware task consists of a minimalistic (and customized) data-path controlled by a finite state machine (FSM). By customizing each of these hardware implementations to their corresponding task, we expect to significantly reduce the dynamic power dissipated by the whole system. Besides, to circumvent the increase in static power caused by the possibly numerous hardware tasks implemented in the chip, we also propose to combine our approach with power gating, so as to supply power to a hardware task only when it needs to be executed [28].

As a proof of concept, a chip has been designed and fabricated in a 65nm CMOS from STMicroelectronics using the CMP facilities. The area is about 1mm² in a QFN52 package. The circuit is a controller part of a wireless sensor network node. It embeds an OpenMSP microcontroller core with SRAM memories for data and programs and some dedicated hardware tasks to control an external radio transceiver such as the TI CC2420 commonly used in the industry.

To reduce energy consumption, low power design techniques such as power gating were used. Two power domains are implemented: one is dedicated to microcontroller and memories, while the goal of the second is to measure the efficiency of our hardware micro-task concept.

The input-output ring around the core is divided into three parts: two parts are digital I/O pads corresponding to a power domain and the third contains analog pads to control the power gating for monitoring. Our goal is to analyze the power benefits of our approach and to compare it with classical microprocessor architectures.

6.1.4.2. Wakeup Time and Wakeup Energy Estimation in Power-Gated Logic Clusters

**Participants:** Olivier Sentieys, Vivek D. Tovinakere.

Run-time power gating for aggressive leakage reduction has brought into focus the cost of mode transition overheads due to frequent switching between sleep and active modes of circuit operation. In order to design circuits for effective power gating, logic circuits must be characterized for overheads they present during mode transitions. We have proposed a method to determine steady-state virtual-supply voltage in active mode and hence present a model for virtual-supply voltage in terms of basic circuit parameters. Further, we derived
expressions for estimation of two mode transition overheads: wakeup time and wakeup energy for a power-gated logic cluster using the proposed model. Experimental results of application of the model to ISCAS85 benchmark circuits show that wakeup time may be estimated within an average error of 16.3% across 22× variation in sleep transistor sizes and 13× variation in circuit sizes with significant speedup in computation time compared to SPICE level circuit simulations [30], [63].

6.1.5. Arithmetic Operators for Cryptography

Participants: Arnaud Tisserand, Thomas Chabrier, Danuta Pamula.

6.1.5.1. ECC Processor with Protections Against SCA

A dedicated processor for elliptic curve cryptography (ECC) is under development. Functional units for arithmetic operations in \( \mathbb{F}_{2^m} \) and \( \mathbb{F}_p \) finite fields and 160–600-bit operands have been developed for FPGA implementation. Several protection methods against side channel attacks (SCA) have been studied. The use of some number systems, especially very redundant ones, allows to change the way some computations are performed and then their effects on side channel traces.

We propose in [83] hardware implementation of the double base number system (DBNS) random recoding of secret keys. This recoding is performed on-the-fly during the elliptic curve cryptography (ECC) scalar multiplication \([k]P\). This leads random behavior of the point operations at the side channel level.

We started a collaboration with the University of Sfax in Tunisia on the use of ECC processor for secure communications in low-cost wireless applications. A first FPGA implementation is under development and we expect to submit our first results in 2012.

6.1.5.2. Arithmetic Operators for High-Performance Cryptography

In [32], we published an extended version of the work started in 2010 on fast algorithms and implementations of \( \mathbb{F}_{2^m} \) finite field multiplication units in FPGA. The proposed and compared methods are based on separated multiplication and reduction steps and analyzed various area and time dependency/efficiency/complexity tradeoffs.

With Mark Hamilton, PhD student in the Code and Crypto group from the University College Cork (UCC), we worked on fast algorithms and implementations of \( \mathbb{F}_p \) finite field multipliers for some specific values of \( p \). The corresponding results have been published in [46].

6.1.6. SoC Modeling and Prototyping on FPGA-based Systems

Participants: François Charot, Kevin Martin, Laurent Perraudau, Charles Wagner.
SoCLib and MutekH are two software development projects to which we contribute. SoCLib (http://www.soclib.fr) is an open platform for modeling and simulation of multiprocessors system-on-chip (MP-SoC). MutekH (http://www.mutekh.org) is a free and portable operating system for embedded platforms, ranging from micro-controller to multiprocessor systems. The use of the configurable and extensible simulation model of the Altera NIOSII processor of the SoCLib library and of the MutekH operating system allows us to easily deploy software applications such as codes from MediaBench, MiBench and Cryptographic Library benchmark sets or multithreaded applications on monoprocessor and multiprocessor simulation platforms. These platforms are used on the one hand for the validation of the processor extensions automatically generated by our compilation tools and on the other hand for the measurement of the speedup achieved using these new extensions.

6.2. Compilation and Synthesis for Reconfigurable Platform

Participants: Steven Derrien, Emmanuel Casseau, Daniel Menard, François Charot, Christophe Wolinski, Olivier Sentieys, Patrice Quinton.

6.2.1. Polyhedral based loop transformations for High-Level synthesis

Participants: Steven Derrien, Antoine Morvan, Patrice Quinton.

After almost two decades of research effort, there now exists a large choice of robust and mature C to hardware tools that are used as production tools by world-class chip vendor companies. Although these tools dramatically slash design time, their ability to generate efficient accelerators is still limited, and they rely on the designer to expose parallelism and to use appropriate data layout in the source program. We believe this can be overcome by tackling the problem directly at the source level, using source-to-source optimizing compilers. More precisely, our aim is to study how polyhedral based program analysis and transformation can be used to address this problem.

In the context of the PhD of Antoine Morvan, we have studied how it was possible to improve the efficiency and applicability of nested loop pipelining (also known as nested software pipelining) in C to hardware tools. Loop pipelining is a key transformation in high-level synthesis tools as it helps maximizing both computational throughput and hardware utilization. Nevertheless, it somewhat looses its efficiency when dealing with small trip-count inner loops, as the pipeline latency overhead quickly limits its efficiency.

Even if it is possible to overcome this limitation by pipelining the execution of a whole loop nest, the applicability of nested loop pipelining has so far been limited to a very narrow subset of loops, namely perfectly nested loops with constant bounds. In this work, we have extended the applicability of nested-loop pipelining to imperfectly nested loops with affine dependencies. We have shown how such loop nest can be analyzed and, under certain conditions, how one can modify the source code in order to allow nested loop pipeline to be applied using a method called polyhedral bubble insertion. Our approach shown encouraging results and led to a publication to the IEEE International Conference on Field Programmable Technology [48] in December 2011.

6.2.2. Reconfigurable Processor Extensions Generation

Participants: Christophe Wolinski, François Charot, Erwan Raffin, Kevin Martin, Antoine Floch.

During this year, we have continued our work on the generation of reconfigurable processor extension using the constraint programming approach. Previously, we showed how all the problems ranging from instruction identification, scheduling and binding to optimized architecture synthesis can be defined and solved using the constraint programming approach. This year, a new pattern scheduling approach has been defined. It enables concurrent match selection and parallel match scheduling on the processor and extension assuming that the execution on an extension is not atomic. It means that the data produced by an extension must not necessarily be sent to the processor just after the end of processing. Thanks to that, a better scheduling can be obtained [71]. The efficient FPGA implementation of processing units require optimization of hardware resources, such as registers and multiplexers. The extension synthesis defined previously has been revisited. For applications from MediaBench, MiBench and MiCrypt benchmark sets, an improvement of 35%, after placement and routing on the Stratix2 Altera FPGA, is observed.
6.2.3. Run-time Reconfigurable Architecture Modeling

Participants: Christophe Wolinski, François Charot, Emmanuel Casseau, Daniel Menard, Antoine Floch, Erwan Raffin, Steven Derrien.

We have continued to work on the modeling problem of the run-time reconfigurable, operator-based, ROMA multimedia architecture. The ROMA processor is composed of a set of \( M \) coarse-grained reconfigurable operators, \( N \) data memories, a configuration memory, two interconnection networks (between operators and between operators and memories), and dedicated controllers designed for each module of the datapath. A centralized controller manages the configuration and the execution steps. The ROMA processor has three different interfaces: a data interface connected to the operator network, a control interface and a debug interface connected to the main controller. The number of operators, the number of memories and their size can be decided according to application requirements. The compilation flow of our framework rests on the use of an architecture abstract model of the targeted ROMA architecture.

During this year we have focused on the definition of the constraint model to deploy an application graph on the pipeline architecture model. The goal is here to minimize the latency of the pipeline. The main changes are at the operator and memory levels. The operators are pipelined and the dual port memories behave like circular buffers. Recall that in the case of the non pipelined model, the goal is to optimize the execution time of the application under resource constraints. We have carried out experiments to evaluate the quality of our method using different pattern libraries (patterns supported by the ROMA SWP coarse-grained reconfigurable operator, patterns extracted from the MediaBench set) \[47\]. In these experiments the model has no limitation in terms of number of operators and number of memories. The optimality of the solutions were proven in 93% of cases. More details can be found in \[29\] and in the Ph.D. thesis of Erwan Raffin \[17\].

In the context of the RecMotifs project, we have continued to work on a specific design flow integrating STMicroelectronics’ compiler flow. This project also allowed us to bring significant evolution to our pattern analysis software tools. The RecMotifs flow consists in a pattern analysis flow for STMicroelectronics graphml files generated by ST compiler. This flow allows pattern description (description of graphml pattern that can be used in the covering pass), type extraction, pattern generation (pattern generation on a graphml file), covering (covering of a graphml file with minimization of the parallel execution time without any resource constraints). Once the pattern analysis has been applied to the graphml files, C code regeneration can be performed using GeCos.

6.2.4. Floating-Point to Fixed-Point Conversion

Participants: Daniel Menard, Karthick Parashar, Olivier Sentieys, Romuald Rocher, Hai-Nam Nguyen.

For the fixed-point conversion process, different optimization algorithms have been tested. The aim is to minimize the implementation cost under accuracy constraint. In \[54\], two new algorithms for the word-length optimization procedure, based on the Greedy Randomized Adaptive Search Procedure (GRASP), are proposed. Compared to existing methods, our proposition yields better results and has a complexity between deterministic methods and stochastic methods.

6.3. Algorithm Architecture Interaction

6.3.1. Flexible hardware accelerators for biocomputing applications

Participants: Steven Derrien, Naeem Abbas, Patrice Quinton.

It is widely acknowledged that FPGA-based hardware acceleration of compute intensive bioinformatics applications can be a viable alternative to cluster (or grid) based approach as they offer very interesting MIPS/watt figure of merits. One of the issues with this technology is that it remains somewhat difficult to use and to maintain (one is rather designing a circuit rather than programming a machine).
Even though there exists C-to-hardware compilation tools (Catapult-C, Impulse-C, etc.), a common belief is that they do not generally offer good enough performance to justify the use of such reconfigurable technology. As a matter of fact, successful hardware implementations of bio-computing algorithms are manually designed at RTL level and are usually targeted to a specific system, with little if any performance portability among reconfigurable platforms.

This research work, which is part of the ANR BioWic project, aims at providing a framework for helping semi-automatic generation of high-performance hardware accelerators. In particular we expect to widen the scope of common design constraints by focusing on system-level criterions that involve both the host machine and the accelerator (workload balancing, communications and data reuse optimisations, hardware utilization rate, etc.). This research work builds upon the CAIRN research group expertise on automatic parallelization for application specific hardware accelerators and has been targeting mainstream bioinformatics applications (HMMER, ClustalW and BLAST).

Our work in 2011 extended the experiment results obtained in 2010 and led to the submission of a paper to IEEE Trans. in Parallel and Distributed Computing (the article being in revision). We also investigated another case study based on a more classical sequence comparison algorithm for which we investigated different style of architectural partitioning. This work led to a paper published in the proceedings of the ARC International Symposium [43].

6.3.2. Range Estimation and Computation Accuracy Optimization

Participants: Daniel Menard, Karthick Parashar, Olivier Sentieys, Romuald Rocher, Pascal Scalart, Aymen Chakhari, Jean-Charles Naud, Emmanuel Casseau, Andrei Banciu.

6.3.2.1. Range Estimation

Efficient range estimation methods are required to optimize the integer part word-length. Our previous works based on the Karhunen-Loève Expansion (KLE) have been extended in [38]. The impulse response between the input and a variable is used to propagate the KLE parameters of the inputs. Range estimation has proven to be a difficult problem for non-linear operations especially when the input data is correlated. A stochastic approach can significantly improve the results compared to the classical methods like the interval and affine arithmetic. The aim is to obtain tight intervals by adapting the bounds to a desired probability of overflows. An approach for the analysis of range uncertainties based on the Polynomial Chaos Expansion (PCE) has been developed. The PCE representation is obtained for every input variable and an analytical description of the variability of the output is determined. Furthermore, the correlation of the inputs is captured using the Nataf transform. The range is computed using a probabilistic analysis from the probability density function (PDF).

6.3.2.2. Accuracy and performance evaluation

The automation of fixed-point conversion requires generic methods to study accuracy degradation. In [51], [73] a new approach using analytical noise power propagation considering conditional structures. These structures are generated from programming language statements such as if-then-else or Switch. The proposed model takes into account two key points in fixed-point design: first, an alternative processing of noise depending on the condition; second, decision errors generated by quantization noise affecting the condition. This method is integrated in the fixed-point conversion process and uses path probabilities of execution alternatives obtained from profiling. This work extends existing analytical approaches for fixed-point conversion. Experimentations of our analytical method show that it has a fairly accurate noise power estimation compared to the real accuracy degradation. An analytical approach is studied to determine accuracy of systems including unsmooth operators. An unsmooth operator represents a function which is not derivable in all its definition interval (for example the sign operator). The classical model is no valid yet since these operators introduce errors that do not respect the Widrow assumption (their values are often higher than signal power). So an approach based on the distribution of the signal and the noise is proposed. It is applied to the sphere decoding algorithm. We also focus on recursive structure where an error influences future decision. So, the Decision Feedback Equalizer is also considered.

6.3.3. Reconfigurable Video Coding

Participants: Emmanuel Casseau, Olivier Sentieys, Arnaud Carer, Cécile Beaumin, Hervé Yviquel.
In the field of multimedia coding, standardization recommendations are always evolving. To reduce design
time, Reconfigurable Video Coding (RVC) standard allows defining new codec algorithms based on a
modular library of components. RVC dataflow-based specification formalism expressly targets multiprocessors
platforms. However software processor cannot cope with high performance and low power requirements.
Hence the mapping of RVC specifications on hardware accelerators is investigated in this work, as well
as the scheduling of the functional units (FU) of the specification. Dataflow programming, such as RVC
applications, express explicit parallelism within an application. Although multi-core processors are now
available everywhere, few applications are able to truly exploit their multiprocessing capabilities. We describe
in [69] a scheduling strategy for executing a dataflow program on multi-core architectures using distributed
schedulers and lock-free communications. Actually, our goal is to design an RVC-dedicated reconfigurable
architecture with various resources. Our previous results lead to the definition of a reconfigurable FIFO
for optimizing cost and performance of RVC dataflow specifications by taking advantage of their dynamic
behavior. We are currently working with Mickael Raulet from IETR INSA Rennes and Dr. Jani Boutellier
from the university of Oulu (Finland), concerning the execution of an RVC decoder on a network of Transport
Triggered Architecture (TTA) processors (proposed by the Tampere University of Technology). Thanks to its
modular structure, TTA can be seen as a nice kind of CPU design to develop Application-Specific Processor.
TTA processor network is connected by hardware channels so it has many similarities with RVC network.
Hervé Yviquel, is expected to have a 4-month stay in 2012 in TUT to provide a functional automated flow to
design TTA-based platform and compile RVC application for this platform.

6.3.4. Multi-Antenna Systems
Participants: Olivier Berder, Pascal Scalart, Quoc-Tuong Ngo.

Considering the possibility for the transmitter to get some Channel State Information (CSI) from the receiver,
antenna power allocation strategies can be performed thanks to the joined optimization of linear precoder (at
the transmitter) and decoder (at the receiver) according to various criteria.

A new exact solution of the maximization of the minimum Euclidean distance between received symbols has
been proposed for two 16-QAM modulated symbols. This precoder shows an important enhancement of this
minimum distance compared to diagonal precoders, which leads to a significant BER performance improve-
ment. This new strategy selects the best precoding matrix among eight different expressions, depending on the
value of the channel angle. Selecting only two of these expressions, this precoder was then generalized to any
rectangular QAM modulation [26].

Not only the minimum Euclidean distance but also the number of neighbors providing it has an important role
in reducing the error probability when a Maximum Likelihood detection is considered at the receiver. Aiming
at reducing this number of neighbors, a new precoder in which the rotation parameter has no influence is
proposed for two independent data streams transmitted. The expression of the new precoding strategy is less
complex and the space of solution is, therefore, smaller [53], [74]. In the paper [52], we proposed the
general neighbor-dmin precoder for three independent data-streams and the simulation results also confirm
a significant bit-error-rate improvement of the new precoder in comparison with other traditional precoding
strategies.

6.3.5. Cooperative Strategies for Low-Energy Wireless Networks
Participants: Olivier Berder, Le Quang Vinh Tran, Olivier Sentieys.

During the last decade, many works were devoted to improving the performance of relaying techniques in ad
hoc networks. One promising approach consists in allowing the relay nodes to cooperate, thus using spatial
diversity to increase the capacity of the system. In wireless distributed networks where multiple antennas can
not be installed in one wireless node, cooperative relay and cooperative Multi-Input Multi-Output (MIMO)
techniques can indeed be used to exploit spatial and temporal diversity gain in order to reduce energy
consumption.
Considering a system having a two-antenna source, two one-antenna relays and a one-antenna destination, MIMO simple cooperative relay model (MSCR) and MIMO full cooperative relay model (MFCR) are proposed in comparison with MIMO normal cooperative relay model (MNCR) where the relays forward signals consecutively to destination. The energy efficiency of these models is investigated by using a realistic power consumption model where the parameters are extracted from the characteristics of CC2420, a wireless sensor transceiver widely used and commercially available. For each transmission ranges, the optimal cooperative scheme in terms of energy efficiency is provided by simulation results [65], [78].

A fair analytical investigation on these cooperative protocols was also performed. A lower bound for the average symbol error probability (ASEP) of full DSTC cooperative relaying system in a Rayleigh fading environment is provided. In the case when the Signal to Noise Ratio (SNR) of the relay-relay link is much greater than that of the source-relay link, the upper bound on ASEP of this system is also derived. The effect of the distance between the relays shows that the performance does not degrade so much as the distance between relays is lower than a half of the source-destination distance. Moreover, we also show that, when the error synchronization range is lower than 0.5, the impact of the transmission synchronization error of the relay-destination link on the performance is not considerable [64].

The energy efficiency of cooperative MIMO and relay techniques is also very useful for the Infrastructure to Vehicle (I2V) and Infrastructure to Infrastructure (I2I) communications in Intelligent Transport Systems (ITS) networks where the energy consumption of wireless nodes embedded on road infrastructure is constrained. Applications of cooperation between nodes to ITS networks are proposed and the performance and the energy consumption of cooperative relay and cooperative MIMO are investigated in comparison with the traditional multi-hop technique. The comparison between these cooperative techniques helps us to choose the optimal cooperative strategy in terms of energy consumption for energy constrained road infrastructure networks in ITS applications [27].

6.3.6. Opportunistic Routing

**Participants:** Olivier Berder, Olivier Sentieys, Ruifeng Zhang, Jean-Marie Gorce [Insa Lyon, INRIA Swing].

However, the aforementioned approaches introduce an overhead in terms of information exchange, increasing the complexity of the receivers. A simpler way of exploiting spatial diversity is referred to as opportunistic routing. In this scheme, a cluster of nodes still serves as relay candidates but only a single node in the cluster forwards the packet. This paper proposes a thorough analysis of opportunistic routing 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 [31]. We then provided a closed-form expression of the lower bound of the energy-delay tradeoff and of energy efficiency for different channel models (additive white Gaussian noise, Rayleigh fast fading and Rayleigh block-fading) in a linear network. These analytical results are also verified in 2-dimensional Poisson networks using simulations. The closed-form expression provides a framework to evaluate the energy-delay performance and to optimize the parameters in physical layer, MAC layer and routing layer from the viewpoint of cross-layer design during the planning phase of a network.

6.3.7. Adaptive techniques for WSN power optimization

**Participants:** Olivier Berder, Daniel Menard, Olivier Sentieys, Mahtab Alam, Trong-Nhan Le.

Wireless sensor networks (WSNs) have obtained a great relevancy in civil as well as military applications such as environment sensing, real-time surveillance and habitat monitoring. It is difficult to design a node that is efficient for all of these different applications. The ideal sensor node would have to dynamically adapt its behavior to various parameters such as the data traffic, the channel conditions, the amount of harvested energy, its battery level, etc. Including the capability to scavenge energy from its environment, the design of an efficient power manager able to address both hardware and software processing seems very promising.
Energy modeling is an important issue for designing and dimensioning low power wireless sensor networks (WSN). In order to help the developers to optimize the energy spent by WSN nodes, a pragmatic and precise hybrid energy model is proposed. This model considers different scenarios that occur during the communication and evaluates their energy consumption based on software profiling as well as the hardware components power profiles. The proposed model is a combination of analytical derivations and real time measurements. These experiments are particularly useful to understand the medium access control (MAC) layer mechanisms, such as wake up or data collisions for the preamble sampling category, and the energy wasted by collisions can be evaluated [18], [35].

An adaptive wake-up-interval scheme for preamble sampling MAC protocols for variable traffic in WSN is then proposed. The wake-up-interval is updated based on the traffic status register (whose content depends on the presence of messages for a particular node). The results show that the sensor node adapts and converges its wake-up-interval to the best trade-off value for fixed and variable traffic patterns. Two optimization parameters (length of traffic status register and initial wake-up-interval value) are also tuned to achieve fast convergence speed for different traffic rates and variations.

A wireless body area sensor network (WBASN) demands ultra-low power and energy-efficient protocols. MAC layer plays a pivotal role for energy management in WBASN, moreover, idle listening is the dominant energy waste in most of the MAC protocols. WBASN exhibits wide range of traffic variations based on different physiological data emanating from the monitored patient. In this context, we proposed a novel energy efficient traffic-aware dynamic (TAD) MAC protocol for WBASN [36]. A comparison with other protocols for three different widely used radio chips, i.e. cc2420, cc1000 and amis52100, is presented. The results show that TAD-MAC outperforms all the other protocols under fixed and variable traffic rates.
5. New Results

5.1. Control-Flow Analysis by Abstract Interpretation

Control-flow analysis (CFA) of functional programs is concerned with determining how the program’s functions call each other. In the case of the lambda calculus, this amounts to computing the flow of lambda expressions in order to determine what functions are effectively called in an application \((e_1 \ e_2)\). This work shows that it is possible to use abstract interpretation techniques to derive systematically a control-flow analysis for a simple higher-order functional language. The analysis approximates the interprocedural control-flow of both function calls and returns in the presence of first-class functions and tail-call optimization. A number of advantages follow from taking this approach:

- The systematic derivation of a CFA for a higher-order functional language from a well-known operational semantics provides the resulting analysis with strong mathematical foundations. Its correctness follows directly from the general theorems of abstract interpretation.
- The approach is easily adapted to different variants of the source language. We demonstrate this by deriving a CFA for functional programs written in continuation-passing style.
- The common framework of these analyses enables their comparison. We take advantage of this to settle a question about the equivalence between the analysis of programs in direct and continuation-passing style.
- The resulting equations can be given an equivalent constraint-based presentation, providing ipso facto a rational reconstruction and a correctness proof of constraint-based CFA.

This work was presented at the Japanese Shonan workshop on Verification of higher-order functional programs in September 2011. A journal article is accepted to appear in Information and Computation.

5.2. Modular SMT Proofs for Fast Reflexive Checking inside Coq

Participants: Frédéric Besson, Pierre-Emmanuel Cornilleau, David Pichardie.

Satisfiability Modulo Theory (SMT) solvers are efficient automatic provers for combination of theories. Those solvers have proved very successful in program verification because they discharge automatically and efficiently challenging verification conditions. SMT solvers are therefore de facto part of the Trusted Computing Base of many program verification methodologies. A consequence is that a soundness bug in a SMT solver can make the whole program verification process unsound.

To tackle this problem, we propose a new methodology for exchanging unsatisfiability proofs between an untrusted SMT solver and a sceptical proof assistant with computation capabilities like Coq. We advocate modular SMT proofs that separate boolean reasoning and theory reasoning; and structure the communication between theories using Nelson-Oppen combination scheme.

We present the design and implementation of a Coq reflexive verifier that is modular and allows for fine-tuned theory-specific verifiers. The current verifier is able to verify proofs for quantifier-free formulae mixing linear arithmetic and uninterpreted functions. Our proof generation scheme benefits from the efficiency of state-of-the-art SMT solvers while being independent from a specific SMT solver proof format. Our only requirement for the SMT solver is the ability to extract unsat cores and generate boolean models. In practice, unsat cores are relatively small and their proof is obtained with a modest overhead by our proof-producing prover. We present experiments assessing the feasibility of the approach for benchmarks obtained from the SMT competition.

This work has been presented at the CPP conference [15] and the international PxTP workshop [21], [20].
5.3. Secure the Clones: Static Enforcement of Policies for Secure Object Copying

Participants: Thomas Jensen, Florent Kirchner, David Pichardie.

Exchanging mutable data objects with untrusted code is a delicate matter because of the risk of creating a data space that is accessible by an attacker. Consequently, secure programming guidelines for Java stress the importance of using defensive copying before accepting or handing out references to an internal mutable object.

However, implementation of a copy method (like clone()) is entirely left to the programmer. It may not provide a sufficiently deep copy of an object and is subject to overriding by a malicious sub-class. Currently no language-based mechanism supports secure object cloning.

We propose a type-based annotation system for defining modular copy policies for class-based object-oriented programs. A copy policy specifies the maximally allowed sharing between an object and its clone. We provide a static enforcement mechanism that will guarantee that all classes fulfill their copy policy, even in the presence of overriding of copy methods, and establish the semantic correctness of the overall approach in Coq.

The mechanism has been implemented and experimentally evaluated on clone methods from several Java libraries. The work as been presented at ESOP [18] this year and is under reviewing for a journal special issue.

5.4. Fault localization and correction in Constraint Programs

Participants: Nadjib Lazaar, Arnaud Gotlieb.

Nowadays, constraint programs are written in high-level modelling languages. Their verification is currently based on trace analysis techniques but does not integrate systematic testing techniques. In this work, we developed a Testing framework for catching the peculiarities of constraint program development, throughout the notions of conformity relations, fault localization and correction.

Within the context of the Nadjib Lazaar’s PhD (defense on 5 Dec. 2011), we explored in 2011 the testing of constraint programs written in OPL and the development of trace-based fault localization and correction techniques [19]. Lazaar’s tool called CPTEST showed impressive experimental results on four hard problems of the CP Community, leading to a publication (in progress) in the Constraints Journal.

5.5. Floating-point constraint solving

Participants: Matthieu Carlier, Arnaud Gotlieb.

Programs including floating-point computations are known to be hard-to-test. Generating test inputs for those programs requires solving constraints over floating-points computations, which led us to the development of specific constraint filtering techniques. In this work, we extended the Marre and Michel property regarding the use of internal floating-point representation to increase the filtering capabilities of addition to the case of multiplication/division. We came up with an optimized implementation of FPSE (our current FP constraint solver) that was able to deal with large C programs that include (non-linear) floating-point computations. We already got a first publication of this work [17].

5.6. Fast inference of polynomial invariants

Participants: David Cachera, Thomas Jensen, Arnaud Jobin, Florent Kirchner.
The problem of automatically inferring polynomial (non-linear) invariants of programs is still a major challenge in program verification. We have proposed an abstract interpretation based method to compute polynomial invariants for imperative programs. Our analysis is a backward propagation approach that computes preconditions for equalities like $g = 0$ to hold at the end of execution. Properties are expressed using ideals, a structure that satisfies the descending chain condition, enabling fixpoints computations to terminate without use of a widening operator. In the general case, termination would be characterized using ideal membership tests and Gröbner bases computations. In order to optimize computational complexity, we propose a specialized analysis dealing with inductive invariants which ensures fast termination of fixpoints computations. The optimized procedure has been shown by experiments to work well in practice, and to be two orders of magnitude faster than state of the art analyzers [23].
5. New Results

5.1. Polychrony as open-source toolset

Participants: Loïc Besnard, Thierry Gautier, Paul Le Guernic.

A major event for us is that the open-source distribution of the Polychrony toolset has been effective since Summer 2011. The Polychrony toolset is described in Section 4.1. Following the considered part of the software, the distribution is made with the GPL V2 or EPL license. One of the objectives of this opening is to make possible a distribution of the software “by apartment” corresponding to a given functionality or to a group of functionalities, for users or developers that would be interested by only a given part of the whole software. To make this possible, a deep restructuration of the whole software has been undertaken. This takes several forms:

- One is related to the polychronous semantics and the transformations that are applied by a compilation process. A typical example is that of the representation type of the Data Control Graph (DCG), for which different levels of representation are distinguished. Some of them are based on the level of representation of the clock hierarchy.
  - The $\text{DCGBasic}$ level is the general type. The DCG represents a program with all dependences set. Clocks are represented as signals of event type.
  - The $\text{DCGPoly}$ level is the subtype of $\text{DCGBasic}$ such that the clock hierarchy in the DCG is the result of the clock calculus. Specific clocks such as $\text{tick}$ are created, but the clock hierarchy, in the general case, has several roots.
  - The $\text{DCGEndo}$ level is the subtype of $\text{DCGPoly}$ such that the clock hierarchy in the DCG is a tree (it is provided with a single root which is $\text{tick}$). The program is endochronous.
  - The $\text{DCGBool}$ level is the subtype of $\text{DCGEndo}$ such that all clock expressions are boolean extractions. Clocks are represented as Boolean signals (no event type is used). Boolean signals representing clocks have themselves clocks represented as Boolean signals (the clock hierarchy still exists).
  - The $\text{DCGSeq}$ level is the subtype of $\text{DCGBool}$ such that all nodes of the graph are statically sorted.
  - The $\text{DCGFlat}$ level is the subtype of $\text{DCGBool}$ such that the clock hierarchy in the DCG is flat: every boolean clock signal is a direct child of the $\text{tick}$. Moreover, each state variable (corresponding to delayed signals) is defined at $\text{tick}$.

- Another aspect of the reorganization is the automatic reconstruction of the toolset from basic components. For that purpose, a new tool, called $pKmake$, has been developed, that allows the architect of the software to describe its structure and construction independently of external tools (such as $\text{emacs}$ that was used previously). It is especially useful for portability reasons, considering the different systems on which the toolset is provided.

- A third aspect that has required special attention is the automatic generation of the documentation of the source, which is realized using $\text{cmake}$, with an automatic management of cross-references.

In the context of the ITEA2 OPEES project, the Polychrony toolset is being provided as base component of the open-source toolchain of the Polarsys platform and Industry Working Group of the Eclipse consortium. A qualification plan will be defined in this context.
5.2. New features of Polychrony

Participants: Loïc Besnard, François Fabre, Thierry Gautier, Paul Le Guernic.

Some new features have been implemented in the Signal toolbox of the Polychrony toolset:

- It is now possible to declare *virtual* objects (types, constants and process models), which are distinguished from external objects, though objects declared as external may also be redefined in the context of declaration. The actual value of an object declared as virtual is provided in the syntactic context of declaration or in a module. A module provides a context of definition for some of the objects described as virtual in the model or the module containing the module importation command. These virtual objects are *overridden* in this way if they are imported (as corresponding objects with the same name) from an imported module, or transitively, from a module imported in an imported module.

- Process models as (static) parameters have been implemented: the formal parameters of the interface of a process model can contain process model parameters, that appear as a formal name of process model typed with a process model type. The call of a process model sets up an expansion context in which an effective process model, designated by its name, is associated with each formal model.

- The connection to the SynDEx tool ([http://www.syndex.org/](http://www.syndex.org/)) has been completed as follows. So far, only the functional part of a given application described in Signal was translated as a corresponding “algorithm” in SynDEx. The multicomponent architecture (typically, processors interconnected through communication medias) and the mapping of the algorithm onto the architecture had to be provided directly within SynDEx. As the polychronous model may be used as intermediate common formalism for applications described in languages where these aspects may be specified (this may be the case in AADL, for instance), they have to be taken into account in the translation. Thus, required elements of the architecture and distribution constraints are described using specific “pragmas” in Signal. These features are then translated into the SynDEx formalism. Using all these information, SynDEx can explore the possible implementations of the algorithm onto the multicomponent.

Moreover, we have redefined the meta-model of Signal in Eclipse, now called SSME (for Signal Syntax meta-model under Eclipse). The SME meta-model, that was used previously, suffered from several drawbacks. It was not fully complete in some parts of the language and, due to design choice, required a strict separation between clock and data flow relations. Thus specific program transformations had to be applied, which did not facilitate traceability. SSME is a full syntax oriented meta-model of Signal, very close to the abstract syntax that is used in the Signal toolbox. Compared to SME, this facilitates model transformations and traceability requirements, which are the primary objectives for its use. The transformation of AADL models into Signal, for instance, now uses the SSME meta-model.

5.3. Extensions of the language and the model

Participants: Loïc Besnard, Thierry Gautier, Paul Le Guernic, Jean-Pierre Talpin.

The different works on using the polychronous model as semantic median model (which has also a syntactic instance) for different effective models (AADL, Simulink via GeneAuto, UML via CCSL...) lead us to study various possible extensions of the semantic model as well as the syntactic one. Some of them have already been defined while for others, the study is still ongoing. In particular, we plan to add to the Signal language a new syntax for automata, partly inspired from AADL mode automata and hierarchic automata existing in other formalisms. An automaton is considered as an instance of a new process model and the “and” composition is the Signal composition.
A fundamental issue that we wish to address in a new way is that of globally asynchronous, locally synchronous (GALS), or globally asynchronous, locally polychronous systems. The idea we have is to extend Signal with a syntactic structure that encapsulates a polychronous (or synchronous) process $P$ in a system, $S$, that creates a continuous temporal domain providing a real-time clock presented in different time units (..., fs, ..., ms, ..., sec, mn, ...). Such a real-time clock can be used as a usual “synchronous” signal in the process $P$ encapsulated in $S$. Systems $S_1$, ..., $S_n$ may be composed (with the standard composition of Signal) in a same system $S$, but the $ms$ of a given system $S_i$ is a priori not synchronous with the $ms$ of another system $S_j$. Then it is possible to specify standard Signal constraints in the system $S$ on these different signals, to express for instance some variation limits of different clocks.

We have also started a new work on causality aspects in order to express and operate more elaborate dependencies than instantaneous dependencies currently computed on the graph of a program. This theoretical work allows one to express dependencies that cross several instants, in a formal framework of word automata and graph algebra.

5.4. Source to source traceability in Polychrony

**Participants:** Loïc Besnard, François Fabre, Thierry Gautier.

To fulfill a mandatory requirement for adoption and qualification of Polychrony environment on the open-source industrial platform of the Polarsys IWG, we have integrated source to source traceability features into the Polychrony toolset. The implementation of traceability is based on the definition of structures of data and algorithms allowing to follow the transformation of objects since the Eclipse modeler of the SME Platform until the generated code. These elements have a direct application with our industrial partners, as, for example, Geensoft with whom, within the framework of the ANR project Spacify, we implemented a simulator of embedded software for satellite applications. We have also integrated such a simulator mode in the Polychrony toolset. Moreover, the error messages from the Signal compiler (Signal Toolbox) are now directly visible on the SME Graphical User Interface and on the Synoptic model (Synoptic is a satellite domain-specific modeling language).

5.5. A simulation infrastructure for CCSL, the timing model of UML MARTE

**Participants:** Huafei Yu, Loïc Besnard, Thierry Gautier, Jean-Pierre Talpin, Paul Le Guernic.

Clock Constraint Specification Language (CCSL) [32] is defined in an annex of the UML MARTE profile [48]. We are interested in the analysis, synthesis and code generation of multi-clocked/polychronous systems specified in CCSL. Timed systems subject to clock expressions or relations can be modeled, specified, analyzed, and simulated within the software environments, such as SCADE [41], TimeSquare [44] and Polychrony. However, code generation from a multi-clocked system is far from obvious. For instance, SCADE always uses a reference or master clock (the fastest); all clocks and all conditions are defined as a functional sampling of this master clock, from the highest specification down to the lowest generated code. In TimeSquare, clock constraints are solved using a heuristic algorithm, which is generally non-deterministic. On the contrary, in Polychrony, a formally defined refinement process yields to the generation of (sequential or concurrent) code by the addition of control variables to get a deterministic behavior satisfying the constraints and allowing the desired amount of concurrency.

The motivation of our work, to address the simulation and code generation of polychronous systems, is to take advantage of the formal framework of Polychrony in the context of a high-level specification formalism, MARTE CCSL [22]. Yet, our work considers a novel approach with regards to previous approaches: to generate executable specifications by considering discrete controller synthesis (DCS) [50], [45], [46]. Clock constraint resolution is addressed by DCS, which does not necessarily require a master clock to address polychronous clocks. In our approach, polychronous (CCSL) specifications are first partitioned: clock relations are considered as control objectives, other constraints are considered as the system to be controlled. The all the constraints are translated into, via SIGNAL, polynomial dynamical systems (PDSs). A PDS represents the transition system of a specification as well as the constraints (invariants) it must satisfy. The Sigali tool
is then used to generate the controller. Finally, the generated controller, together with the original system, is composed to complete the code generation for simulation. In our approach, the temporal semantics of CCSL is mapped onto a polychronous model of computation, on which effective synthesis is carried out to meet constraint requirements. This approach provides both a useful mapping in theory and a flow, which is practical in the generation of reactive controllers.

5.6. The CESAR demonstrator and reference technology platform

Participants: Huafeng Yu, Yue Ma, Loïc Besnard, Thierry Gautier, Jean-Pierre Talpin, Paul Le Guernic.

The design of embedded systems from multiple views and heterogeneous models is ubiquitous in avionics as, in particular, different high-level modeling standards are adopted for specifying the structure, hardware and software components of a system. The system-level simulation of such composite models is necessary but difficult task, allowing to validate global design choices as early as possible in the system design flow. Inspired by the Ptolemy [40], MoBIES [31], SML-Sys [47], etc., we propose an approach to the issue of composing, integrating and simulating heterogeneous models in a system co-design flow [21]. First, the functional behavior of an application is modeled with synchronous data-flow and Statechart diagrams using Simulink/Gene-Auto [54], [55]. These high-level, synchronous and asynchronous, models are then translated into a common model, based on a polychronous model of computation, allowing for a Globally Asynchronous Locally Synchronous (GALS) interpretation of the composed models. This translation is implemented as an automatic model transformation within Polychrony. Simulation, including profiling, value change dump demonstration [24], Syndex adequation [43], etc., is carried out based on the common model within Polychrony.

Polychrony has been integrated to the Reference Technology Platform (RTP) V2 and V3 of CESAR to serve as a framework for co-modeling and architecture exploration. ModelBus [49] is used for the integration of Polychrony into the RTP. ModelBus [25], an integration platform based on Service-Oriented Architecture (SOA), connects different services offered by tools connected to ModelBus. In the demonstration, we participated in the pilot application of Sub-Project 3 (SP3), whose aim is to use the RTP to define a complete software design flow for the doors management system (DMS) of an Airbus A350 in the framework of ModelBus. In the pilot application of the DMS, functional components are modeled in the synchronous model of computation of Simulink, whereas the architecture is modeled in the asynchronous model of computation of AADL [14], [18]. These high-level models are transformed into Signal programs via SME models. Additional models, which are used in the simulation of a closed system, are coded manually in Signal and synchronously composed with the Signal programs transformed from Simulink and AADL models. Finally, C or Java code is generated from Signal programs. Simulation can then be carried out for the purpose of performance evaluation and VCD (Value Change Dump) based demonstration in RTP V2. In RTP V3, Syndex adequation is also integrated to demonstrate real-time scheduling and distribution. Our whole model transformation and simulation chain has been implemented with Galileo Eclipse and attached to ModelBus as a provider of registered remote service. This demonstration also shows the integration of Polychrony with other tools, such as OSATE (AADL), Simulink, Gene-Auto, TimeSquare, ATL, Kermeta, etc.

5.7. Modeling AADL in a polychronous model of computation

Participants: Loïc Besnard, Thierry Gautier, Paul Le Guernic, Jean-Pierre Talpin, Huafeng Yu.

Architecture Analysis and Design Language (AADL) is an SAE standard aimed at high level design and evaluation of architecture of embedded systems. We are interested in the analysis, simulation and verification of timed systems specified in AADL. Polychrony is well suited for the GALS architecture, and it enables deterministic specifications and formal analysis for the design of safety-critical systems. In order to benefit from the advantages provided by Polychrony, a proposition of a methodology for system-level modeling and validation of embedded systems specified in AADL via the polychronous model of computation is proposed.
By studying the different timing semantics of AADL and Polychrony, we have proposed an approach that automatically translates AADL models to a polychronous model of computation (SSME model). In the Polychrony framework, the Signal program can be generated, and an executable model can be obtained. The systems can be analyzed by tools and technologies associated with Polychrony allowing early simulation, testing and verification.

We implemented a plug-in for Eclipse framework to perform model transformation from AADL to SSME (new meta model of Signal). This transformation is implemented in Java. The following new features have been developed this year:

- Temporal interpretation of AADL model. Due to the different timing semantics between AADL and Signal, we keep the ideal view of instantaneous computations of polychronous model, moving computing latencies and communication delays to specific memory process, that introduce delays and well suited synchronizations. Each component modeled in Polychrony is composed of a behavior process (which models the functional behaviors) and a property process (which models the temporal properties).

- Architecture restructures. The architecture of the transformation is optimized. Functional architecture and meta architecture are described to give a global view of the transformation. The translation is recursive. Each AADL component is separated into a java class. The hierarchy of classes are reserved.

- Library developments. We define a Signal library containing the Signal process models representing some basic AADL concepts.

- Documentation. A new technical documentation of the transformation from AADL to SSME has been developed to accompany its implementation. This document aims to provide a global view of our implementation, from a high-level structural view to low-level implementation technical details of components.

- Programming language updates. This version of model transformation uses Java as the programming language. It avoids the disadvantages of dependent on other model transformation languages, and it provides more conveniences and flexibility. The new version is integrated as a plug-in in the Eclipse platform.

- Papers published. Three papers [14], [18], [21] are published this year.

### 5.8. Composing Simulink and AADL

**Participants:** An Phung-Khac, Jean-Pierre Talpin, Benoit Combemale, Jean-Marc Jezequel.

The goal of this work is to improve an import function of the Polychrony environment proposed by the team. Particularly, Polychrony comprises a co-modeling tool supporting the import a high-level Simulink (functional) and AADL (architectural) specifications [21]. This import function is currently implemented by two different transformations, namely Simulink-to-Signal, and AADL-to-Signal. To integrate the Signal programs resulting from these transformations, some Signal interfaces are manually implemented. The composition of Simulink and AADL models thus depends on system designers who implement the interfaces, making difficult its maintenance and validation. To deal with this issue, the model composition approach proposed by the Triskell team, namely ModMap [37], could be used to build a new Simulink and AADL model composition framework.

In ModMap, model composition is considered as a pair of a mapping and an interpretation. A mapping aligns concepts of two meta-models, while the interpretation describes the composition goal. As a model mapping framework, ModMap provides an extensible modeling language supporting the definition of generic mappings and the definition of interpretations. Together with this language, the ModMap kernel is also implemented as an extensible set of mapping processing functions. Model composition frameworks are then built by extending the language and the kernel according to specific composition purposes.
As mentioned above, we intend to apply the ModMap approach to the development of the Simulink and AADL model composition framework. To this end, we need to extend the ModMap mapping language to obtain another one that allows system designers to align elements between Simulink and AADL models regarding the purpose of co-simulation in Signal. Then, a transformation, namely ModMap-to-Signal, needs to be implemented by extending the ModMap kernel. This transformation uses mappings provided by system designers as inputs to generate Signal interfaces. The three transformations (i.e., Simulink-to-Signal, AADL-to-Signal, and ModMap-to-Signal) form the new model composition framework. Compared to the previous one, this framework will more automated. On the other hand, existing transformations will also be reused.

5.9. From affine-related dataflow models to Safety-critical Java

Participants: Adnan Bouakaz, Jean-Pierre Talpin, Jan Vitek.

The objective of this work is to investigate a dataflow concurrency model in order to help specifying, analyzing, and synthesizing functionally deterministic and schedulable SCJ applications. Indeed, the SCJ shared-memory concurrency model makes proving functional determinism and schedulability of applications quite hard if not impossible.

The new model is called the firing related dataflow (FRDF) model in which actors are connected to each other by means of bounded channels. The operational semantics of this model is based on the notion of firing relations. Each actor is associated with a firing clock (an infinite set of activation ticks). The proportionality of the rates of two clocks is expressed by a firing relation. A special and enough expressive case of firing relations is the class of affine relations. Some results about the canonical form of affine relations are already developed by the ESPRESSO team.

Our first study was about synthesizing affine relations between firing clocks in such a way that overflow and underflow exceptions cannot occur during execution. This synthesis is conducted by minimizing the overall of buffer sizes. It is proven that the operational semantics of the dataflow graph based on the computed affine relations is equivalent to the Kahn semantics. This implies that functional determinism is guaranteed.

The previous analysis step (called affine relations synthesis) aims to produce an abstract schedule of the dataflow graph. The computed schedule is abstract in the sense that it is independent from the implementation code of actors and from the target machine. Executing the graph on a mono-processor system using EDF scheduling algorithm is investigated in our study. We synthesize the timing characteristics of each actor (i.e. its period and phase) in such a way the set of tasks is schedulable. In this timing synthesis, we use the worst-case execution times computed from the Java implementation code of actors.

Our objective is to automatically generate a SCJ application from a dataflow specification. Currently, we work on increasing the expressivity of the underlying dataflow model together with providing the necessary analysis tool for generating deterministic and schedulable SCJ code.

5.10. Translation validation of Polychronous Equations with an iLTS Model-checker

Participants: Van-Chan Ngo, Jean-Pierre Talpin, Loïc Besnard.

Synchronous languages such as SIGNAL have been introduced and used successfully for the design and implementation of embedded and critical real-time systems. They rely on the fact that programs are modeled as data-flow equations or finite state machines that allow formal reasoning on designs. In consequence of that, a full toolset of synchronous languages provides formal transformation, automatic code generation, formal verification...
In general, the synchronous language’s compiler takes several translations from the source program before generating the target code (e.g. C/C++ or Java code), thus we present an approach to verify these translations of synchronous language compiler. Our approach adopts the translation validation notion [49]. The idea of translation validation is the following: rather than proving in advance that the compiler always produces correct translations, each individual translation (e.g. every run of the compiler) is followed by a validation phase which verifies that the final output of this run correctly implements the input source program. This method avoids the drawback of freezing the potential improvements and/or developments of the compiler of the traditional compiler verification. For every small change in the compiler, the verification must be redoing the proof, that is an extremely complex task.

The validation phase is made automated which consists of:
(i) Represent both the input source and output target SIGNAL programs as Polynomial Dynamical Systems - PDSs.
(ii) Propose a refinement relation for the PDS models of the source and target programs.
(iii) Use a syntactic simulation-based proof method which automatically verifies the refinement. This automated proof is done by extending the functionality of the model checker SIGALI in the Polychrony toolset.

5.11. PDSs for translation validation: from SIGNAL to C

Participants: Van-Chan Ngo, Jean-Pierre Talpin, Loïc Besnard.

Synchronous programming languages provide a formal and abstract model of concurrency to facilitate the implementation of concurrent embedded software by automating the most complex tasks of verification, validation and code generation. They also guarantee the reliability of the design/implementation of concurrent embedded software by providing either the proof of compiler’s correction or the validation of each run of the compiler. Adopting the translation validation approach [49], we provide an automatic process to formally verify the code C generation task of the SIGNAL’s compiler.

The verification framework will take the SIGNAL program and the generated C code program as the input and proves whether the generated C code correctly implements the SIGNAL program. It also allows to automatically generate the refinement and counterexamples of the generated C code.

Polynomial dynamical system - PDS is used as a common semantic framework to model the behavior of both the SIGNAL program and its generated C code. First, the generated C code is translated into the target SIGNAL program [34] thanks to the intermediate SSA forms. An appropriate relation called refinement for PDSs is proposed to represent the correct implementation relation between the SIGNAL program and its generated C code. The generated code C correctly implements the SIGNAL program if and only if there is a refinement for their PDSs and we say that the generated C code’s PDS refines the SIGNAL program’s PDS. A proof method which allows to generate the refinement or counterexamples, and then proposes a refining process for the generated C code.

5.12. Synchronous symbolic translation systems for translation validation

Participants: Van-Chan Ngo, Jean-Pierre Talpin.

We propose a framework for verification of the correct implementation of the SIGNAL compiler’s generation code task. In order to present the formal semantics of SIGNAL and generated code programs we introduce synchronous symbolic transition system (SSTS) which is the computational model of our formal verification approach. We denote \( \mathcal{D}V = \prod_{i \in [1,n]} \mathcal{D}v_i \) as the domain of a set of variables \( V = (v_1, \ldots, v_n) \). A set of states \( P \subseteq \mathcal{D}V \) is defined as a predicate over the set of variables \( V \) such that the predicate is held in \( P \). An assignment \( A \) is a function \( A: \mathcal{D}V \mapsto \mathcal{D}V \) that the values of the variable set \( V \). A SSTS is a tuple \( L = (V, \Theta, \Gamma, \xi) \) where:

- \( V = (v_1, \ldots, v_n) \) is a set of variables,
- \( \Theta \subseteq \mathcal{D}V \) is a predicate on \( V \) defining the initial condition on the variable set,
- \( \Gamma \) is a finite set of symbolic transitions \( \gamma = (P_\gamma, A_\gamma) \) where:
\[ P_\gamma \subseteq D_V \] is a predicate on \( V \), which guards \( \gamma \)

\[ A_\gamma : D_V \rightarrow D_V \] is the assignment function of \( \gamma \)

- \( E \subseteq V \) is a set of externally observable variables.

The generated code correctly implements the SIGNAL program if and only if there is a refinement for their SSTs and we say that the generated code’s SST refines the SIGNAL program’s SST. This framework also works with SIGNAL programs which is considered as infinite state systems. To obtain the verification results, we apply abstraction interpretation techniques [39] which provide over-approximations of the refinement relation between the input SIGNAL program’s model and the output generated code’s model.

5.13. An integrated environment for Esterel/Quartz and Polychrony/Signal

Participants: Jens Brandt, Ke Sun, Jean-Pierre Talpin.

The design of modern embedded software architectures relies on models and programs built and reused from engineering teams with specific skills and know-how. Each of these skills and backgrounds correspond to specific tools and processes that help implement the viewpoint under consideration with mathematically grounded foundations.

It is not uncommon, for instance, that the design of the only functional views of a system may require the use of tools as heterogeneous and exotic as Catia, Scade, Matlab or Rhapsody. The same holds for design objectives that may range from that of mapping the functional design on specific hardware architectures to that of virtual prototyping for simulation or performance or energy usage evaluation.

Co-modeling itself encompasses the variety of engineering activities that cross the border between the functional and physical views of system design. It is typically the system architects, who will put together functional components and explore different metrics for an effective and efficient mapping on target systems.

We wish to further and scale the framework and experiments developed within the CESAR and OPEES projects in that respect, by thinking a new, domain-specific language, built from synchronous modules designed with Quartz, an imperative synchronous programming language, and connected by data-flow networks described in Signal, the polychronous data-flow language at the core of Polychrony. The combination of viewpoints or paradigms offered by these two design environments provides powerful abstractions and easy to use concepts in order to address two design challenges of utmost importance:

- To provide a natural and dependable specification of elementary synchronous functionalities, most of them algorithmic and control-intensive, in the imperative framework offered by Quartz.
- To synthesize the scheduling of computations and communications among these functionalities starting from the multi-clocked synchronous abstractions offered by the Signal data-flow language.

The remaining long-term goal will then be homogenize this programming framework by further extending it with the capability to control polychromatic networks, seen as modes of execution, with a Quartz module, which would control mode changes.

5.14. Quality assessment and qualification of Polychrony on the open-source Polarsys IWG platform

Participants: Loïc Besnard, Thierry Gautier, Paul Le Guernic, Jean-Pierre Talpin.
Since the open-source release of Polychrony and in the context of the ITEA2 OPEES project, we are collaborating with CS to the integration of Polychrony on the Polarsys platform. This integration proceeds according to guidelines and requirements under definition within the OPEES project and aims first, at putting them to the test. The qualification process of Polychrony in Polarsys consists of checking the maturity level of its implementation and documentation (a standard software engineering assessment) but is also concerned with its capabilities to be composed, inter-operated and mapped with other components on the platform to form a application-specific design toolchain, by using model-driven engineering technologies for model transformation and orchestration. The last phase of this assessment is with regards to its qualifiability, as a simulation tool, as a verification tool, as a code generation tools, which needs to adhere standards such as these defined the in DO178 documents. In parallel, the quality assessment of Polychrony is complemented with a case study, of the APOTA network protocol, whose aim is to document, as a tutorial, the use and added-value of the toolset for its future users on the Polarsys platform.
6. New Results

6.1. Petri Nets and their Synthesis

Participant: Philippe Darondeau.

6.1.1. Petri Net Reachability Graphs: Decidability Status of FO Properties

In [24], we investigate the decidability and complexity status of model-checking problems on unlabelled reachability graphs of Petri nets by considering first-order, modal and pattern-based languages without labels on transitions or atomic propositions on markings. We consider several parameters to separate decidable problems from undecidable ones. Not only are we able to provide precise borders and a systematic analysis, but we also demonstrate the robustness of our proof techniques.

6.1.2. Separability in Persistent Petri Nets

We prove in [14] that the separability of plain, bounded, reversible and persistent Petri nets, a class of nets that extends the well-known live and bounded marked graphs. We establish first a weak form of separability, already known to hold for marked graphs, in which every firing sequence is simulated by a firing sequence of k parallel instances identical firing counts. We establish on top of this a strong form of separability, in which every firing sequence of is simulated by identical firing sequences.

6.1.3. Petri Net Distributability

A Petri net is distributed if, given an allocation of transitions to (geographical) locations, no two transitions at different locations share a common input place. A system is distributable if there is some distributed Petri net implementing it. We address in [23] the question of which systems can be distributed, while respecting a given allocation. The paper states the problem formally and discusses several examples illuminating — to the best of the authors’ knowledge — the current status of this work.

6.2. Heterogeneous Systems

Participants: Eric Badouel, Albert Benveniste, Timothy Bourke, Benoît Caillaud.

6.2.1. Hybrid Modeling

Hybrid modeling tools like Simulink have evolved from simulation platforms into development platforms on which testing, verification and code generation are also performed. It is critical to ensure that the results of simulation, compilation and verification are consistent. Synchronous languages have addressed these issues but only for discrete systems. Reprising earlier work [32], we present in [21] a hybrid modeler built from a synchronous language and an off-the-shelf numerical solver. The main novelty is a language with hierarchical automata that can be arbitrarily mixed with data-flow and ordinary differential equations (ODEs). A type system statically ensures that discrete state changes are aligned with zero-crossing events and that the function passed to the numerical solver has no side-effects during integration. Well-typed programs are compiled by source-to-source translation into synchronous code which is then translated into sequential code using an existing synchronous language compiler.
Starting from a minimal, yet full-featured, Lustre-like synchronous language, we present in [22] a conservative extension where data-flow equations can be mixed with ordinary differential equations (ODEs) with possible reset. A type system is proposed to statically distinguish discrete computations from continuous ones and to ensure that signals are used in their proper domains. We propose a semantics based on non-standard analysis which gives a synchronous interpretation to the whole language, clarifies the discrete/continuous interaction and the treatment of zero-crossings, and also allows the correctness of the type system to be established. The extended data-flow language is realized through a source-to-source transformation into a synchronous subset, which can then be compiled using existing tools into routines that are both efficient and bounded in their use of memory. These routines are orchestrated with a single off-the-shelf numerical solver using a simple but precise algorithm which treats causally-related cascades of zero-crossings. We have validated the viability of the approach through experiments with the Sundials library.

6.2.2. Distributed Structured Documents

Evaluation of attributes w.r.t. an attribute grammar can be obtained by inductively computing a function expressing the dependencies of the synthesized attributes on inherited attributes. This higher-order functional approach to attribute grammars leads to a straightforward implementation using a higher-order lazy functional language like Haskell. The resulting evaluation functions are, however, not easily amenable to optimization rules. We present in [12] an alternative first-order functional interpretation of attribute grammars where the input tree is replaced with an extended cyclic tree each node of which is aware of its context viewed as an additional child tree. By the way, we demonstrate that these cyclic representations of zippers (trees with their context) are natural generalizations of doubly-linked lists to trees over an arbitrary signature.

6.3. Component-Based Design

Participants: Eric Badouel, Albert Benveniste, Benoît Caillaud, Benoît Delahaye, Sophie Pinchinat.

6.3.1. The Modal Interface Theory

In [18], we present the modal interface theory, a unification of interface automata and modal specifications, two radically dissimilar models for interface theories. Interface automata is a game-based model, which allows the designer to express assumptions on the environment and which uses an optimistic view of composition: two components can be composed if there is an environment where they can work together. Modal specifications are a language theoretic account of a fragment of the modal mu-calculus logic with a rich composition algebra which meets certain methodological requirements but which does not allow the environment and the component to be distinguished. The present paper contributes a more thorough unification of the two theories by correcting a first attempt in this direction by Larsen et al., drawing a complete picture of the modal interface algebra, and pushing the comparison between interface automata, modal automata and modal interfaces even further.

6.3.2. A Stochastic Interface Theory

Notions of specification, implementation, satisfaction, and refinement, together with operators supporting stepwise design, constitute a specification theory. In [16], we construct such a theory for Markov Chains (MCs) employing a new abstraction of a Constraint MC. Constraint MCs permit rich constraints on probability distributions and thus generalize prior abstractions such as Interval MCs. Linear (polynomial) constraints suffice for closure under conjunction (respectively parallel composition). This is the first specification theory for MCs with such closure properties. We discuss its relation to simpler operators for known languages such as probabilistic process algebra. Despite the generality, all operators and relations are computable.

6.3.3. Contract-Based Compositional Analysis of Stochastic Systems

A contract allows to distinguish hypotheses made on a system (the guarantees) from those made on its environment (the assumptions). In [17], we focus on models of Assume/Guarantee contracts for (stochastic) systems. We consider contracts capable of capturing reliability and availability properties of such systems. We also show that classical notions of Satisfaction and Refinement can be checked by effective methods thanks to a reduction to classical verification problems. Finally, theorems supporting compositional reasoning and enabling the scalable analysis of complex systems are also studied.
6.3.4. Modal event-clock specifications for timed component-based design

On the one hand, modal specifications are classic, convenient, and expressive mathematical objects to represent interfaces of component-based systems. On the other hand, time is a crucial aspect of systems for practical applications, e.g. in the area of embedded systems. And yet, only few results exist on the design of timed component-based systems. In [13], we propose a timed extension of modal specifications, together with fundamental operations (conjunction, product, and quotient) that enable reasoning in a compositional way about timed system. The specifications are given as modal event-clock automata, where clock resets are easy to handle. We develop an entire theory that promotes efficient incremental design techniques.

6.4. Scheduling and Supervisory Control

Participants: Eric Badouel, Benoît Caillaud, Philippe Daronneau.

6.4.1. Model Identification and Synthesis of Discrete-Event Systems

Book chapter [28] focuses on two important and tightly related problems, namely the identification and synthesis of discrete-event systems. Particular attention is devoted to two main formalisms in this area, i.e., finite state automata and Petri nets. The goal of this chapter is to provide a collection of references in this framework, and discuss the main research areas where such problems have been investigated. Due to the extensive literature, only some of the results are discussed in a certain detail, such as the basic ideas related to the theory of regions and the synthesis of labeled Petri nets, while other results are simply mentioned and the reader is addressed to the specific contributions for more details.

6.4.2. Assembling Sessions

Sessions are a central paradigm in Web services to implement decentralized transactions with multiple participants. Sessions enable the cooperation of workflows while at the same time avoiding the mixing of workflows from distinct transactions. Languages such as BPEL, ORC, AXML that implement Web Services usually realize sessions by attaching unique identifiers to transactions. The expressive power of these languages makes the properties of the implemented services undecidable. In [25], we propose a new formalism for modelling web services. Our model is session-based, but avoids using session identifiers. The model can be translated to a dialect of Petri nets that allows the verification of important properties of web services.

6.4.3. Towards Distributed Control of Discrete Event Systems

To initiate a discussion on the modeling requirements for distributed control of discrete-event systems, a partially-automated region based methodology is presented in [26]. The methodology is illustrated via a well-known example from distributed computing: the dining philosophers.

6.4.4. Communicating Decentralized Control

Frameworks that incorporate communication into decentralized supervisory control theory address the following problem: find locations in the evolution of the plant behavior where some supervisors send information so that a supervisor that was unable to make the correct control decision prior to receiving external information, is now capable of making the correct control decision. We propose in [19] a solution to this problem and identify an earliest and a latest placement where such communication results in the synthesis of a correct control solution. In addition to a first and last communication opportunity, there may be a selection of intermediate possibilities where communication would produce the correct control solution. We present a computable procedure to identify a broader range of suitable communication locations.

6.4.5. Residuation of tropical series: rationality issues

In [20], the decidability of existence, rationality of delay controllers and robust delay controllers are investigated for systems with time weights in the tropical and interval semirings. Depending on the (max,+)- or (min,+) -rationality of the series specifying the controlled system and the control objective, cases are identified where the controller series defined by residuation is rational, and when it is positive (i.e., when delay control is feasible). When the control objective is specified by a tolerance, i.e. by two bounding rational series, a nice case is identified in which the controller series is of the same rational type as the system specification series.
6.5. Games, Logic and System Synthesis

Participants: Bastien Maubert, Sophie Pinchinat.

6.5.1. Opacity Issues in Games with Imperfect Information

In [27], we study the class of games with opacity condition, which are two-player games with imperfect information in which one of the players only has imperfect information, and where the winning condition relies on the information he has along the play. Those games are relevant for security aspects of computing systems: a play is opaque whenever the player who has imperfect information never "knows" for sure that the current position is one of the distinguished "secret" positions. We study the problems of deciding the existence of a winning strategy for each player, and we call them the opacity-violate problem and the opacity-guarantee problem. Focusing on the player with perfect information is new in the field of games with imperfect-information because when considering classical winning conditions it amounts to solving the underlying perfect-information game. We establish the EXPTIME-completeness of both above-mentioned problems, showing that our winning condition brings a gap of complexity for the player with perfect information, and we exhibit the relevant opacity-verify problem, which noticeably generalizes approaches considered in the literature for opacity analysis in discrete-event systems. In the case of blindfold games, this problem relates to the two initial ones, yielding the determinacy of blindfold games with opacity condition and the PSPACE-completeness of the three problems.

6.5.2. Hardness of preorder checking for basic formalisms

We investigate in [15] the complexity of preorder checking when the specification is a flat finite-state system whereas the implementation is either a non-flat finite-state system or a standard timed automaton. In both cases, we show that simulation checking is Exptime-hard, and for the case of a non-flat implementation, the result holds even if there is no synchronization between the parallel components and their alphabets of actions are pairwise disjoint. Moreover, we show that the considered problems become Pspace-complete when the specification is assumed to be deterministic. Additionally, we establish that comparing a synchronous non-flat system with no hiding and a flat system is Pspace-hard for any relation between trace containment and bisimulation equivalence.
6. New Results

6.1. Octagonal Domains for Continuous Constraints (continuous)

**Participants:** Marie Pelleau, Charlotte Truchet, Fréderic Benhamou.

Domains in Continuous Constraint Programming (CP) are generally represented with intervals whose \( n \)-ary Cartesian product (box) approximates the solution space. We propose a new representation for continuous variable domains based on octagons [28]. We generalize local consistency and split to this octagon representation, and we propose an octagonal-based branch and prune algorithm [22]. Experimental results in IBEX on the COCONUT benchmarks suite show promising performance improvements on several classical benchmarks.

The corresponding paper *Octagonal Domains for Continuous Constraints* got the Best Student Paper Award at the 17th International Conference on Principles and Practice of Constraint Programming (CP 2011) [22].

6.2. A Constraint Seeker (classification, constraints and application)

**Participants:** Nicolas Beldiceanu, Helmut Simonis.

We design a Constraint Seeker application which provides a web interface to search for global constraints in the global constraint catalog, given positive and negative ground examples. Based on the given instances the tool returns a ranked list of matching constraints, the rank indicating whether the constraint is likely to be the intended constraint of the user. A systematic evaluation is provided over the complete global constraint catalog.

The corresponding paper *A Constraint Seeker: Finding and Ranking Global Constraints from Examples* [16] was published at the 17th International Conference on Principles and Practice of Constraint Programming (CP 2011) and the corresponding tool is available as a web service at [http://www.emn.fr/z-info/sdemasse/gccat/](http://www.emn.fr/z-info/sdemasse/gccat/).

6.3. Soft Problems (modelling and filtering)

**Participants:** Thierry Petit, Alexis De Clerq, Nicolas Beldiceanu, Narendra Jussien.

1. **Side-constrained problems** We experimentally shown that solving some classes of over-constrained problems requires an efficient (global) propagation of side constraints on variables representing violations. This work completes our previous studies, which highlighted the interest of a variable-based representation of violations for sake of modelling.

   The corresponding paper *Global Propagation of Side Constraints for Solving Over-constrained Problems* was published in the *Annals of Operations Research* journal [14].

2. **Soft cumulative scheduling** We proposed a new constraint for handling cumulative problems with exceeds of capacities, in the case where the time horizon is fixed and the capacity can vary over time. Sweep and Edge-finding algorithms for classical cumulative problems have been modified so as to provide a filtering algorithm for our constraint. Experiments shown that instances involving several hundreds of activities can be solved with our approach.

   The corresponding paper *Filtering Algorithms for Discrete Cumulative Problems with Over-loads of Resource* was published at the 17th International Conference on Principles and Practice of Constraint Programming (CP 2011), [17].
6.4. Constraints with costs (modelling and filtering)  
Participants: Thierry Petit, Nicolas Beldiceanu.

1. **Distribution of costs** We presented a new cardinality constraint dedicated to sequences of totally ordered cost. This constraint is useful to impose a precise (fair) distribution of the values taken by the cost variables in a given sequence, for instance in a bin packing with safety rules or in cumulative scheduling with overloads of resource. We came up with a generalized arc-consistency filtering algorithm, whose time complexity is linear in the sum of the number of variables and the number of values in the union of their domains.

The corresponding paper *the Ordered Distribute Constraint* was published in the International Journal on Artificial Intelligence Tools [15].

2. **The objective sum constraint.** Constraint toolkits generally provide a sum constraint whose propagation is poor to solve optimization problems. Therefore, solving real-world problems requires to develop ad-hoc techniques for handling sums, based on the particular properties of each problem. We proposed a generic technique which improves the standard sum constraint by exploiting the propagation of a set of constraints defined on the variables involved in a sum.

The corresponding paper *The Objective Sum Constraint* was published in the 8th International Conference on Integration of Artificial Intelligence and Operations Research Techniques in Constraint Programming for Combinatorial Optimization Problems (CPAIOR 2011) [25].

3. **the increasing sum constraint** Given a sequence of variables $X$ of length $n$, we consider the increasing sum constraint, which imposes the variables of $X$ to be sorted in non strictly order, and that the sum of the variables of $X$ is equal to $s$. We propose an linear time bound-consistency algorithm for increasing sum. This work is related to problems with variable symmetries, when some of the symmetric variables are involved in sum constraints.

The paper *A Theta(n) Bound-Consistency Algorithm for the Increasing Sum Constraint* was published at the 17th International Conference on Principles and Practice of Constraint Programming (CP 2011), [24].

These works were all done in collaboration with J.-C. Régin (Univ. Nice).

6.5. Efficient Filtering Algorithms for Generic Constraints (filtering)  
Participants: Nicolas Beldiceanu, Xavier Lorca, Thierry Petit.

**Counting constraints** We identified a family of counting constraints for which performing a complete filtering is a tractable problem. We provided a generalized arc-consistency algorithm and its specialization to some well-known global constraints. For some of them the obtained time complexity is linear in the sum of domain sizes, which improve or equals the best known results in the literature.

The corresponding paper *A Generalized Arc-Consistency Algorithm for a Class of Counting Constraints* was published at the 22th International Joint Conference on Artificial Intelligence (IJCAI’11) [23].

6.6. Efficient Filtering Algorithms and Heuristics for Dedicated Constraints (filtering)  
Participants: Jean-Guillaume Fages, Xavier Lorca, Arnaud Malapert, Narendra Jussien.

1. **Revisiting the tree constraint** We revisit the tree constraint introduced at CPAIOR 2005 in [35] which partitions the nodes of a $n$-nodes, $m$-arcs directed graph into a set of node-disjoint anti-arborescences for which only certain nodes can be tree roots. We introduce in a new filtering algorithm that enforces generalized arc-consistency in $O(n+m)$ time while the original filtering algorithm reaches $O(nm)$ time. This result allows to tackle larger scale problems involving graph partitioning in CHOCO.
2. **An Optimal Constraint Programming Approach to the Open-Shop Problem** We present an optimal constraint programming approach for the Open-Shop problem, which integrates recent constraint propagation and branching techniques with new upper bound heuristics for the Open-Shop problem. Randomized restart policies combined with nogood recording allow to search diversification and learning from restarts. This approach closed all remaining problems of the Brucker et al. and Guéret and Prins benchmarks with cpu times that are orders of magnitude lower than the best known metaheuristics.

The corresponding paper *An Optimal Constraint Programming Approach to the Open-Shop Problem* was published in the INFORMS Journal on Computing [13]. This work was done in collaboration with H. Cambazard (4C, INP Grenoble), C. Guéret (IRCCyN), A. Langevin (Ecole Polytechnique Montréal) and L.-M. Rousseau (Ecole Polytechnique Montréal).

6.7. **Explanations for Constraint Programming (solver)**

**Participants:** Narendra Jussien, Charles Prud’homme.

Constraint programming, despite its numerous successes in practice, suffers from not being really user-friendly when used by software engineers. Indeed, when faced with a *no solution* message from a constraint solver, it is hard yet impossible to identify the cause of this message: is it from a bad modelling, an ill-written constraint, a bug in the solver, .... Explanations for constraint programming have addressed this issue but are not yet widely used in the CP community. Recent work in the field tend to demonstrate that providing explanation-based user-oriented features can be done quite easily in modern constraint solvers. The objective of this line of work is to specify an user-oriented explanation-module for flexible solver architectures. This line of work is financed through a Google focused research grant. First results provide a complete solver independent specification of explanation algorithms, data structure for encoding nogoods and treatment algorithms. A reference implementation is being made within the new version of our solver CHOCO.

6.8. **Bin repacking (constraints and application)**

**Participants:** Sophie Demassay, Xavier Lorca, Fabien Hermenier.

A datacenter can be viewed as a dynamic bin packing system where servers host applications with varying resource requirements and varying relative placement constraints. When those needs are no longer satisfied, the system has to be reconfigured. Virtualization allows to distribute applications into Virtual Machines (VMs) to ease their manipulation. In particular, a VM can be freely migrated without disrupting its service, temporarily consuming resources both on its origin and destination.

We introduce the Bin Repacking Scheduling Problem in this context. This problem is to find a final packing and to schedule the transitions from a given initial packing, accordingly to new resource and placement requirements, while minimizing the average transition completion time. Our CP-based approach uses CHOCO and is implemented into Entropy, an autonomous VM manager which detects reconfiguration needs, generates and solves the CP model, then applies the computed decision. CP provides the awaited flexibility to handle heterogeneous placement constraints and the ability to manage large datacenters with up to 2000 servers and 10000 VMs.

The corresponding paper *Bin-Repacking Scheduling in Virtualized Datacenters* was published at the 17th International Conference on Principles and Practice of Constraint Programming (CP 2011), [19].

6.9. **A Global Constraint for Multi-Agent Localization (constraints and application)**

**Participants:** Gilles Chabert, Sophie Demassay.
This work has been initiated in the context of the Angels research project, in which G. Chabert has been involved during two years. The idea was to provide a new method for inter-localizing a group of autonomous underwater robots, traditional Kalman-based methods being inadequate in this context due to the highly nonlinear models derived from the sensing technology (electric fish robots).

We proposed, through a rough discretization of the signals, to consider the problem as a whole and under a combinatorial form. The level of the signal is basically associated to a cardinality of surrounding objects. This led to a global constraint, namely a conjunction of among constraint with interval value domains and in multiple dimension (objects are variables with several components).

Conjunction of among constraints had been already studied but not in the case of interval value domains. We therefore conducted a theoretical study and proved that the problem was tractable in the one-dimensional case, but not in higher dimension. We have also investigated different decompositions and filtering algorithms. This work is submitted to CPAIOR 2012. An INRIA research report has also been issued in June 2011, where the robotics aspects are described.
6. New Results

6.1. Verification

6.1.1. Analysis of partially observed recursive discrete-event systems

Participants: Sébastien Chédor, Thierry Jéron, Hervé Marchand, Christophe Morvan.

Monitoring of recursive discrete-event systems under partial observation is an important issue with major applications such as the diagnosability of faulty behaviors and the detection of information flow. We consider regular discrete-event systems, that is recursive discrete-event systems definable by deterministic graph grammars. This setting is expressive enough to capture classical models of recursive systems such as the pushdown systems. Hence they are infinite-state in general and standard powerset constructions for monitoring do not apply anymore. We exhibit computable conditions on these grammars together with non-trivial transformations of graph grammars that enable us to construct a monitor. This construction is applied to diagnose faulty behaviors, to detect information flow in regular discrete-event systems, and to generate tests.

6.1.2. Analysis of timed systems

6.1.2.1. Approximate determinization of timed automata

Participants: Nathalie Bertrand, Thierry Jéron, Amélie Stainer.

Timed automata are frequently used to model real-time systems. Their determinization is a key issue for several validation problems. However, not all timed automata can be determinized, and determinizability itself is undecidable. In [18], we propose a game-based algorithm which, given a timed automaton, tries to produce a language-equivalent deterministic timed automaton, otherwise a deterministic over-approximation. Our method subsumes two recent contributions: it is at once more general than an existing (non terminating) determinization procedure by Baier et al. (2009) and more precise than the approximation algorithm of Krichen and Tripakis (2009). Moreover, an extension of the method allows to deal with invariants and $\epsilon$-transitions, and to consider other useful approximations: under approximation, and combination of under- and over-approximations which are particularly useful in testing (see 6.2.1).

6.1.2.2. Frequency analysis for timed automata

Participants: Nathalie Bertrand, Amélie Stainer.

The languages of infinite timed words accepted by timed automata are traditionally defined using Büchi-like conditions. These acceptance conditions focus on the set of locations visited infinitely often along a run, but completely ignore quantitative timing aspects. In [15] we propose a natural quantitative semantics for timed automata based on the so-called frequency, which measures the proportion of time spent in the accepting states. We study various properties of timed languages accepted with positive frequency, and in particular the emptiness and universality problems.

6.1.3. Petri nets reachability graphs

Participant: Christophe Morvan.

Petri nets are a general model for concurrency, the structure of their reachability graph is mostly unknown. In [19] we have investigated the decidability and complexity status of model-checking problems on unlabelled reachability graphs of Petri nets by considering first-order, modal and pattern-based languages without labels on transitions or atomic propositions on markings. We consider several parameters to separate decidable problems from undecidable ones. These results illustrate the intrinsic complexity of the structure of these graphs.
6.2. Active and passive testing

6.2.1. Off-line test selection with test purposes for non-deterministic timed automata

Participants: Nathalie Bertrand, Thierry Jéron, Amélie Stainer.

In [17], we propose novel off-line test generation techniques for non-deterministic timed automata with inputs and outputs (TAIOs) in the formal framework of the tioco conformance theory. In this context, a first problem is the determinization of TAIOs, which is necessary to foresee next enabled actions, but is in general impossible. The determinization problem is addressed in [18] thanks to an approximate determinization using a game approach (see 6.1.2.1). We adapt this procedure here to over- and under-approximation, in order to preserve tioco and guarantee the soundness of generated test cases. A second problem is test selection for which a precise description of timed behaviors to be tested is carried out by expressive test purposes modeled by a generalization of TAIOs. Finally, using a symbolic co-reachability analysis guided by the test purpose, test cases are generated in the form of TAIOs equipped with verdicts.

6.2.2. Test generation using pushdown automata

Participant: Puneet Bhateja.

IOLTS (input output labeled transition system) is a versatile model and is frequently used in model based testing to model the functional behavior of an IUT (implementation under test). However when a system is tested remotely, its observed behavior can be different from its actual functional behavior. In a previous paper, we defined a notion of remotely observed behavior of an IOLTS in terms of its actual behavior. Paper [14] contributes by proposing a methodology to simulate a PDA (pushdown automaton) from the given IOLTS such that the simulated PDA precisely expresses the remotely observed behavior of the IOLTS. The simulated PDA can be thought of as an automatic test generator for remote testing.

6.2.3. Test case selection in asynchronous testing

Participants: Puneet Bhateja, Thierry Jéron.

Conformance testing has a rich underlying formal theory called IOLTS-based conformance testing. Depending upon whether the implementation-under-test (IUT) interacts with its environment directly, or indirectly through a medium, IOLTS-based conformance testing can be classified as synchronous testing or asynchronous testing, respectively. So far the problem of test case selection has been addressed mostly in the context of synchronous testing. In this work we contribute by addressing this problem in the context of asynchronous testing. Though an asynchronously communicating process can be simulated by a synchronously communicating process, the fact that the simulating process is infinite state even if the simulated process is finite state made the problem challenging.

6.2.4. A tagging protocol for asynchronous testing

Participant: Puneet Bhateja.

Conformance testing has a rich underlying theory popularly called IOCO-test theory. In the realm of IOCO-test theory, this paper addresses the issue of testing a component of an asynchronously communicating distributed system. Testing a system which communicates asynchronously (i.e., through some medium) with its environment is more difficult than testing a system which communicates synchronously (i.e., directly without any medium). What impedes asynchronous testing is that the actual behavior of the implementation under test (IUT) appears distorted and infinite to the tester. This impediment consequently renders the problem of generating a complete test suite, from the given specification of the IUT, infeasible. To this end, paper [13] proposes a tagging protocol which when implemented by the asynchronously communicating distributed system will enable the generation of a complete test suite, from the specification of any of its component. Further, this paper describes how to generate the test suite from the given specification of the component.

6.2.5. Abstracting time and data for conformance testing of real-time systems

Participants: Thierry Jéron, Hervé Marchand.
Current approaches to model-based conformance testing of real-time systems are mostly based either on finite state machines/transition systems or on timed automata. However, most real-time systems manipulate data while being subjected to time constraints. The usual solution consists in enumerating data values (in finite domains) while treating time symbolically, thus leading to the classical state explosion problem. Paper [12] with W.L. Andrade and P. Machado (Fed. Univ. Campina Grande) proposes a new model of real-time systems as an extension of both symbolic transition systems and timed automata, in order to handle both data and time requirements symbolically. We then adapt the tioco conformance testing theory to deal with this model and describe a test case generation process based on a combination of symbolic execution and constraint solving for the data part and symbolic analysis for timed aspects.

6.2.6. Ensuring security properties

6.2.6.1. Runtime enforcement monitors: composition, synthesis, and enforcement abilities

**Participant:** Yliès Falcone.

Runtime enforcement is a powerful technique to ensure that a program will respect a given set of properties. In [9] we extend previous work on this topic in several directions. Firstly, we propose a generic notion of enforcement monitors based on a memory device and finite sets of control states and enforcement operations. Moreover, we specify their enforcement abilities w.r.t. the general Safety-Progress classification of properties. Furthermore, we propose a systematic technique to produce a monitor from the automaton recognizing a given safety, guarantee, obligation or response property. Finally, we show that this notion of enforcement monitors is more amenable to implementation and encompasses previous runtime enforcement mechanisms.

6.2.6.2. What can you verify and enforce at runtime?

**Participant:** Yliès Falcone.

The underlying property, its definition and representation play a major role when monitoring a system. Having a suitable and convenient framework to express properties is thus a concern for runtime analysis. It is desirable to delineate in this framework the sets of properties for which runtime analysis approaches can be applied to. [8] presents a unified view of runtime verification and enforcement of properties in the Safety-Progress classification. Firstly, we extend the Safety-Progress classification of properties in a runtime context. Secondly, we characterize the set of properties which can be verified (monitorable properties) and enforced (enforceable properties) at runtime. We propose in particular an alternative definition of ”property monitoring” to the one classically used in this context. Finally, for the delineated sets of properties, we define specialized verification and enforcement monitors.

6.3. Control synthesis

6.3.1. Controllers for probabilistic systems

**Participant:** Nathalie Bertrand.

*Partially Observable Markov Decision Processes (POMDP for short) have been extensively studied in several research communities, among which AI and model-checking. In [16] we address the problem of the minimal information a user needs at runtime to achieve a simple goal, modeled as reaching an objective with probability one. More precisely, to achieve her goal, the user can either choose at each step to use partial information only, or pay a fixed cost and receive full information. The natural question is then to minimize the cost the user needs to fulfill its objective. This optimization question gives rise to two different problems, whether we consider to minimize the worst case cost, or the average cost. On the one hand, concerning the worst case cost, we show that efficient techniques from the model checking community can be adapted to compute the optimal worst case cost and give optimal strategies for the users. On the other hand, we show that the optimal average price (a question typically considered in the AI community) cannot be computed in general, nor can it be approximated in polynomial time even up to a large approximation factor.*
6.3.2. Supervisory control for synchronous systems

6.3.2.1. Controller synthesis and programming language

Participant: Hervé Marchand.

In [24], we define a mixed imperative/declarative programming language: declarative contracts are enforced upon imperatively described behaviors. We rely on the notion of Discrete Controller Synthesis (DCS), a formal technique stemming from control theory and the supervisory control of discrete event systems. We target the application domain of adaptive and reconfigurable computing systems: our language can serve programming closed-loop adaptation controllers, enabling flexible execution of functionalities w.r.t. changing resource and environment conditions. We give a synthetic presentation of the language, its semantics and compilation, and we illustrate its use with the example of a robot system.

6.3.2.2. Symbolic supervisory control of infinite transition systems under partial observation using abstract interpretation

Participant: Hervé Marchand.

In [11], we propose algorithms for the synthesis of state-feedback controllers with partial observation of infinite state discrete event systems modelled by Symbolic Transition Systems. We provide models of safe memoryless controllers both for potentially deadlocking and deadlock free controlled systems. The termination of the algorithms solving these problems is ensured using abstract interpretation techniques which provide an overapproximation of the transitions to disable. We then extend our algorithms to controllers with memory and to online controllers. We also propose improvements in the synthesis of controllers in the finite case which, to our knowledge, provide more permissive solutions than what was previously proposed in the literature. Our tool SMACS gives an empirical validation of our methods by showing their feasibility, usability and efficiency.

6.3.2.3. Decentralized control of infinite systems

Participant: Hervé Marchand.

In [10], we propose algorithms for the synthesis of decentralized state-feedback controllers with partial observation of infinite state systems, which are modeled by Symbolic Transition Systems. We first consider the computation of safe controllers ensuring the avoidance of a set of forbidden states and then extend this result to the deadlock free case. The termination of the algorithms solving these problems is ensured by the use of abstract interpretation techniques, but at the price of overapproximations, in particular, in the computation of the states which must be avoided. We then extend our algorithms to the case where the system to be controlled is given by a collection of subsystems (modules). This structure is exploited to locally compute a controller for each module. Our tool SMACS gives an empirical evaluation of our methods by showing their feasibility, usability and efficiency.

6.3.2.4. Polychronous controller synthesis from MARTE CCSL timing specifications

Participant: Hervé Marchand.

The UML Profile for Modeling and Analysis of Real-Time and Embedded systems (MARTE) defines a mathematically expressive model of time, the Clock Constraint Specification Language (CCSL), to specify timed annotations on UML diagrams and thus provides them with formally defined timed interpretations. Thanks to its expressive capability, the CCSL allows for the specification of static and dynamic properties, of deterministic and non-deterministic behaviors, or of systems with multiple clock domains. Code generation from such multicolocated specifications (for the purpose of synthesizing a simulator, for instance) is known to be a difficult issue. We address it in [23] by using the approach of controller synthesis. In our framework, a timed CCSL specification is regarded as a property whose satisfaction should be enforced for any UML diagram carrying it as annotation. To do so, CCSL statements are first translated into dynamical polynomial systems. Such systems can be manipulated using the model-checker Sigali to synthesize an executable property (a controller) which enforces the satisfaction of the specified timing constraints on the UML diagram with which it is executed.
6.3.3. Control of distributed systems

Participant: Hervé Marchand.

In this work, we consider the control of distributed systems composed of subsystems communicating asynchronously; the aim is to build local controllers that restrict the behavior of a distributed system in order to satisfy a global state avoidance property. We model our distributed systems as communicating finite state machines with reliable unbounded FIFO queues between subsystems. Local controllers can only observe the behavior of their local subsystem and do not see the queue contents. To refine their control policy, the controllers can use the FIFO queues to communicate by piggybacking extra information (some timestamps and their state estimates) to the messages sent by the subsystems [21]. We provide an algorithm that computes, for each local subsystem (and thus for each controller), during the execution of the system, an estimate of the current global state of the distributed system. The local estimate is updated at each message reception. We then define synthesis algorithms allowing to compute the local controllers. Our method relies on the computation of (co)reachable states. Since the reachability problem is undecidable in our model, we use abstract interpretation techniques to obtain regular overapproximations of the possible FIFO queue contents, and hence of the possible current global states. An implementation of our algorithms provides an empirical evaluation of our method [22].
5. New Results

5.1. Reactive trajectories for molecular dynamics

Participants: Frédéric Cérou, Arnaud Guyader, Florent Malrieu.

See 3.3 and 4.2.

This is a collaboration with Tony Lelièvre (CERMICS, Ecole des Ponts ParisTech).

Consider a one-dimensional Brownian motion in a double well potential. It is known that, as its variance goes to zero, the Brownian particle has to wait for a longer and longer time to jump from a well to the other. This metastable behavior is described by the Freidlin–Wentzell theory. We are investigating the length of the paths between the last passage time in the bottom of a well and the hitting time of the other one (reactive trajectory). In the case of an Ornstein–Uhlenbeck process between the wells, we obtained the remarkable result that the time length of the reactive trajectories converges in distribution when the noise intensity goes to zero, to a Gumbel random variable shifted by a deterministic term growing like minus the logarithm of the noise intensity. Our numerical simulations are also in good accordance with this theoretical result. We are also very close to extend this result to more general one dimensional diffusion processes.

5.2. Learning the optimal importance distribution

Participants: François Le Gland, Rudy Pastel.

See 3.3 and 4.2.

This is a collaboration with Jérôme Morio (ONERA Palaiseau).

As explained in 3.3, multilevel splitting ideas can be useful even to solve some static problems, such as evaluating the (small) probability that a random variable exceeds some (extreme) threshold. Incidentally, it can also be noticed that a population of particles is available at each stage of the algorithm, which is distributed according to the original distribution conditioned on exceeding the current level. Furthermore, this conditional distribution is known to be precisely the optimal importance distribution for evaluating the probability of exceeding the current level. In other words, the optimal importance distribution is learned automatically by the algorithm, as a by–product, and therefore can be used to produce an importance sampling estimate with very low variance. Building on this idea, several other iterative methods have been studied, that learn the optimal importance distribution at each stage of the algorithm, such as nonparametric adaptive importance sampling (NAIS) [69], or the cross–entropy (CE) method [65], [33]. These methods have been applied to a practical example from the aerospace industry, the evaluation of collision probabilities between two satellites, or between a satellite and space debris.

5.3. Impact of dimension in particle filtering and the Laplace method

Participants: François Le Gland, Paul Bui–Quang.

See 3.1.

This is a collaboration with Christian Musso (ONERA Palaiseau).
Particle filtering is a widely used Monte Carlo method to approximate the posterior probability distribution in non-linear filtering, with an error scaling as $1/\sqrt{N}$ in terms of the sample size $N$, but otherwise independently of the underlying state dimension. However, it has been observed for a long time in practice that particle filtering can be quite inefficient when the dimension of the system is high. In a simple static linear Gaussian model, it has been possible indeed to check that the error on the estimation of the predicted likelihood, a quantitative indicator of the consistency between the prior distribution and the likelihood function, increases exponentially with the dimension \[30\]. This preliminary result has been extended to a non-linear / non-Gaussian model, using the Laplace method \[23\]. The Laplace method, which approximates multidimensional integrals accurately, has also been used to compute the asymptotic variance of the importance weights, as the sample size $N$ goes to infinity, and to analyze its dependence on the dimension.

5.4. Stochastic wind-wave modelling

**Participant:** Valérie Monbet.

This is a collaboration with Pierre Ailliot (UBO) and Christophe Maisondieu (IFREMER).

Directional wave spectra generally exhibit several peaks due to the coexistence of wind sea generated by local wind conditions and swells originating from distant weather systems. This paper proposes a new algorithm for partitioning such spectra and retrieving the various systems which compose a complex sea-state. It is based on a sequential Monte-Carlo algorithm which allows to follow the time evolution of the various systems. The main particularity of the algorithm is that the dimension of the hidden state can change from time to time and a model selection step is included. The proposed methodology is validated on both synthetic and real spectra and the results are compared with a method commonly used in the literature.

5.5. Sequential data assimilation: ensemble Kalman filter vs. particle filter

**Participants:** François Le Gland, Valérie Monbet.

*See 6.4.*

Surprisingly, very little was known about the asymptotic behaviour of the ensemble Kalman filter \[38\], \[39\], \[40\], whereas on the other hand, the asymptotic behaviour of many different classes of particle filters is well understood, as the number of particles goes to infinity. Interpreting the ensemble elements as a population of particles with mean-field interactions, and not only as an instrumental device producing an estimation of the hidden state as the ensemble mean value, it has been possible to prove the convergence of the ensemble Kalman filter, with a rate of order $1/\sqrt{N}$, as the number $N$ of ensemble elements increases to infinity \[25\]. In addition, the limit of the empirical distribution of the ensemble elements has been exhibited, which differs from the usual Bayesian filter. The next step has been to prove (by induction) the asymptotic normality of the estimation error, i.e. to prove a central limit theorem for the ensemble Kalman filter.
6. New Results

6.1. identification of linear systems

6.1.1. Modular identification and damage detection for large structures

**Participants:** Michael Döhler, Laurent Mevel.

In Operational Modal Analysis (OMA) of large structures it is often needed to process sensor data from multiple non-simultaneously recorded measurement setups, especially in the case of large structures. In this work a new efficient variant of the PreGER algorithm is presented that avoids the numerical explosion of the calculation by using a modular approach, where the data from the measurement setups is processed setup by setup and not at the same time [16].

6.1.2. Fast multi order subspace identification algorithm

**Participants:** Michael Döhler, Laurent Mevel.

Stochastic subspace identification methods are an efficient tool for system identification of mechanical systems in Operational Modal Analysis (OMA), where modal parameters are estimated from measured vibrational data of a structure. System identification is usually done for many successive model orders, as the true system order is unknown and identification in results at different model orders need to be compared to distinguish true structural modes from spurious modes in so-called stabilization diagrams. An algorithm to estimate the system matrices at multiple model orders has been derived [20].

6.1.3. Evaluation of confidence intervals and computation of sensitivities for subspace methods

**Participants:** Michael Döhler, Xuan Lam, Laurent Mevel.

In Operational Modal Analysis, the modal parameters (natural frequencies, damping ratios and mode shapes) obtained from Stochastic Subspace Identification (SSI) of a structure, are afflicted with statistical uncertainty. A variant of this approach has been derived for the Eigenvalue-Realization-Algorithm (ERA) [25]. Another version has been proposed for the merging subspace algorithm [17], [17]. This approach has been validated on large scale examples [14].

6.2. damage detection for mechanical structures

6.2.1. Damage detection and localisation

**Participants:** Michael Döhler, Laurent Mevel.

Statistical methods using output-only data have been shown to offer a robust solution to the damage detection task. These techniques have also been combined with sensitivities extracted from finite element models to offer information on the location of damage accounting for uncertainties in the finite element sensitivities. In some applications, however, the formulation of the finite element model makes implementation impractical and this motivates the search for model-free damage localization alternatives. One option is to use experimentally extracted sensitivities but their computation requires a set of constants (usually absorbed in the normalization of the eigenvectors) that are not available in output only identification. The noted limitation can be circumvented by adding a known perturbation to the mass distribution and repeating the output only identification, a procedure that can be practical in some cases. Linking a null-space based subspace damage index with experimentally extracted sensitivities allows us to infer on the position of damage without formulating a
finite element model and without the need for input measurements. The performance of the algorithm is illustrated on simulated data [19]. Damage detection has also been applied to a large scale example of an European project [23], [15].

6.2.2. Robust subspace damage detection

**Participants:** Michael Döhler, Laurent Mevel.

Subspace methods enjoy some popularity, especially in mechanical engineering, where large model orders have to be considered. In the context of detecting changes in the structural properties and the modal parameters linked to them, some subspace-based fault detection residual has been recently proposed and applied successfully. However, most works assume that the unmeasured ambient excitation level during measurements of the structure in the reference and possibly damaged condition stays constant, which is not possible in any application. This work addresses the problem of robustness of such fault detection methods. A subspace-based fault detection test is derived that is robust to excitation change but also to numerical instabilities that could arise easily in the computations [21].

6.3. Instability monitoring of aeronautical structures

6.3.1. Subspace identification for hinged-blades helicopters

**Participants:** Ahmed Jhinaoui, Laurent Mevel.

In this work, an extension of the output-only subspace identification, to the class of linear periodically time-varying (LPTV) systems, is proposed. The goal is to identify a useful information about the system’s stability using the Floquet theory which gives a necessary and sufficient condition for stability analysis [24].

6.3.2. Optimal input design for identification and detection

**Participants:** Alireza Esna Ashari, Laurent Mevel.

Output only techniques rely on the presence on unknown turbulence, which may or may not be enough to excite the system. A new approach for applying artificial input to the system for maximizing detection and identifiability has been developed. This work considers the problem of auxiliary input design for subspace-based fault detection methods. In several real applications, particularly in the damage detection of mechanical structures and vibrating systems, environment noise is the only input to the system. In some applications, white noise produces low-quality output data for the subspace-based fault detection method. In those methods, a residual is calculated to detect the fault based on the output information. However, some modes of the system may not influence the outputs and the residual appropriately if the input is not exciting enough for those modes. In this work, rotated inputs method is implemented to excite the system modes. In addition to produce a residual more sensitive to the weak modes, it is possible to detect system order changes due to the fault using the rotated inputs. Simulation results demonstrate the efficiency of injecting these auxiliary inputs to improve the subspace-based fault detection methodology [22]. This work is funded by FP7-NMP Large Scale Integrated Project IRIS.
5. New Results

5.1. Asymptotic preserving schemes

**Participant:** Nicolas Crouseilles.

In [18], we extend the micro-macro decomposition based numerical schemes developed previously to the collisional Vlasov-Poisson model in the diffusion and high-field asymptotics. In doing so, we first write the Vlasov-Poisson model as a system that couples the macroscopic (equilibrium) part with the remainder part. A suitable discretization of this micro-macro model enables to derive an asymptotic preserving scheme in the diffusion and high-field asymptotics. In addition, two main improvements are presented: On the one hand a self-consistent electric field is introduced, which induces a specific discretization in the velocity direction, and represents a wide range of applications in plasma physics. On the other hand, as suggested in a previous reference, we introduce a suitable reformulation of the micro-macro scheme which leads to an asymptotic preserving property with the following property: It degenerates into an implicit scheme for the diffusion limit model when $\varepsilon \to 0$, which makes it free from the usual diffusion constraint $\Delta t = O(\Delta x^2)$ in all regimes. Numerical examples are used to demonstrate the efficiency and the applicability of the schemes for both regimes.

In [45], a Two-Scale Macro-Micro decomposition of the Vlasov equation with a strong magnetic field is derived. This consists in writing the solution of this equation as a sum of two oscillating functions with circumscribed oscillations. The first of these functions has a shape which is close to the shape of the Two-Scale limit of the solution and the second one is a correction built to offset this imposed shape.

5.2. Resolution of the quasi-neutrality equation

**Participant:** Nicolas Crouseilles.

In reference [39], different parallel algorithms are proposed for the numerical resolution of the quasi-neutrality equation in the GYSELA code. A set of benchmarks on a parallel machine has permitted to evaluate the performance of the different versions of the quasi-neutrality solver. In particular, in [40], these improvements are combined with memory optimization which enable a scalability of the GYSELA code up to 64k cores.

In [20], a new discretization scheme of the gyrokinetic quasi-neutrality equation is proposed. It is based on Isogeometric Analysis; the IGA which relies on NURBS functions, seems to accommodate arbitrary coordinates and the use of complicated computation domains. Moreover, arbitrary high order degree of basis functions can be used. Here, this approach is successfully tested on elliptic problems like the quasi-neutrality equation.

5.3. High order schemes for Vlasov-Poisson system

**Participant:** Nicolas Crouseilles.

In [44], we derive the order conditions for fourth order time splitting schemes in the case of the 1D Vlasov-Poisson system. Computations to obtain such conditions are motivated by the specific Poisson structure of the Vlasov-Poisson system: this structure is similar to Runge-Kutta-Nyström systems. The obtained conditions are proved to be the same as RKN conditions derived for ODE up to the fourth order. Numerical results are performed and show the benefit of using high order splitting schemes in that context.
In [19], we present a discontinuous Galerkin scheme for the numerical approximation of the one-dimensional periodic Vlasov-Poisson equation. The scheme is based on a Galerkin-characteristics method in which the distribution function is projected onto a space of discontinuous functions. We present comparisons with a semi-Lagrangian method to emphasize the good behavior of this scheme when applied to Vlasov-Poisson test cases.

The CEMRACS is an annual summer research session promoted by the SMAI. The 15th edition of 2010 has been organized by N. Crouseilles, H. Guillard, B. Nkonga and E. Sonnendrücker around "Numerical modeling of fusion plasmas". The volume [38] gathers artless resulting from research projects initiated during the CEMRACS 2010.

5.4. Second order averaging for the nonlinear Schrödinger equation with strong anisotropic potential

**Participants:** Florian Méhats, François Castella.

In [10], we consider the three dimensional Gross-Pitaevskii equation (GPE) describing a Bose-Einstein Condensate (BEC) which is highly confined in vertical \( z \) direction. The confining potential induces high oscillations in time. If the confinement in the \( z \) direction is a harmonic trap – an approximation which is widely used in physical experiments – the very special structure of the spectrum of the confinement operator implies that the oscillations are periodic in time. Based on this observation, it can be proved that the GPE can be averaged out with an error of order \( \epsilon \), which is the typical period of the oscillations. In this article, we construct a more accurate averaged model, which approximates the GPE up to errors of order \( O(\epsilon^2) \). Then, expansions of this model over the eigenfunctions (modes) of the confining operator \( H_z \) in the \( z \)-direction are given in view of numerical applications. Efficient numerical methods are constructed to solve the GPE with cylindrical symmetry in 3D and the approximation model with radial symmetry in 2D, and numerical results are presented for various kinds of initial data.

5.5. A problem of moment realizability in quantum statistical physics

**Participant:** Florian Méhats.

This work [34] is a generalization of the results previously obtained by F. Méhats and O. Pinaud, in J. Stat. Phys. (2010), in a one-dimensional setting: we revisit the problem of the minimization of the quantum free energy (entropy + energy) under local constraints (moments) and prove the existence of minimizers in various configurations. While the above quoted article addressed the 1D case on bounded domains, we treat in the present paper the multi-dimensional case as well as unbounded domains and non-linear interactions as Hartree/Hartree-Fock. Moreover, whereas this article dealt with the first moment only, namely the charge density, we extend the results to the second moment, the current density.

5.6. Orbital stability of spherical galactic models

**Participant:** Florian Méhats.

In [33] we consider the three dimensional gravitational Vlasov Poisson system which is a canonical model in astrophysics to describe the dynamics of galactic clusters. A well known conjecture is the stability of spherical models which are nonincreasing radially symmetric steady states solutions. This conjecture was proved at the linear level by several authors in the continuation of the breakthrough work by Antonov in 1961. In a previous work, we derived the stability of anisotropic models under *spherically symmetric perturbations* using fundamental monotonicity properties of the Hamiltonian under suitable generalized symmetric rearrangements first observed in the physics literature. In this work, we show how this approach combined with a *new generalized* Antonov type coercivity property implies the orbital stability of spherical models under general perturbations.
5.7. The Schrödinger Poisson system on the sphere

Participant: Florian Méhats.

In [31] we study the Schrödinger-Poisson system on the unit sphere $S^2$ of $\mathbb{R}^3$, modeling the quantum transport of charged particles confined on a sphere by an external potential. Our first results concern the Cauchy problem for this system. We prove that this problem is regularly well-posed on every $H^s(S^2)$ with $s > 0$, and not uniformly well-posed on $L^2(S^2)$. The proof of well-posedness relies on multilinear Strichartz estimates, the proof of ill-posedness relies on the construction of a counterexample which concentrates exponentially on a closed geodesic. In a second part of the paper, we prove that this model can be obtained as the limit of the three dimensional Schrödinger-Poisson system, singularly perturbed by an external potential that confines the particles in the vicinity of the sphere.

5.8. A boundary matching micro-macro decomposition for kinetic equations

Participant: Florian Méhats.

In [32], we introduce a new micro-macro decomposition of collisional kinetic equations which naturally incorporates the exact space boundary conditions. The idea is to write the distribution function $f$ in all its domain as the sum of a Maxwellian adapted to the boundary (which is not the usual Maxwellian associated with $f$) and a reminder kinetic part. This Maxwellian is defined such that its ‘incoming’ velocity moments coincide with the ‘incoming’ velocity moments of the distribution function. Important consequences of this strategy are the following. i) No artificial boundary condition is needed in the micro/macro models and the exact boundary condition on $f$ is naturally transposed to the macro part of the model. ii) It provides a new class of the so-called ‘Asymptotic preserving’ (AP) numerical schemes: such schemes are consistent with the original kinetic equation for all fixed positive value of the Knudsen number $\epsilon$, and if $\epsilon \leftarrow 0$ with fixed numerical parameters then these schemes degenerate into consistent numerical schemes for the various corresponding asymptotic fluid or diffusive models. Here, the strategy provides AP schemes not only inside the physical domain but also in the space boundary layers. We provide a numerical test in the case of a diffusion limit of the one-group transport equation, and show that our AP scheme recovers the boundary layer and a good approximation of the theoretical boundary value, which is usually computed from the so-called Chandrasekhar function.

5.9. 1D quintic nonlinear equation with white noise dispersion

Participant: Arnaud Debussche.

Under certain scaling the nonlinear Schrödinger equation with random dispersion converges to the nonlinear Schrödinger equation with white noise dispersion. The aim of these works is to prove that this latter equation is globally well posed in $L^2$ or $H^1$. In [28], we improve the Strichartz estimates obtained previously for the Schrödinger equation with white noise dispersion in one dimension. This allows us to prove global well posedness when a quintic critical nonlinearity is added to the equation. We finally show that the white noise dispersion is the limit of smooth random dispersion.

5.10. Weak approximation of stochastic partial differential equations: the nonlinear case

Participant: Arnaud Debussche.

In [22] we study the error of the Euler scheme applied to a stochastic partial differential equation. We prove that as it is often the case, the weak order of convergence is twice the strong order. A key ingredient in our proof is Malliavin calculus which enables us to get rid of the irregular terms of the error. We apply our method to the case a semilinear stochastic heat equation driven by a space-time white noise.

5.11. Ergodic BSDEs under weak dissipative assumptions

Participant: Arnaud Debussche.
In [27] we study ergodic backward stochastic differential equations (EBSDEs) dropping the strong dissipativity assumption needed previously. In other words we do not need to require the uniform exponential decay of the difference of two solutions of the underlying forward equation, which, on the contrary, is assumed to be non degenerate. We show existence of solutions by use of coupling estimates for a non-degenerate forward stochastic differential equations with bounded measurable non-linearity. Moreover we prove uniqueness of “Markovian” solutions exploiting the recurrence of the same class of forward equations. Applications are then given to the optimal ergodic control of stochastic partial differential equations and to the associated ergodic Hamilton-Jacobi-Bellman equations.

5.12. Asymptotic first exit times of the Chafee-Infante equation with small heavy tailed noise

Participant: Arnaud Debussche.

Motivated by paleoclimatological issues, we determine in [26] asymptotic first exit times for the Chafee-Infante equation forced by heavy-tailed Levy diffusions from reduced domains of attraction in the limit of small intensity. We show that in contrast to the case of Gaussian diffusion the expected first exit times are polynomial in terms of the intensity.

5.13. Stochastic Cahn-Hilliard equation with double singular nonlinearities and two reflections

Participant: Arnaud Debussche.

In [25] we consider a stochastic partial differential equation with two logarithmic nonlinearities, two reflections at 1 and -1, and a constraint of conservation of the space average. The equation, driven by the derivative in space of a space-time white noise, contains a bi-Laplacian in the drift. The lack of a maximum principle for the bi-Laplacian generates difficulties for the classical penalization method, which uses a crucial monotonicity property. Being inspired by the works of Debussche, Goudenège, and Zambotti, we obtain existence and uniqueness of a solution for initial conditions in the interval \((-1, 1)\). Finally, we prove that the unique invariant measure is ergodic, and we give a result of exponential mixing.

5.14. Diffusion limit for a stochastic kinetic problem

Participants: Arnaud Debussche, Erwan Faou.

In [29] we consider numerical approximations of stochastic differential equations by the Euler method. In the case where the SDE is elliptic or hypo-elliptic, we show a weak backward error analysis result in the sense that the generator associated with the numerical solution coincides with the solution of a modified Kolmogorov equation up to high order terms with respect to the stepsize. This implies that every invariant measure of the numerical scheme is close to a modified invariant measure obtained by asymptotic expansion. Moreover, we prove that, up to negligible terms, the dynamic associated with the Euler scheme is exponentially mixing.

5.15. Convergence of stochastic gene networks to hybrid piecewise deterministic processes

Participant: Arnaud Debussche.

In [47] we consider numerical approximations of stochastic differential equations by the Euler method. In the case where the SDE is elliptic or hypo-elliptic, we show a weak backward error analysis result in the sense that the generator associated with the numerical solution coincides with the solution of a modified Kolmogorov equation up to high order terms with respect to the step-size. This implies that every invariant measure of the numerical scheme is close to a modified invariant measure obtained by asymptotic expansion. Moreover, we prove that, up to negligible terms, the dynamic associated with the Euler scheme is exponentially mixing.
5.16. **Exponential mixing of the 3D stochastic Navier-Stokes equations driven by mildly degenerate noise**

**Participant:** Arnaud Debussche.

In [11], we prove the strong Feller property and exponential mixing for 3D stochastic Navier-Stokes equation driven by mildly degenerate noises (i.e. all but finitely many Fourier modes are forced) via Kolmogorov equation approach.

5.17. **Ergodicity results for the stochastic Navier-Stokes equations: an introduction**

**Participant:** Arnaud Debussche.

In this survey article [46], we review recent progresses in the study of ergodicity for the stochastic Navier-Stokes equations. The first part introduces general concept, the second deals with the 2D case and the 3D case is treated in the third part.

5.18. **Local Martingale and Pathwise Solutions for an Abstract Fluids Model**

**Participant:** Arnaud Debussche.

In the first article [23], we establish the existence and uniqueness of both local martingale and local pathwise solutions of an abstract nonlinear stochastic evolution system. The primary application of this abstract framework is to infer the local existence of strong, pathwise solutions to the 3D primitive equations of the oceans and atmosphere forced by a nonlinear multiplicative white noise. In the second article [24], global existence is obtained.

5.19. **Geometric numerical integration and Schrödinger equations**

**Participant:** Erwan Faou.

The goal of geometric numerical integration is the simulation of evolution equations by preserving their geometric properties over long times. This question is of particular importance in the case of Hamiltonian partial differential equations typically arising in many application fields such as quantum mechanics or wave propagations phenomena. This implies many important dynamical features such as energy preservation and conservation of adiabatic invariants over long times. In this setting, a natural question is to know how and to which extent the reproduction of such long time qualitative behavior is ensured by numerical schemes.

Starting from numerical examples, these notes [37] try to provide a detailed analysis in the case of the Schrödinger equation in a simple setting (periodic boundary conditions, polynomial nonlinearities) approximated by symplectic splitting methods. This text analyzes the possible stability and instability phenomena induced by space and time discretization, and provides rigorous mathematical explanations for them.

5.20. **On the influence of the geometry on skin effect in electromagnetism**

**Participant:** Erwan Faou.

In [14], we consider the equations of electromagnetism set on a domain made of a dielectric and a conductor subdomain in a regime where the conductivity is large. Assuming smoothness for the dielectric-conductor interface, relying on recent works we prove that the solution of the Maxwell equations admits a multiscale asymptotic expansion with profile terms rapidly decaying inside the conductor. This skin effect is measured by introducing a skin depth function that turns out to depend on the mean curvature of the boundary of the conductor. We then confirm these asymptotic results by numerical experiments in various axisymmetric configurations. We also investigate numerically the case of a nonsmooth interface, namely a cylindrical conductor.
5.21. Reconciling alternate methods for the determination of charge distributions: A probabilistic approach to high-dimensional least-squares approximations

Participant: Erwan Faou.

In [17], we propose extensions and improvements of the statistical analysis of distributed multipoles (SADM) algorithm put forth by CHIPOT in 1998 for the derivation of distributed atomic multipoles from the quantum-mechanical electrostatic potential. The method is mathematically extended to general least-squares problems and provides an alternative approximation method in cases where the original least-squares problem is computationally not tractable, either because of its ill-posedness or its high-dimensionality. The solution is approximated employing a Monte Carlo method that takes the average of a random variable defined as the solutions of random small least-squares problems drawn as subsystems of the original problem. The conditions that ensure convergence and consistency of the method are discussed, along with an analysis of the computational cost in specific instances.

5.22. Hamiltonian interpolation of splitting approximations for nonlinear PDEs

Participant: Erwan Faou.

In [30], we consider a wide class of semi linear Hamiltonian partial differential equations and their approximation by time splitting methods. We assume that the nonlinearity is polynomial, and that the numerical trajectory remains at least uniformly integrable with respect to an eigenbasis of the linear operator (typically the Fourier basis). We show the existence of a modified interpolated Hamiltonian equation whose exact solution coincides with the discrete flow at each time step over a long time. While for standard splitting or implicit-explicit schemes, this long time depends on a cut-off condition in the high frequencies (CFL condition), we show that it can be made exponentially large with respect to the step size for a class of modified splitting schemes.

5.23. Energy cascades for NLS on the torus

Participant: Erwan Faou.

In the work [16], we consider the nonlinear Schrödinger equation with cubic (focusing or defocusing) nonlinearity on the multidimensional torus. For special small initial data containing only five modes, we exhibit a countable set of time layers in which arbitrarily large modes are created. The proof relies on a reduction to multiphase weakly nonlinear geometric optics, and on the study of a particular two-dimensional discrete dynamical system.

5.24. A Nekhoroshev type theorem for the nonlinear Schrödinger equation on the d-dimensional torus

Participant: Erwan Faou.

In [49] we prove a Nekhoroshev type theorem for the nonlinear Schrödinger equation

\[ iu_t = -\Delta u + V \ast u + \partial_\Pi g(u, \Pi) , \quad x \in \mathbb{T}^d, \]

where \( V \) is a typical smooth Fourier multiplier and \( g \) is analytic in both variables. More precisely we prove that if the initial datum is analytic in a strip of width \( \rho > 0 \) whose norm on this strip is equal to \( \epsilon \) then, if \( \epsilon \) is small enough, the solution of the nonlinear Schrödinger equation above remains analytic in a strip of width \( \rho/2 \), with norm bounded on this strip by \( C \epsilon \) over a very long time interval of order \( \epsilon^{-\alpha/\ln \epsilon^{\beta}} \), where \( 0 < \beta < 1 \) is arbitrary and \( C > 0 \) and \( \alpha > 0 \) are positive constants depending on \( \beta \) and \( \rho \).
5.25. Sobolev stability of plane wave solutions to the cubic nonlinear Schrödinger equation on a torus

**Participant:** Erwan Faou.

In [48], it is shown that plane wave solutions to the cubic nonlinear Schrödinger equation on a torus behave orbitally stable under generic perturbations of the initial data that are small in a high-order Sobolev norm, over long times that extend to arbitrary negative powers of the smallness parameter. The perturbation stays small in the same Sobolev norm over such long times. The proof uses a Hamiltonian reduction and transformation and, alternatively, Birkhoff normal forms or modulated Fourier expansions in time.

5.26. Approximate travelling wave solutions to the 2D Euler equation on the torus

**Participants:** Erwan Faou, Nicolas Crouseilles.

In [43], we consider the two-dimensional Euler equation with periodic boundary conditions. We construct approximate solutions of this equation made of localized travelling profiles with compact support propagating over a stationary state depending on only one variable. The direction or propagation is orthogonal to this variable, and the support is concentrated around flat points of the stationary state. Under regularity assumptions, we prove that the approximation error can be made exponentially small with respect to the width of the support of the travelling wave. We illustrate this result by numerical simulations.

5.27. Markov chains competing for transitions: applications to large-scale distributed systems

**Participant:** François Castella.

We consider in [12] the behavior of a stochastic system composed of several identically distributed, but non independent, discrete-time absorbing Markov chains competing at each instant for a transition. The competition consists in determining at each instant, using a given probability distribution, the only Markov chain allowed to make a transition. We analyze the first time at which one of the Markov chains reaches its absorbing state. We obtain its distribution and its expectation and we propose an algorithm to compute these quantities. We also exhibit the asymptotic behavior of the system when the number of Markov chains goes to infinity. Actually, this problem comes from the analysis of large-scale distributed systems and we show how our results apply to this domain.

5.28. Analysis of a large number of Markov chains competing for transitions

**Participant:** François Castella.

This text [41] generalizes the previous one [12] in the following sense.

In the situation on the previous article, we analyze the first time at which one of the Markov chains reaches its absorbing state. When the number of Markov chains goes to infinity, we analyze the asymptotic behavior of the system for an arbitrary probability mass function governing the competition. We give conditions for the existence of the asymptotic distribution and we show how these results apply to cluster-based distributed storage when the competition is handled using a geometric distribution.

5.29. Splitting methods with complex coefficients for some classes of evolution equations

**Participant:** Philippe Chartier.
We are concerned in [13] with the numerical solution obtained by splitting methods of certain parabolic partial differential equations. Splitting schemes of order higher than two with real coefficients necessarily involve negative coefficients. In a previous paper, Castella et al. demonstrated the possibility to overcome this second-order barrier by considering splitting methods with complex-valued coefficients and built up methods of orders 3 to 14. In this paper, we reconsider the technique employed therein and show that it is inherently bound to order 14 and largely sub-optimal with respect to error constants. As an alternative, we solve directly the algebraic equations arising from the order conditions and construct several methods of orders 4, 6, 8 and 16 that are the most accurate ones available at present time.

5.30. Higher-order averaging, formal series and numerical integration

**Participant:** Philippe Chartier.

The paper [42] considers non-autonomous oscillatory systems of ordinary differential equations with \(d = 1\) non-resonant constant frequencies. Formal series like those used nowadays to analyze the properties of numerical integrators are employed to construct higher-order averaged systems and the required changes of variables. With the new approach, the averaged system and the change of variables consist of vector-valued functions that may be written down immediately and scalar coefficients that are universal in the sense that they do not depend on the specific system being averaged and may therefore be computed once and for all. The new method may be applied to obtain a variety of averaged systems. In particular we study the quasi-stroboscopic averaged system characterized by the property that the true oscillatory solution and the averaged solution coincide at the initial time. We show that quasi-stroboscopic averaging is a geometric procedure because it is independent of the particular choice of co-ordinates used to write the given system. As a consequence, quasi-stroboscopic averaging of a canonical Hamiltonian (resp. of a divergence-free) system results in a canonical (resp. in a divergence-free) averaged system. We also study the averaging of a family of near-integrable systems where our approach may be used to construct explicitly formal first integrals for both the given system and its quasi-stroboscopic averaged version. As an application we construct three first integrals of a system that arises as a nonlinear perturbation of coupled harmonic oscillators with one slow frequency and four resonant fast frequencies.

The stroboscopic averaging method (SAM) is a technique for the integration of highly oscillatory differential systems \(\dot{y} = f(y, t)\) with a single high frequency. The method may be seen as a purely numerical way of implementing the analytical technique of stroboscopic averaging which constructs an averaged differential system \(\dot{Y} = F(Y)\) whose solutions \(Y\) interpolate the sought highly oscillatory solutions \(y\). SAM integrates numerically the averaged system without using the analytic expression of \(F\); all information on \(F\) required by the algorithm is gathered on the fly by numerically integrating the originally given system in small time windows. SAM may be easily implemented in combination with standard software and may be applied with variable step sizes. Furthermore it may also be used successfully to integrate oscillatory DAEs. The paper [15] provides an analytic and experimental study of SAM and two related techniques: the LISP algorithms of Kirchgraber and multirevolution methods.
6. New Results

6.1. Fluid motion estimation

6.1.1. Multiscale PIV method based on turbulent kinetic energy decay

**Participants:** Patrick Héas, Dominique Heitz, Etienne Mémin.

We have proposed a new multiscale PIV method based on turbulent kinetic energy decay. The technique is based on scaling power laws describing the statistical structure of turbulence. A spatial regularization constrains the solution to behave through scales as a self similar process via second-order structure function and a given power law. The real parameters of the power-law, corresponding to the distribution of the turbulent kinetic energy decay, have been estimated from a simple hot-wire measurement. The method has been assessed in a turbulent wake flow and grid turbulence through comparisons with HWA measurements and other PIV approaches. Results have indicated that the present method is superior because it accounts for the whole dynamic range involved in the flows.

6.1.2. Stochastic uncertainty models for motion estimation

**Participants:** Thomas Corpetti, Etienne Mémin.

In this work we have proposed a stochastic formulation of the brightness consistency used principally in motion estimation problems. In this formalization the image luminance is modeled as a continuous function transported by a flow known only up to some uncertainties. Stochastic calculus enables to built then conservation principles which take into account the motion uncertainties. These uncertainties defined either from isotropic or anisotropic models can be estimated jointly to the motion estimates. Such a formulation besides providing estimates of the velocity field and of its associated uncertainties allows us to define a natural linear scale space multiresolution framework. The corresponding estimator implemented within a local least squares approach has shown to improve significantly the results of the corresponding deterministic estimator (Lucas and Kanade estimator). This fast local motion estimator has been shown to provide results that are in the same order of accuracy than state-of-the-art dense fluid flow motion estimator for particle images. This work has been published in a conference proceeding and has been accepted for publication in the journal IEEE Trans. on Image Processing [23], [16]. We intend to pursue this formalization to define dense motion estimators that allows handling in the same way luminance conservation under motion uncertainty principles.

6.1.3. 3D flows reconstruction from image data

**Participants:** Ioana Barbu, Dominique Heitz, Cédric Herzet, Etienne Mémin.

Our work focusses on the design of new tools for the problem of 3D reconstruction of a turbulent flow motion. This task includes both the study of physically-sound models on the observations and the fluid motion, and the design of low-complexity and accurate estimation algorithms. On the one hand, state-of-the-art methodologies such as “sparse representations” will be investigated for the characterization of the observation and fluid motion models. Sparse representations are well-suited to the representation of signals with very few coefficients and offer therefore advantages in terms of computational and storage complexity. On the other hand, the estimation problem will be placed into a probabilistic Bayesian framework. This will allow the use of state-of-the-art inference tools to effectively exploit the strong time-dependence of the fluid motion. In particular, we will investigate the use of “ensemble Kalman” filter to devise low-complexity sequential estimation algorithms.
This year, we have more particularly focussed on the problem of reconstructing the particle positions from several two-dimensional images. Our approach is based on the exploitation of a particular family of sparse representation algorithms, namely the so-called “pursuit algorithms”. Indeed, the pursuit procedures generally allow a good trade-off between performance and complexity. Hence, we have performed a thorough study comparing the reconstruction performance and the complexity of different state-of-the-art algorithms to that achieved with pursuit algorithms. This work has led to the publication of two conference papers in experimental fluid mechanics [21], [34].

6.1.4. Motion estimation techniques for turbulent fluid flows

Participants: Patrick Héas, Dominique Heitz, Cédric Herzet, Etienne Mémin.

Based on physical laws describing the multi-scale structure of turbulent flows, this article proposes a regularizer for fluid motion estimation from an image sequence. Regularization is achieved by imposing some scale invariance property between histograms of motion increments computed at different scales. By reformulating this problem from a Bayesian perspective, an algorithm is proposed to jointly estimate motion, regularization hyper-parameters, and to select the most likely physical prior among a set of models. Hyper-parameter and model inference is conducted by likelihood maximization, obtained by marginalizing out non-Gaussian motion variables. The Bayesian estimator is assessed on several image sequences depicting synthetic and real turbulent fluid flows. Results obtained with the proposed approach in the context of fully developed turbulence improve significantly the results of state-of-the-art fluid flow dedicated motion estimators. This work has been published in several conferences and in the journal Tellus, Serie A [18].

6.1.5. Wavelet basis for multi-scale motion estimation

Participants: Pierre Dérian, Patrick Héas, Cédric Herzet, Souleymane Kadri Harouna, Etienne Mémin.

This work aims at exploring wavelet representations for fluid motion estimation from consecutive images. This scale-space representation, associated to a simple gradient-based optimization algorithm, sets up a natural multi-resolution framework for the optical flow estimation well suited to medium range velocity magnitude. Moreover, a very simple closure mechanism, approaching locally the solution by high-order polynomials, is provided by truncating the wavelet basis at fine scales. Well-known turbulence regularities and multifractal behaviors on the reconstructed motion field can also be imposed on the wavelet coefficients. Accuracy and efficiency of the proposed method has been evaluated on scalar and particles image sequences of turbulent fluid flows. Particularly good results have been observed for particle image velocimetry. This offers a very interesting alternative to traditional PIV techniques. This work has been published in computer vision or turbulence conferences [26], [25].

6.1.6. Divergence-free wavelet basis and high-order regularization

Participants: Pierre Dérian, Patrick Héas, Souleymane Kadri Harouna, Etienne Mémin.

Expanding on a wavelet basis the solution of an inverse problem provides several advantages. Wavelet bases yield a natural multiresolution analysis that may alleviate the use of Gauss Newton strategy for medium range motion amplitude. The continuous representation of the solution with wavelets enables analytical calculation of regularization integrals over the spatial domain. By choosing differentiable wavelets, high-order derivative regularizers can be designed, either taking advantage of the wavelet differentiation properties or via the basis’s mass and stiffness matrices. Moreover, differential constraints on vector solutions, such as the divergence-free volume preserving constraint, can be handled with biorthogonal wavelet bases. Numerical results on synthetic and real images of incompressible turbulence show that divergence-free wavelets and high-order regularizers are particularly relevant in the context of incompressible fluid flows. This work has been partly published in a conference proceeding [25].

6.1.7. Divergence-free wavelet basis and high-order regularization

Participants: Pierre Dérian, Patrick Héas, Souleymane Kadri Harouna.
This work presents a method for regularization of inverse problems. The vectorial bi-dimensional unknown is assumed to be the realization of an isotropic divergence-free fractional Brownian Motion (fBm). The method is based on fractional Laplacian and divergence-free wavelet bases. The main advantage of these bases is to enable an easy formalization in a Bayesian framework of fBm priors, by simply sampling wavelet coefficients according to Gaussian white noise. Fractional Laplacians and the divergence-free projector can naturally be implemented in the Fourier domain. An interesting alternative is to remain in the spatial domain. This is achieved by the analytical computation of the connection coefficients of divergence-free fractional Laplacian wavelets, which enables to easily rotate this simple prior in any sufficiently “regular” wavelet basis. Taking advantage of the tensorial structure of a separable fractional wavelet basis approximation, isotropic regularization is then computed in the spatial domain by low-dimensional matrix products. The method is successfully applied to fractal image restoration and turbulent optic-flow estimation.

### 6.1.8. Bayesian inference of hyper-parameters and models in motion estimation

**Participants:** Patrick Héas, Cédric Herzet, Etienne Mémin.

Bayes rule provides a nice framework for motion estimation from image sequences. We rely on a hierarchical modeling linking the image intensity function variable, the motion field variable, hyper-parameters composed of the likelihood and prior model inverse variances and of robust parameters, and finally the observation and prior model. The variable dependence can thus be expressed as a 4-level hierarchy. Applying the Bayes rule on this hierarchy, we obtain three levels of inference, which enable us to obtain, by marginalizing out intermediate variables, a direct dependence of the variable of interest to the image intensity function. Thus, the estimates of regularization parameters, of robust parameters associated to semi-quadratic norms of a family of M-estimators, and of observation and prior models are inferred in a maximum likelihood sense while maximizing jointly the motion field a posteriori probability. The quality of the method is demonstrated on synthetic and real two-dimensional turbulent flows and on several computer vision scenes of the "Middlebury" data-base. This work has been accepted for publication in IEEE transaction on Image Processing (IP) [17].

### 6.1.9. Method to quantify the uncertainty of motion measurement

**Participants:** Patrick Héas, Dominique Heitz, Cédric Herzet.

Measurement uncertainty is a general concept associated with any measurement that can be used to quantify the confidence of the estimation. The ‘Guide to the expression of uncertainty in measurement’ (GUM) provides a framework to account for all uncertainties and then to propagate them. However, in particle image velocimetry (PIV), measurement uncertainty estimation is a tricky task since it has to be done through the propagation of distributions via Monte Carlo simulations for each velocity component and all pixel location in the image. Considering a standard Bayesian formulation of the optical flow problem together with a Gaussian assumption the uncertainty associated to the estimated velocity field has been provided and described. First PIV measurement uncertainty estimations and discussions have been recently published in a conference issue [36].

### 6.1.10. Sparse-representation algorithms

**Participant:** Cédric Herzet.

We have pursued the study of efficient sparse decomposition algorithms. In particular, we have addressed the problem of finding good sparse representations into a probabilistic framework [24]. First, we have showed that one of the standard formulations - the Lagrangian formulation - of this problem can be interpreted as a limit case of a maximum a posteriori (MAP) problem involving Bernoulli-Gaussian variables. Then, we have proposed different tractable implementations of this MAP problem and explained some well-known pursuit algorithms (MP, OMP, StOMP, CoSaMP and SP) as particular cases of the proposed algorithms. Experimentations led on synthetic data show a good general behavior of the proposed methods. Exploiting further this probabilistic framework, we have then considered the design of soft pursuit algorithms. In particular, instead of making hard decisions on the support of the sparse representation and the amplitude of the non-zero coefficients, our soft procedures iteratively update probability on the latter values. The proposed algorithms are designed within the framework of the mean-field approximations and resort to the so-called variational Bayes EM algorithm to implement an efficient minimization of a Kullback-Leibler criterion.
6.2. Tracking and data assimilation

6.2.1. Stochastic filtering for fluid motion tracking

Participants: Sébastien Béyou, Anne Cuzol, Sai Gorthi, Etienne Mémin.

We investigated the study of a recursive Bayesian filter for tracking velocity fields of fluid flows. The filter combines an Ito diffusion process associated to 2D vorticity-velocity formulation of Navier-Stokes equation and discrete image error reconstruction measurements. In contrast to usual filters designed for visual tracking problems, our filter combines a continuous law for the description of the vorticity evolution with discrete image measurements. We resort to a Monte-Carlo approximation based on particle filtering. The designed tracker provides a robust and consistent estimation of instantaneous motion fields along the whole image sequence.

When the likelihood of the measurement can be modeled as Gaussian law, we have also investigated the use of the so-called ensemble Kalman filtering for fluid tracking problems. This kind of filters introduced for the analysis of geophysical fluids is based on the Kalman filter update equation. Nevertheless, unlike traditional Kalman filtering setting, the covariances of the estimation errors, required to compute the so-called Kalman gain, relies on an ensemble of forecasts. Such a process gives rise to a Monte Carlo approximation for a family of non-linear stochastic filters enabling to handle state spaces of large dimension. We have recently proposed an extension of this technique that combines sequential importance sampling and the propagation law of an ensemble Kalman filter. This technique leads to an ensemble Kalman filter with an improved efficiency. This year we have investigated the introduction of a nonlinear direct image measurements operator within this ensemble Kalman scheme. This modification of the filter provides very good results on 2D numerical and experimental flows even in the presence of strong noises. We are currently assessing its application to oceanic satellite images for the recovering of ocean streams. We are studying also the impact on the stochastic dynamics of turbulent noise defined as auto-similar Gaussian random fields and the introduction within an incremental ensemble analysis scheme of multiscale motion measurements. This work has been published in conference issues [28], [27], [35].

6.2.2. Reduced-order models for flows representation from image data

Participants: Patrick Héas, Cédric Herzet, Etienne Mémin, Véronique Souchaud.

One of the possibilities to neglect the influence of some degrees of freedom over the main characteristics of a flow consists in representing it as a sum of $K$-orthonormal spatial basis functions weighted with temporal coefficients. To determine the basis function of this expansion one of the usual approaches relies on the Karhunen-Loeve decomposition (referred as proper orthogonal decomposition – POD – in the fluid mechanics domain). In practice, the spatial basis functions, also called modes, are the eigen vectors of an empirical auto-correlation matrix which is built from “snapshots” of the considered physical process.

In this axis of work we focus on the case where one does not have a direct access to snapshots of the considered physical process. Instead, the POD has to be built from the partial and noisy observation of the physical process. Instances of such scenarios include situations where real instantaneous vector-field snapshots are estimated from a sequence of images. We have been working on several approaches dealing with such a new paradigm. A first approach consists in extending standard penalized motion-estimation algorithms to the case where the sought velocity field is constrained to span a low-dimensional subspace [38]. Giving a probabilistic interpretation to this problem, we have designed novel optimization procedures in the framework of maximum a posteriori estimation problem. This work has led to the publication of a paper in the Gretsi conference. We are currently working on an EM-algorithm implementation of this approach.

In a second approach, we are considering the design of the POD as the solution of a minimum least squares estimation problem based on the distribution of the (unknown) velocity field given a sequence of images. This alternative formulation allowed us to take explicitly the uncertainty on the velocity field into account into our optimization process. We are currently working on several practical implementations of this problem, relying on Monte-Carlo integration and Krylov subspaces.
In a third axis we have studied two variational data assimilation techniques for the estimation of low order dynamical models for fluid flows. Both methods are built from optimal control recipes and rely on POD representation associated to Galerkin projection of the Navier Stokes equations. The proposed techniques differ in the control variables they involve. The first one introduces a weak dynamical model defined only up to an additional uncertainty time dependent function whereas the second one, handles a strong dynamical constraint in which the coefficients of the dynamical system constitute the control variables. Both choices correspond to different approximations of the relation between the reduced basis on which is expressed the motion field and the basis components that have been neglected in the reduced order model construction. The techniques have been assessed on numerical data and for real experimental conditions with noisy Image Velocimetry data. This work has been presented in several conferences. A journal paper has been recently accepted with minor changes to the journal of computational Physics.

6.2.3. Optimal control techniques for the coupling of large eddy dynamical systems and image data

Participants: Dominique Heitz, Etienne Mémin, Cordelia Robinson, Yin Yang.

This work aims at investigating the use of optimal control techniques for the coupling of Large Eddies Simulation (LES) techniques and 2D image data. The objective is to reconstruct a 3D flow from a set of simultaneous time resolved 2D image sequences visualizing the flow on a set of 2D plans enlightened with laser sheets. This approach will be experimented on shear layer flows and on wake flows generated on the wind tunnel of Irstea Rennes. Within this study we wish also to explore techniques to enrich large-scale dynamical models by the introduction of uncertainty terms or through the definition of subgrid models from the image data. This research theme is related to the issue of turbulence characterization from image sequences. Instead of predefined turbulence models, we aim here at tuning from the data the value of coefficients involved in traditional LES subgrid models or in longer-term goal to learn empirical subgrid models directly from image data. An accurate modeling of this term is essential for Large Eddies Simulation as it models all the non resolved motion scales and their interactions with the large scales.

First tests have been conducted with two-dimensional Direct Numerical Simulations (DNS) of mixing layer coupled with noisy observations. By modifying the initial condition of the system, the proposed method recovers the state of an unknown function with good accuracy. This work has been published in the International Symposium on Turbulence and Shear Flow Phenomena (TSFP) 2011 [ 29 ].

6.2.4. Free surface flows reconstruction and tracking

Participants: Benoît Combes, Dominique Heitz, Etienne Mémin.

Characterising a free-surface flow (space and time-dependent velocity and geometry) given observations/measures at successive times is an ubiquitous problem in fluid mechanic and in hydrology. Observations can consist of e.g. measurements of velocity, or like in this work of measurements of the geometry of the free-surface. Indeed, recently developed depth/range sensors allow to capture directly a rough 3D geometry of surfaces with high space and time resolution. The main purpose of this study is to evaluate the ability of the Kinect sensor to estimate time-dependent 3D free-surface geometries. Then, based on these observations and on a stochastic data assimilation method, we want to estimate both time dependent geometry and displacement field associated to a free-surface flow from a simple temporal sequence of Kinect data. This year we have demonstrated on real data the possibility to measure free surface flow geometry with kinect sensor and on synthetic data to estimate both time dependent geometry and displacement field associated to a free-surface flow from a simple temporal sequence of Kinect-like data. This work has been published in the new conference Flow Volume Reconstruction [ 22 ]. We intend to extend such a study to hydrological applications.

6.2.5. Stochastic filtering technique for the tracking of closed curves

Participants: Christophe Avenel, Etienne Mémin.
We have proposed a filtering methodology for the visual tracking of closed curves. Opposite to works of the literature related to this issue, we consider here a curve dynamical model based on a continuous time evolution law with different noise models. This led us to define three different stochastic differential equations that capture the uncertainty relative to curve motions. This new approach provides a natural understanding of classical level-set dynamics in terms of such uncertainties. These evolution laws have been combined with various color and motion measurements to define probabilistic state space models whose associated Bayesian filters can be handled with particle filters. This on going work will be continued within extensive curve tracking experiments and extended to the tracking of other very high dimensional entities such as vector fields and surfaces. This work has been published in conference proceedings and a journal article is conditionally accepted to minor changes for publication in a meteorological journal.

6.2.6. Sequential smoothing for fluid motion
Participants: Anne Cuzol, Etienne Mémin.

In parallel to the construction of stochastic filtering techniques for fluid motions, we have proposed a new sequential smoothing method within a Monte-Carlo framework. This smoothing aims at reducing the temporal discontinuities induced by the sequential assimilation of discrete time data into continuous time dynamical models. The time step between observations can indeed be long in environmental applications for instance, and much longer than the time step used to discretize the model equations. While the filtering aims at estimating the state of the system at observations times in an optimal way, the objective of the smoothing is to improve the estimation of the hidden state between observation times. The method is based on a Monte-Carlo approximation of the filtering and smoothing distributions, and relies on a simulation technique of conditioned diffusions. The proposed smoother can be applied to general non linear and multidimensional models. It has been applied to a turbulent flow in a high-dimensional context, in order to smooth the filtering results obtained from a particle filter with a proposal density built from an Ensemble Kalman procedure.

6.2.7. Stochastic fluid flows dynamics under Gaussian uncertainty
Participant: Etienne Mémin.

In this research axis we aim at devising stochastic Eulerian expression for the description of fluid flow evolution laws incorporating uncertainty on the particles location. Such an uncertainty modeled through the introduction of a random term allows taking into account approximations or truncation effects performed within the dynamics analytical constitution steps. This includes for instance the modeling of unresolved scales interaction in large eddies simulation (LES) or in Reynolds average numerical simulation (RANS), but also uncertainties attached to non uniform grid discretization. This model is mainly based on a stochastic version of the Reynolds transport theorem. Within this framework various simple expressions of the mean drift component can be exhibited for different models of the random field carrying the uncertainties we have on the flow. We aim at using such a formalization within image based data assimilation framework and to derive appropriate stochastic versions of geophysical flow dynamical modeling.

6.2.8. Variational assimilation of images for large scale fluid flow dynamics with uncertainty
Participants: Souleymane Kadri Harouna, Etienne Mémin.

In this work we explore the assimilation of a large scale representation of the flow dynamics with image data provided at a finer resolution. The velocity fields at large scales is described as a regular smooth components whereas the complement component is a highly oscillating random velocity field defined on the image grid but living at all the scales. Following this route we have started to assess the performances of a variational assimilation technique with direct image data observation. Preliminary results obtained for a wavelet based 2D Navier Stokes implementation and images of a passive scalar transported by the flow are very encouraging.

6.3. Analysis and modeling of turbulent flows

6.3.1. Mixing layers between a uniform flow and a shear flow
Participant: Dominique Heitz.
We have addressed the analysis and modelling of non canonical turbulent mixing layers between a uniform flow and a shear flow. From a parametric study by bidimensional direct numerical simulations two mixing layer configurations between a uniform flow and a shear flow have been selected. These two configurations share the same shear flow but have a different uniform flow.

The shear flow was obtained with curved gauze. However the theoretical shear parameter predicted by the literature is different from the value obtained by experiments. In order to study these discrepancies, the flow through a gauze was studied by particle image velocimetry. This allowed the general modeling of the uniform flow through curved wire gauze, leading to linear mean velocity profiles. From a hot-wire anemometry study of the two flow configurations it was observed that one flow behaves like a mixing layer whereas the other flow yields a wake behaviour. The mixing layer indicates an increasing turbulent kinetic energy along its longitudinal development, while the wake exhibits an asymmetry.

6.3.2. Hot-wire anemometry at low velocities

**Participant:** Dominique Heitz.

A new dynamical calibration technique has been developed for hot-wire probes. The technique permits, in a short time range, the combined calibration of velocity, temperature and direction calibration of single and multiple hot-wire probes. The calibration and measurements uncertainties were modeled, simulated and controlled, in order to reduce their estimated values. This year a patent application has been submitted.

6.3.3. Experimental studies for the assessment of turbulence statistical models

**Participants:** Patrick Héas, Dominique Heitz, Etienne Mémin.

[In collaboration with G. Artana and P. Minini (Univ. Bueno Aires)]

Selecting directly from images the most likely scaling motion priors enables the recovery of physical quantities related to the energy flux and the flow regularity. Such measurements are of major interest for turbulence studies. In particular, determining the energy flux across scales and characterizing intermittency is very important to assess the relevance of the statistical models proposed for atmospheric turbulence. Although, the measurement of flux and atmospheric flow regularity has already been obtained previously using in situ data, it required an important measurement campaign lasting several years based on sensors placed on airplanes. Therefore, the proposed motion estimation technique described above represents an attractive tool since it enables the direct estimation of these quantities from a couple of images. A paper concerning an atmospheric turbulence study using Meteosat Second Generation (MSG) images has been accepted in the journal Tellus A. Experimental studies of three-dimensional turbulence behind a grid or in the wake of a cylinder has been performed and will be submitted in the journal Experiments in fluids. New experiments for the assessment of turbulence statistical models are currently going on in collaboration with the laboratory of fluid mechanics and turbulence scientists in Argentina. They focus on two-dimensional turbulence of soap films visualized with a Schlieren imagery system. The goal of this work is to validate experimentally the theoretical model predicting non-intermittent inverse energy cascades in pure two-dimensional flows.

6.4. Visual servoing approach for fluid flow control

6.4.1. Fully exploitation of the controlled degrees of freedom of the 2D plane Poiseuille flow

**Participants:** Christophe Collewet, Xuan Quy Dao.

This works concerns the Phd of Xuan-Quy Dao and can be seen as an extension of the works carried out by Roméo Tatsambon. Since visual measurements are used, we propose here to use advanced visual servoing techniques to fully exploit the controlled degrees of freedom of the 2D plane Poiseuille flow. To achieve this goal we propose to design a control law based on partitioned visual servo control (this approach has been first proposed in the robotics community by [46] but in a very different context). Therefore, we have shown that, following this way and contrary to the literature concerning drag reduction, it becomes easy to simultaneously reduce the drag and the kinetic energy density of the flow. That is of great importance since the controlled flow is in an unstable state and may become turbulent when the kinetic energy density is growing. This key
problem is not well taken into account in the literature. Indeed, either the drag or the kinetic energy density is reduced, but never both of them. Moreover, we have shown that in practice, the way the drag is reduced does not influence the way the kinetic energy density is reduced. In addition, since dense visual measurements are used, our approach is very robust against measurement noise.

6.4.2. Visual servoing for the 3D plane Poiseuille flow

Participant: Christophe Collewet.

We focus here on the 3D plane Poiseuille flow which is much more realistic than the 2D case. In that case, it can be shown that the reduced linearized flow is in a stable configuration. However, it is possible to find some bad initial conditions which causes the flow to present high transient energy growths. Indeed, a small perturbation velocity value in the reduced linearized system leads to a transient effect which is characterized by a growth in a short-time behaviour of the kinetic density energy, before a decay occurs. Practically, this transient effect, if not controlled, can cause transition to turbulence. Usually, the streamwise shear stress component at a point belonging to the wall is used as the output of the system in order to control it in a closed-loop fashion. We have proposed a vision-based approach to control this flow (see section 3.4). Our approach has revealed to be the most efficient approach in comparison to existing ones. Indeed, the transient energy is highly reduced. In addition, as in the 2D plane Poiseuille flow, the initialization problem is not of concern in the vision-based approach. In addition, our approach is robust to measurement noise when a large number of flow measurements is available, which is possible in real practical situations. This work has been published in several conferences [33], [20], [32] and has been recently accepted for publication in the International Journal of Flow Control [19].

6.5. National Initiatives

6.5.1. ANR-COSINUS PREVASSEMBLE: Ensemble methods for assimilation of observations and for prevision in Meteorology and Oceanography

Participants: Sébastien Béyou, Anne Cuzol, Etienne Mémin.

duration 36 months.

The purpose of this project is to further study ensemble methods and to develop their use for both assimilation of observations and prediction. Among the specific questions to be studied are the theory of Particle Filters and Ensemble Kalman Filters, the possibility of taking temporal correlation into account in ensemble assimilation, the precise assessment of what can and cannot be achieved in ensemble prediction, and the objective validation of ensemble methods.

The partners of this project are Laboratoire de Météorologie Dynamique/ENS (leader), Météo-France and three INRIA groups (ALEA, ASPI, FLUMINANCE).

6.5.2. ANR SYSCOMM MSDAG: MultiScale Data Assimilation in Geophysics

Participants: Pierre Dérian, Patrick Héas, Dominique Heitz, Cédric Herzet, Etienne Mémin.

duration 36 months.

Changing scale is a well-known topic in physics (geophysics, fluid mechanics and turbulence, theoretical and statistical physics, mechanics, porous media, etc.) It has lead to the creation of powerful sophisticated mathematical tools: renormalization, homogenization, etc. These ideas are also used in numerical analysis (the so-called multigrid approach) for solving efficiently partial differential equations. Data assimilation in geophysics is a set of methods that allows to combine optimally numerical models in large spaces with large dataset of observations. At the confluence of these two topics, the goal of this project is to study how to embed the change of scales (a multiscale point of view) issue into the framework of geophysical data assimilation, which is a largely unexplored subject.

The partners of this 3 years project are the CEREA/CLIME INRIA group (leader), the LSCE/CEA, the INRIA groups MOISE and FLUMINANCE.
6.5.3. ANR SYSCOMM GeoFluids:
Participants: Patrick Héas, Dominique Heitz, Etienne Mémin, Véronique Souchaud.

duration 48 months.

The project Geo-FLUIDS focuses on the specification of tools to analyse geophysical fluid flows from image sequences. Geo-FLUIDS aims at providing image-based methods using physically consistent models to extract meaningful features describing the observed flow and to unveil the dynamical properties of this flow. The main targeted application domains concern Oceanography and Meteorology. The project consortium gathers the INRIA research groups: FLUMINANCE (leader), CLIME, IPSO, and MOISE. The group of the “Laboratoire de Météorologie Dynamique” located at the ENS Paris, the IFREMER-CERSAT group located at Brest and the METEOFRANCE GMAP group in Toulouse.

6.5.4. Brittany concil ARED IMAGEO:
Participants: Cédric Herzet, Etienne Mémin, Véronique Souchaud.

duration 36 months. This project of the Brittany concil, which finances the PhD thesis of Véronique Souchaud, aims at studying methods for the estimation of reduced order modeling of fluid flows evolution laws from image sequences. The goal consists here at defining the estimation of a reduced basis describing the flow evolution as a motion estimation problem.

6.6. International Initiatives

6.6.1. INRIA Associate Teams

6.6.1.1. HURACAN

Title: Analysis and control of fluid flows from image sequences
INRIA principal investigator: Etienne Memin
International Partner:
Institution: Universidad de Buenos Aires (Argentina)
Laboratory: Fluid Dynamics Laboratory

International Partner:
Institution: IRSTEA (France)
Duration: 2010 - 2012
See also: http://huracan.inria.fr

The HURACAN associated team is centered on the analysis and the control of fluid flows from image sequences. The research objectives of this team are organized into two distinct work axes. The first one aims at defining and studying visual servoing techniques for fluid flows control. In addition to the definition of efficient visual servoing schemes this axis of work gathers research issues related to fluid flows velocity measurement from images and to flows excitation through plasma actuators. The second research axis focuses on the coupling between large scales representations of geo-physical flows and image data. More precisely, it aims at studying means to define directly from the image sequences the small scales terms of the dynamics. This research axis includes the study of coupling models and data defined at different scales, problems of multiscale velocities estimation respecting turbulence phenomenological laws and issues of experimental validation.
6. New Results

6.1. Parallelism and convergence in Krylov methods

Participants: Édouard Canot, Jocelyne Erhel, Désiré Nuentsa Wakam, Bernard Philippe.

This work is done in the context of the Cinemas2 and the Libraero contracts, 7.2 and 8.1.3. It is also done in collaboration with the joint INRIA/ NCSA laboratory on petascale computing.

A Ph-D thesis was defended this year [12].

6.1.1. Some properties of Krylov methods

Participant: Jocelyne Erhel.

A survey was presented at a conference and published in a book chapter [37] [24].

Solving a linear system is at the heart of many scientific and engineering applications. Generally, this operation is the most time and memory consuming part of the simulation. This paper focuses on some properties of Krylov iterative methods. Iterative methods of Krylov type require less memory than direct methods, but the number of iterations increases rapidly with the size of the system. The convergence rate and the accuracy of the results depend on the condition number which can blow up at large scale. Therefore, it is essential to combine these methods with a preconditioner; the idea is to solve another system, close to the original one, but which is easier to solve; also, on parallel computers, it must be scalable. In Krylov iterative methods, the matrix is not transformed but the kernel operation is the matrix-vector product; thus it is possible to use matrix-free versions without storing the matrix. However, preconditioning will sometimes require the matrix. Krylov methods are described in many books. In this survey, we choose the framework of polynomial and projection methods. We first give general properties. Then, we study specific methods for the three different types of matrices: the case of SPD matrices is analyzed first, followed by the case of symmetric indefinite matrices. The general case of nonsymmetric matrices is studied with the description of several Krylov methods. Finally, some practical issues, preconditioning and parallelism are discussed.

6.1.2. Generation of Krylov subspace bases

Participant: Bernard Philippe.

This work was done in collaboration with L. Reichel, from University of Kent, USA.

It has been published in a journal [19].

Many problems in scientific computing involving a large sparse square matrix A are solved by Krylov subspace methods. This includes methods for the solution of large linear systems of equations with A, for the computation of a few eigenvalues and associated eigenvectors of A, and for the approximation of nonlinear matrix functions of A. When the matrix A is non-Hermitian, the Arnoldi process commonly is used to compute an orthonormal basis for a Krylov subspace associated with A. The Arnoldi process often is implemented with the aid of the modified Gram–Schmidt method. It is well known that the latter constitutes a bottleneck in parallel computing environments, and to some extent also on sequential computers. Several approaches to circumvent orthogonalization by the modified Gram–Schmidt method have been described in the literature, including the generation of Krylov subspace bases with the aid of suitably chosen Chebyshev or Newton polynomials. We review these schemes and describe new ones. Numerical examples are presented.

6.1.3. Parallel preconditioned GMRES with Multiplicative Schwarz

Participants: Édouard Canot, Jocelyne Erhel, Désiré Nuentsa Wakam, Bernard Philippe.

This work was published in a journal [18].
This paper presents a robust hybrid solver for linear systems that combines a Krylov subspace method as accelerator with a Schwarz-based preconditioner. This preconditioner uses an explicit formulation associated to one iteration of the multiplicative Schwarz method. The Newton-basis GMRES, which aim at expressing a good data parallelism between subdomains is used as accelerator. In the first part of this paper, we present the pipeline parallelism that is obtained when the multiplicative Schwarz preconditioner is used to build the Krylov basis for the GMRES method. This is referred as the first level of parallelism. In the second part, we introduce a second level of parallelism inside the subdomains. For Schwarz-based preconditioners, the number of subdomains is kept small to provide a robust solver. Therefore, the linear systems associated to subdomains are solved efficiently with this approach. Numerical experiments are performed on several problems to demonstrate the benefits of using these two levels of parallelism in the solver, mainly in terms of numerical robustness and global efficiency.

6.1.4. Adaptive deflation in preconditioned GMRES algorithm using a combined preconditioning

Participants: Jocelyne Erhel, Désiré Nuentsa Wakam, Bernard Philippe.

This work has been presented at a conference and a workshop [35], [27] and submitted to the proceedings of DD20 [45]. The software module DGMRES is integrated in the Petsc distribution.

Many scientific libraries are currently based on the GMRES method as a Krylov subspace iterative method for solving large linear systems. The restarted formulation known as GMRES\( (m) \) has been extensively studied and several approaches have been proposed to reduce the negative effects due to the restarting procedure. A common effect in GMRES\( (m) \) is a slow convergence rate or a stagnation in the iterative process. In this situation, it is less attractive as a general solver in industrial applications. In this work, we propose an adaptive deflation strategy which retains useful information at the time of restart to avoid stagnation in GMRES\( (m) \) and improve its convergence rate. We give a parallel implementation in the PETSc package. The provided numerical results show that this approach can be effectively used in the hybrid direct/iterative methods to solve large-scale systems.

6.1.5. Adaptive deflation in preconditioned GMRES algorithm using an augmented subspace

Participants: Jocelyne Erhel, Désiré Nuentsa Wakam, Bernard Philippe.

This work has been presented at a conference [31] and submitted to the journal ETNA [46].

The GMRES iterative method is widely used as Krylov subspace accelerator for solving sparse linear systems when the coefficient matrix is nonsymmetric and indefinite. The Newton basis implementation has been proposed on distributed memory computers as an alternative to the classical approach with the Arnoldi process. The aim of our work here is to introduce a modification based on deflation techniques. This approach builds an augmented subspace in an adaptive way to accelerate the convergence of the restarted formulation. In our numerical experiments, we show the benefits of using this implementation with hybrid direct/iterative methods to solve large linear systems.

6.1.6. Using deflated preconditioned GMRES for industrial CFD problems

Participant: Désiré Nuentsa Wakam.

This work has been submitted to the journal Computers and Fluids [47].

This paper deals with the solution of large and sparse linear systems arising from design optimization in Computational Fluid Dynamics. From the algebraic decomposition of the input matrix, a hybrid robust direct/iterative solver is often defined with a Krylov subspace method as accelerator, a domain decomposition method as preconditioner and a direct method as subdomain solver. The goal of this paper is to reduce the memory requirements and indirectly the computational cost at different steps of this scheme. To this end, we use a grid-point induced block approach for the data storage and the partitioning part, a Krylov subspace method based on the restarted GMRES accelerated by deflation, a preconditioner formulated with the restricted additive Schwarz method and an aerodynamic/turbulent fields split at the subdomain level. Numerical results are presented with industrial test cases to show the benefits of these choices.
6.2. Parallel numerical algorithms

6.2.1. High Performance Scientific Computing

Participant: Bernard Philippe.

This work was done in collaboration with several authors, from US, Greece, etc. A book will appear on this subject in 2012 [39] and a chapter of this book is devoted to a historical perspective [38].

This comprehensive text/reference, inspired by the visionary work of Prof. Ahmed H. Sameh, represents the state of the art in parallel numerical algorithms, applications, architectures, and system software. Articles in this collection address solutions to various challenges arising from concurrency, scale, energy efficiency, and programmability. These solutions are discussed in the context of diverse applications, ranging from scientific simulations to large-scale data analysis and mining.

As exascale computing is looming on the horizon while multicore and GPU’s are routinely used, we survey the achievements of Ahmed H. Sameh, a pioneer in parallel matrix algorithms [38]. Studying his contributions since the days of Illiac IV as well as the work that he directed and inspired in the building of the Cedar multiprocessor and his recent research, unfolds a useful historical perspective in the field of parallel scientific computing.

6.2.2. Updating the Diagonalization of a Symmetric Matrix

Participant: Bernard Philippe.

This work is done in the context of the DIAMS project.

Two methods are compared: Jacobi method and first order correction of the spectral projectors [25],[26].

6.2.3. Counting eigenvalues in domains of the complex field

Participant: Bernard Philippe.

This work is done in collaboration with E. Kamgnia, from the University of Yaounde 1, Cameroon, in the context of the MOMAPLI project at LIRIMA.

It has been submitted to a journal [43].

A procedure for counting the number of eigenvalues of a matrix in a region surrounded by a closed curve is presented. It is based on the application of the residual theorem. The quadrature is performed by evaluating the principal argument of the logarithm of a function. A strategy is proposed for selecting a path length that insures that the same branch of the logarithm is followed during the integration. Numerical tests are reported for matrices obtained from conventional matrix test sets.

6.2.4. Rescaling for time integration

Participant: Jocelyne Erhel.

This work is done in collaboration with N. Makhoul and N. Nassif, from the American University of Beirut, Lebanon.

It is published in a journal [17].
This paper considers the mathematical framework of a sliced-time computation method for explosive solutions to systems of ordinary differential equations: \( Y(t) \in \mathbb{R}^k \): \( \frac{dY}{dt} = F(Y), \) \( 0 < t, Y(0) = Y_0, \) that have finite or infinite explosion time. The method used generates automatically a sequence of non uniform slices \( \{[T_{n-1}, T_n]|n \geq 1\} \) determined by an end-of-slice condition that controls the growth of the solution within each slice. It also uses rescaling of the variables, whereas: \( t = T_{n-1} + \beta_n s \) and \( Y(t) = Y(T_{n-1}) + D_n Z_n(s), \) \( D_n \mathbb{R}^{k \times k} \) and \( \beta_n \) being respectively an invertible diagonal matrix and a rescaling time factor. Thus, the original system is transformed into a sequence of slices-dependent initial-value shooting problems: \( \frac{dZ_n}{ds} = G_n(Z_n), \) \( 0 < s \leq s_n, \) \( Z_n(0) = 0. \) A suitable selection of \( \beta_n \) and \( D_n \) leads the rescaled systems to verify a concept of uniform similarity, allowing to disable the extreme stiffness of the original ODE problem.

Then, on each time slice, the uniformly rescaled systems are locally solved using a 4th order explicit Runge-Kutta scheme, within a computational tolerance of \( \epsilon_{\text{loc}}. \) After sequentially implementing the local solver on a total of \( N \) slices, a global tolerance \( \epsilon_{\text{glob}} \) would result in approximating the solution \( Y(t) \) of the original system.

The proper definition of Uniform Similarity leads to deriving, under a stability assumption, a relationship between \( \epsilon_{\text{loc}}, \epsilon_{\text{glob}} \) and \( N. \) Numerical experiments are conducted for infinite and finite times explosive discrete reaction diffusion problems. These experiments validate the theoretical results and attest for the efficiency of the method in terms of stability and high accuracy.

6.3. Numerical models and simulations applied to physics

6.3.1. Heat and mass transfer in soil and prehistoric fires

Participant: Édouard Canot.

This work is done in the context of the Arphymat project, in collaboration with Archeosciences, IPR and Lebanese International University (LIU), Lebanon.

This work is published in a journal [16].

This paper is devoted to the simulation of water forced evaporation in a porous saturated medium in a 3D-axisymmetric domain by resolution of partial differential algebraic equations (PDAE) that are encountered in different engineering applications. The goal of this paper is an attempt to present effective realizations, in order to determine the minimal duration of burning for prehistoric occupations. This multidisciplinary work includes scientists in Mathematics, Physics and Archaeology. The model proposed here couples the heat conduction in a water saturated soil with the water steam flow in the medium. We propose an efficient and robust global numerical method, based on a method of lines and differential algebraic equations (DAE) solvers, combined with a Newton method using a powerful sparse linear solver. After a brief overview of classes for numerical techniques applied for moving boundary problems, the Apparent Heat Capacity method (AHC) is used, and in order to validate our codes, a comparison with experiments is done.

Recent work concerns the optimal choice of the temperature interval across which the phase change occurs in the apparent capacity method, because we have to make a compromise between the smoothness of the solution and its accuracy.

6.3.2. Rheology of granular systems flowing out of silo

Participant: Édouard Canot.

This work is done in the framework of a project funded by the Region Bretagne. A PhD thesis (Merline Djouwe-Tankeo), coadvised with Patrick Richard, who is from the Physics Institute at the University of Rennes (IPR), started in February 2009 and will be defended in January 2012.

It has been presented at a conference [36] and a paper is submitted.
We first studied the granular flows by the "discrete elements" method in silo geometries. By changing the micro-mechanical properties of the grains (restitution and friction), we showed that they had a significant influence on the flow discharge. Although models such as "discrete elements" provide access to all the individual properties of the grains, they have one major drawback: the computation time is very important that prohibits the modeling of geophysical and industrial situations. To overcome this problem, we used the "continuous medium" approach, which consider that the granular medium studied follows a rheology recently proposed in the literature. After discussing the numerical implementation, we have studied this rheology for steady and fully developed flows with a semi-analytical method in two configurations: a shear cell and a channel. This allowed us to highlight the differences between a granular medium and a Newtonian fluid.

6.4. Models and simulations for transport in porous media

6.4.1. Transport in highly heterogeneous porous medium

Participants: Jocelyne Erhel, Géraldine Pichot, Nadir Soualem.

This work is done in collaboration with A. Beaudoin, from University of Poitiers (Pprime) and J.-R. de Dreuzy, from Geosciences Rennes. It is done in the context of the Micas project (8.1.2).

It has been presented at a conference and a paper is in preparation [28].

We study the transport of an inert species in a 2D heterogeneous porous medium via a Random Walk Particle Tracking (RWPT) method. The main objective is to derive the macroscopic properties of the transport by the means of Monte-Carlo simulations in large domains. Conditions to reach asymptotic macro-dispersion coefficients are given. We also present our on-going research about the RWPT method in presence of discontinuities within the domain.

6.4.2. Transport in discontinuous porous medium

Participants: Jocelyne Erhel, Géraldine Pichot.

This work is done in collaboration with A. Lejay, from Inria Nancy. It is done in the context of the Micas project (8.1.2).

It is published in the proceedings of a conference and submitted in a journal [30], [44].

We study a diffusion process in a 1D discontinuous medium using a random walk approach. Our main contribution is to encompass two existing numerical methods in the unified framework of the Skew Brownian Motion. This theoretical approach allows to detail and justify the derived algorithms. Numerical simulations are performed on two test cases to show that the algorithms can deal with the discontinuity in the diffusion coefficient.

6.4.3. Reactive transport

Participants: Édouard Canot, Jocelyne Erhel, Souhila Sabit, Nadir Soualem.

This work is done in the context of the MOMAS GNR (8.1.1) and the contract with Andra (7.1).

It has been presented at a workshop and a paper is in preparation [33]. The software GRT3D (see section 5.6) is described in a report [48].

We have developed a method coupling transport and chemistry, based on a method of lines such that spatial discretization leads to a semi-discrete system of algebraic differential equations (DAE system). The main advantage is to use a complex DAE solver, which controls simultaneously the timestep and the convergence of Newton nonlinear iterations [53]. Analysis done with several numerical experiments showed that most of CPU time is spent in solving the linear systems of Newton iterations. We have reduced this computational time by reducing the size of the system; numerical experiments with large 2D domains show the efficiency.
6.5. Models and simulations for flow in fractured media

This work is done in collaboration with J.-R. de Dreuzy, from the department of Geosciences at the University of Rennes 1 (who is on leave until 2013 at UPC, Barcelona, Spain). It is done in the context of the Micas project (8.1.2).

A Ph-D thesis was defended this year [13].

6.5.1. Domain decomposition method for flow in 3D networks of fractures

Participants: Jocelyne Erhel, Baptiste Poirriez.

This work was presented at a conference and published in the proceedings of another conference [29], [32]. A paper is in preparation.

This paper aims at solving efficiently the linear system arising from flow computations in Discrete Fracture Networks (DFN). We define a partition of fractures into connected sets and apply a Schur domain decomposition method. Conjugate Gradient is preconditioned by Neumann-Neumann and deflation. Preliminary results with one network show the ability of our method to reduce both the number of iterations and the computational time.

6.5.2. Mortar method for flow in 3D networks of fractures

Participants: Jocelyne Erhel, Géraldine Pichot.

This work is published in a journal [20].

The simulation of flow in fractured media requires handling both a large number of fractures and a complex interconnecting network of these fractures. Networks considered in this paper are 3D domains made up of 2D fractures intersecting each other and randomly generated. Due to the stochastic generation of fractures, intersections can be highly intricate. The numerical method must generate a mesh and define a discrete problem for any Discrete Fracture Network (DFN). A first approach [51] is to generate a conforming mesh and to apply a mixed hybrid finite element method. However the resulting linear system becomes very large when the network contains many fractures. Hence a second approach [52] is to generate a non conforming mesh, using an independent mesh generation for each fracture. Then a Mortar technique applied to the mixed hybrid finite element method deals with the non-matching grids. When intersections do not cross nor overlap, pairwise Mortar relations for each intersection are efficient [52]. But for most of random networks, discretized intersections involve more than two fractures. In this paper, we design a new method generalizing the previous one and applicable for stochastic networks. The main idea is to combine pairwise Mortar relations with additional relations for the overlapping part. This method still ensures the continuity of fluxes and heads and still yields a symmetric positive definite linear system. Numerical experiments show the efficiency of the method applied to complex stochastic fracture networks. We also study numerical convergence when reducing the mesh step. This method makes it easy to perform mesh optimization and appears as a very promising tool to simulate flow in multiscale fracture networks.

6.6. Uncertainty quantification in hydrogeology

This work is done in collaboration with A. Debussche, from ENS-Cachan-Rennes and Ipso INRIA team. It is done in the context of the Micas project (8.1.2).

A PhD thesis was defended this year [11].

6.6.1. Strong and weak error estimates for elliptic partial differential equations with random coefficients

Participant: Julia Charrier.

This work has been presented at a workshop and is published in a journal [23], [14].
We consider the problem of numerically approximating the solution of an elliptic partial differential equation with random coefficients and homogeneous Dirichlet boundary conditions. We focus on the case of a lognormal coefficient, we have then to deal with the lack of uniform coercitivity and uniform boundedness with respect to the randomness. This model is frequently used in hydrogeology. We approximate this coefficient by a finite dimensional noise using a truncated Karhunen-Loève expansion. We give then estimates of the corresponding error on the solution, both a strong error estimate and a weak error estimate, that is to say an estimate of the error committed on the law of the solution. We obtain a weak rate of convergence which is twice the strong one. Besides this, we give a complete error estimate for the stochastic collocation method in this case, where neither coercitivity nor boundedness are stochastically uniform. To conclude, we apply these results of strong and weak convergence to two classical cases of covariance kernel choices: the case of an exponential covariance kernel on a box and the case of an analytic covariance kernel, yielding explicit weak and strong convergence rates.

6.6.2. Numerical analysis of a multilevel Monte Carlo method for elliptic PDEs with random coefficients

Participant: Julia Charrier.

This work has been presented at a conference and is submitted in a journal [42], [22].

We consider a finite element approximation of elliptic partial differential equations with random coefficients. Such equations arise, for example, in uncertainty quantification in subsurface flow modelling. Models for random coefficients frequently used in these applications, such as log-normal random fields with exponential covariance, have only very limited spatial regularity, and lead to variational problems that lack uniform coercitivity and boundedness with respect to the random parameter. In our analysis we overcome these challenges by a careful treatment of the model problem almost surely in the random parameter, which then enables us to prove uniform bounds on the finite element error in standard Bochner spaces. These new bounds can then be used to perform a rigorous analysis of the multilevel Monte Carlo method for these elliptic problems that lack full regularity and uniform coercitivity and boundedness. To conclude, we give some numerical results that confirm the new bounds.

6.6.3. Numerical analysis of the advection-diffusion of a solute in random media

Participant: Julia Charrier.

This work is submitted in a journal [41].

We consider the problem of numerically approximating the solution of the coupling of the flow equation in a random porous medium, with the advection-diffusion equation. More precisely, we present and analyse a numerical method to compute the mean value of the spread of a solute introduced at the initial time, and the mean value of the macro-dispersion, defined at the temporal derivative of the spread. We propose a Monte-Carlo method to deal with the uncertainty, i.e. with the randomness of the permeability field. The flow equation is solved using finite element. The advection-diffusion equation is seen as a Fokker-Planck equation, and its solution is approximated thanks to a probabilistic particular method. The spread is indeed the expected value of a function of the solution of the corresponding stochastic differential equation, and is computed using an Euler scheme for the stochastic differential equation and a Monte-Carlo method. Error estimates on the mean spread and on the mean dispersion are established, under various assumptions, in particular on the permeability random field.

6.6.4. Model reduction for a 1D stochastic elliptic PDE

Participants: Jocelyne Erhel, Mestapha Oumouni.

This work is done in collaboration with Z. Mghazli, from the university of Kenitra, Morocco, in the context of the Co-Advise and Hydromed projects (8.2.1, 8.3.4).

This work has been presented at a conference and published in a journal [15] [34].
In this paper, we present an efficient method to approximate the expectation of the response of a one-dimensional elliptic problem with stochastic inputs. In conventional methods, the computational effort and cost of the approximation of the response can be dramatic. Our method presented here is based on the Karhunen–Loève (K-L) expansion of the inverse of the diffusion parameter, allowing us to build a base of random variables in reduced numbers, from which we construct a projected solution. We show that the expectation of this projected solution is a good approximation, and give an a priori error estimate. A numerical example is presented to show the efficiency of this approach.

6.6.5. Inverse problems in hydrogeology

Participant: Sinda Khalfallah.

This work is done in collaboration with A. ben Abda, from LAMSIN, Tunisia, in the context of the Hydromed and Co-Advise projects (8.2.1, 8.3.4). It is also done in collaboration with B. T. Johansson, from University of Birmingham, GB.

This work has been submitted to a journal [40].

This work is an initial study of a numerical method for identifying multiple leak zones in saturated unsteady flow. Using the conventional saturated groundwater flow equation, the leak identification problem is modelled as a Cauchy problem for the heat equation and the aim is to find the regions on the boundary of the solution domain where the solution vanishes, since leak zones corresponds to null pressure values. To reconstruct the solution to the Cauchy problem in a stable way, we modify and employ an iterative regularizing method proposed recently. In this method, one solves mixed well-posed problems (obtained by changing the boundary conditions) for the heat operator as well as for its adjoint, to get a sequence of approximations to the original Cauchy problem. The mixed problems are solved using a Finite element method (FEM), and the numerical results obtained show that the leak zones can be accurately identified also when there is noise in the data.
6. New Results

6.1. Aggregation methods for optical flow computation

Participants: Charles Kervrann, Denis Fortun.

We address the problem of optical flow estimation, that is recovering the dense apparent motion of the pixels in a sequence of images. It is a fundamental computer vision task at the basis of a large variety of applications: object tracking, video compression, motion segmentation, movement detection, 3D reconstruction . . . Most of state-of-the-art methods rely on a common global variational framework [22], [31]. Computing optical flow amounts to minimizing a global energy. Our experiments demonstrated that the restriction of the minimization to local regions yields significant improvements of the estimation. Motivated by this fact, we developed a novel method to take advantage of this local approach by deriving the global estimate of the flow field from an aggregation of several local estimates. Our work can thus be seen as a general semi-local framework which can be used to improve the performance of any global variational method (Figure 3 ). We evaluated the performance of our approach on real and synthetic sequences. We investigate adaptations of this methodology to time-lapse fluorescence microscopy and we have recently performed comparisons with usual correlation techniques.

6.2. Lifetime estimation of moving vesicles in FLIM microscopy

Participants: Charles Kervrann, Philippe Roudot.

Fluorescence lifetime imaging microscopy (FLIM) is a widely spread imaging technique for sensing fluorophore environment in a living biological sample (like pH, ions...). Fluorescence lifetime (i.e. the average time a fluorophore stays in excited state before relaxing to its ground state possibly emitting a photon) is particularly useful to detect the Förster resonance energy transfer (FRET) which quantifies spatial proximity between molecules. We have proposed a statistical framework that exploits the intensity model of the frequency-domain FLIM output to jointly estimate trajectories and lifetimes of tracked vesicles. The proposed tracker, inspired from template cross-correlation or gaussian fitting, combines lifetime estimation and robust M-estimation in a efficient and fast way. Estimation of movement and lifetime are decoupled and alternatively performed, while particle/spot detection is performed on the first frame (Figure 4 ). To improve the results on real image sequences depicting moving vesicles, the background (cytoplasmic auto-fluorescence) model parameters and the scale parameters involved in the M-estimation procedure are estimated in our approach.

Partner: F. Waharte and J. Boulanger (UMR 144 CNRS PICT IBiSA Institut Curie)

6.3. Repetitive and transient event detection in fluorescence video-microscopy

Participants: Charles Kervrann, Pierre Hellier.

Progresses in imaging dynamic behaviours of molecules including fast video microscopy and the application of evanescent wave microscopy have allowed to image intracellular vesicular movements, exocytosis and endocytosis of fluorescent-tagged proteins. For an unbiased quantification of repetitive and transient events, we have proposed an approach which is versatile enough, to be applicable to diverse although complementary modes of microscopy. The proposed detection method described in [21], [11], [16] can be decomposed into three main steps: i) a first pre-processing step is dedicated to the normalization of the image sequence; ii) the second step is the patch-based detection procedure to detect unusual patterns; iii) a third post-processing step allows us to cluster and count detected events in space and time. In a more recent case study, we have used this approach to analyse image sequences depicting M10 cells stably expressing Langerin-YFP and to get deeper insights in the recycling pathway and dynamics of this molecule.
Figure 3. Average Angular Error (AAE) of vector flows obtained with the global variational method and three aggregation methods applied to three image sequences of the Middlebury (http://vision.middlebury.edu/flow/) database: row #1: first frame of the sequence; row #2: global variational method; row #3: mean aggregation; row #4: confidence-based weighted average; row #5: graph cut aggregation.
6.4. Atlas creation of fluorescence microscopy images

**Participan**: Pierre Hellier.

In this work, we consider the analysis of fluorescence images over time to account for two artifacts: fluorescence decreasing over time and geometric misalignment. A single exponential function is most commonly used to represent measured fluorescence decay profiles due to photobleaching. Accordingly, homologous points need to be geometrically aligned over time. Unfortunately, the living cell exhibits slow motion over time. We have considered the iterative estimation of both geometrical alignment and intensity correction by the creation of a 3D atlas.

6.5. Averaging of 3D volumes and denoising for the analysis of cryo-electron tomograms

**Participant**: Charles Kervrann.

Trichocysts are large vesicles secreted by the ciliated protozoa, Paramecium. They are characterized by the presence of large three-dimensional crystals of proteins. Under chemical or physical stimuli, or facing a predator, trichocysts undergo an exocytosis, right after fusion of their membrane with the unicellular organism plasma membrane. The crystalline mesh changes its conformation from a condensed to an extended shape in a few milliseconds. Nowadays, cryo-electron tomography (cryo-ET) allows one to visualize those crystals and so, to analyse their three-dimensional organization. However, two main impediments remain with this method. Samples are very sensitive to electron radiation involving the spreading out of the electron dose on the whole tilt series, causing the emergence of background noise in the images. Moreover, a lack of data occurs during image acquisition, called the “missing wedge”, due to uncovered angles at the moment of the acquisition of the tilt series. After tomogram reconstruction of four trichocysts, we have tested usual denoising methods (anisotropic diffusion, Fourier coefficient thresholding) and an unpublished patch-based denoising method.
inspired from the nD-safir software (5.1). The denoising methods improved the alignment of different crystal sub-volumes. The sub-volume averaging allowed us to fill in partially the “missing wedge” and then, to obtain a more faithful three-dimensional crystal reconstruction [17], [15].

**Partners:** E. Pollet, A. Guesdon and D. Chrétien (UMR 6026 CNRS University of Rennes 1)

### 6.6. Analysis of spatio-temporal dynamics of cytoplasmic actin under geometrical confinement

**Participant:** Charles Kervrann.

The generation, cell-cycle regulation and maintenance of such cellular functions are often correlated with symmetry breaking and spatiotemporal reorganization of F-actin assembly. In this study, we analyzed the spatiotemporal evolution of actin filaments using Xenopus meiotic extracts artificially confined within a geometry mimicking the cell boundary. It turns out the confinement of the cytoplasm generates symmetry breaking in the spatial organization of actin filaments. Combination of quantitative image analysis and biochemical perturbations show that both spatial localization of F-actin nucleators and actin turnover play a decisive role in generating symmetry breaking. In this project, we proposed to combine an optical flow-based tracker (Kanade-Lucas-Tomasi tracker [40], [36]) to a photobleaching correction method in order to extract quantitative spatiotemporal characteristics of the actin dynamics.

**Partners:** Z. Gueroui (BioPhysics team, UMR 8640 Ecole Normale Superieure, Paris) and M. Pinot (UMR 144 CNRS Institut Curie, Paris)

### 6.7. Analysis of lateral organization of ordered domains at the plasma membrane surface

**Participant:** Charles Kervrann.

Modifications of plasma membrane physical properties are also known to be involved in the perception and response to environmental modifications such as temperature, mechanical and osmotic stress in various organisms. We have analysed a recently designed probe, di-4-ANEPPDHQ, that can change its fluorescent properties depending on whether it is residing in ordered or disordered phases of the tobacco cell plasma membranes. We performed a spatial analysis (covariograms, Markov Random Fields) of small (<200 nm) ordered domains observed in multispectral confocal microscopy. We focused on binary images, assumed to be realizations of a MRF-Ising model, depicting the spatial organization of ordered domains. Maximum pseudo-likelihood methods were investigated to estimate parameters able to describe the spatial properties of ordered domains. We analyzed the modifications of the whole plasma membrane fluidity, and the distribution of ordered domains occurring in the few minutes following addition of the elicitor of defense reaction cryptogein to tobacco cells.

**Partners:** P. Gerbeau-Pissot, F. Simon-Plas (UMR 1088 PME INRA, Dijon) and K. Kiêu (MIA Unit INRA, Jouy-en-Josas)
SYMBIOSE Project-Team

6. New Results

6.1. Advanced tools for data management

Participants: Olivier Collin [contact], Dominique Lavenier, François Moreews, Olivier Sallou, Anthony Bretaudieu, Jonathan Piat.

- **Annotation and databases**: The AnnotQTL server is a tool designed to gather the functional annotation of genes from several institutional databases for a specific chromosomal region. [14] [Online publication: http://dx.doi.org/10.1093/NAR/GKR361]. SigReannot-mart is a query environment populated with regularly updated annotations for different oligo sets. It stores the results of the SigReannot pipeline that has mainly been used on farm and aquaculture species [17] [Online publication: http://database.oxfordjournals.org/content/2011/bar025]. BioMart Central Portal is a first of its kind, community-driven effort to provide unified access to dozens of biological databases spanning genomics, proteomics, model organisms, cancer data, ontology information and more. [11] [Online publication: http://database.oxfordjournals.org/content/2011/bar041].

- **Bioinformatics Workflow for Intensive Computation**
  SLICEE proposes to abstract the calls to the cluster scheduler by handling command submission and takes care of exploiting the data parallelism. It enables an easy implementation maintaining and sharing for bioinformatics workflows using intensive computation resources [32]. OBIWEE is a virtual cluster deployment tool associated with SLICEE. It can be deployed either on private cloud or a public cloud architecture. It helps at facing the increasing demand for bioinformatics intensive treatments, in a context of large dissemination of sequencing technologies usages. [29] http://vapor.gforge.inria.fr/. We also developed a library of bioinformatics softwares implemented on manycore structures such as GPU [27].

6.2. Sequences assembly, alignement and comparison

Participants: Dominique Lavenier [contact], Claire Lemaitre, Pierre Peterlongo, Fabrice Legeai, Guillaume Chapuis, Rayan Chikhi, Nicolas Maillet, Delphine Naquin, Raluca Uricaru, Pavlos Antoniou, Thomas Derrien.

- **Hardware accelerator**: Designing FPGA-based accelerators is a difficult and time-consuming task that can be eased by High Level Synthesis Tools. A C-to-hardware methodology has been used to develop an efficient systolic array for the genomic sequence alignment problem. [42], [25] [Online publication 1: http://www.eetimes.com/design/programmable-logic/4217568/How-to-accelerate-genomic-sequence-alignment-4X-using-half-an-FPGA?Ecosystem=programmable-logic] [Online publication 2: http://www.springerlink.com/content/37l00567qm18h146/]

- **De novo assembly of NGS data**: A novel framework has been introduced for de novo assembly of next-generation sequencing data. The new paired string graphs and localized assembly models are implemented in the Monument assembler [24]. [Online publication: http://www.springerlink.com/content/f5g305j5k73x3k14/]

- **International competition of de novo genome assembly**: The Symbiose team (IRISA/CNRS/ENS Cachan Brittany) participated to this competition. [9]. [Online publication: http://genome.cshlp.org/content/early/2011/09/16/gr.126599.111.abstract]

- **Indexation of NGS data**: A novel data structure is described for indexing NGS data. The structure is coupled with filtering algorithms that enable memory-efficient and parallel indexing. [23]

- **Breakpoints in genomes**: We analysed the correlation between 3D chromatin interaction data and breakpoint regions resulting from evolutionary rearrangements in the human genome. We found that two loci distant in the human genome but adjacent in the mouse genome are significantly more often observed in close proximity in the human nucleus than expected. [21]. [Online publication: http://www.biomedcentral.com/1471-2164/12/303]
• **Repeat detection**: A tool has been presented for detecting long similar fragments that occur two or more times in a set of biological sequences. This is achieved by using a filter as a preprocessing step, and by using the information that the filter has gathered also in the successive inference phase. [26]. [Online publication: http://www.stringology.org/event/2011/p08.html]

• **Targeted assembly of NGS data**: Mapsembler is an iterative targeted assembler which processes large datasets of reads on commodity hardware. Mapsembler checks for the presence of given regions of interest in the reads and reconstructs their neighborhood, either as a plain sequence (consensus) or as a graph (full sequence structure). [39]

• **Transcriptome assembly and annotation**: We established and analyzed two catalogues of transcripts by assembling EST sequences, and performed their functional annotations using the gene ontology for the following 2 species: *spodoptera littoralis* [15] and *cabomba* [20].

• **Substitution matrices**: A general and simple methodology has been proposed to build new matrices fitted to specific compositional bias of proteins. It was then applied to the large scale comparison of Mollicute AT-rich genomes [16]. [Online publication: http://www.biomedcentral.com/1471-2105/12/457/]

6.3. **Genome Structure**

**Participants**: Jacques Nicolas [contact], Dominique Lavenier, François Coste, Catherine Belleannée, Olivier Sallou, Fabrice Legeai, Guillaume Rizk, Guillaume Chapuis, Matthias Gallé, Anthony Bretaudeau.

• **GPU accelerated RNA folding algorithm**: The main kernel of the widely used RNA folding package Unafold has been accelerated using GPU boards by reordering computations to enable tiled computations and good data reuse [37], [2].

• **GPU accelerated QTL algorithm**: Our GPU/multicore implementation performs up to 20 times faster than the previous multicore implementation and allows extensive QTL analysis to be conducted in a reasonable time, while maintaining the same level of precision [35].

• **Hierarchical structure of genomes**: In [7], we proposed to split the classical smallest grammar problem into two tasks: (1) choosing the constituents of the grammar and (2) finding the *smallest grammar parsing* given these constituents. This defines properly the search space for this problem and, as we have shown how to solve in polynomial time the second task, this opens doors for new algorithms finding smaller grammars as shown on a generic compression benchmark (up to 10%). In [6], we have worked on the scalability to propose a new algorithm able to handle whole genomes: on this kind of sequences, the size reduction is still about 10% for a comparable execution time with respect to state-of-the-art algorithms.

• **Data compression**: By using grammars with rigid patterns as words, we were able to achieve a compression rate up to 25% better compared to the previous best DNA grammar-based coder, and just below state-of-the-art dedicated DNA compressors [1].

• **CRISPR Modeling and identification**: CRISPR (Clustered regularly interspaced short palindromic repeats) are small repeats present in a number of bacterial and archaeal species. We proposed the most complete database on these elements (http://crispi.genouest.org), elaborating for the first time a complete study of the palindromic nature of these repeats. The analysis has made an extensive use of our Logol Parser to decipher stem-loop structures [40].

• **Aphid genetics**: We participated in a genetic study aiming at comparing the rates of evolution of genes enclosed in aphid sexual chromosome (X) to autosomal genes. In order to do so, we provided particular microsatellites for the selection of genomic sequences, as well as tools for studying their genomic environment. [13] [Online publication: http://mbe.oxfordjournals.org/content/early/2011/10/12/molbev.msr252]
6.4. Protein Sequences and Structures

Participants: Rumen Andonov [contact], Antonio Mucherino, François Coste, Jacques Nicolas, Andres Burgos, Gaëlle Garet, Pavel Senin, Mathilde Le Boudic-Jamin.

- **Branch & Prune Algorithm**: We proposed an extension of the Branch & Prune (BP) algorithm for the Discretizable Molecular Distance Geometry Problem (DMDGP) which is able to exploit all symmetries of the research domain of the corresponding combinatorial optimization problem [30].

- **Modeling protein sequences with long distance correlations** To initiate this new line of research, we have set up a framework to learn context-free grammars on protein sequences based on the identification of conservation blocks and substitutability of non-terminals. A first implementation of the learning algorithms showed the interest of this approach [38].

- **Maximum Contact Map Overlap (CMO)** is a popular measure for quantifying the similarity between protein structures. A new integer programming model was presented for CMO and an exact branch-and-bound algorithm was designed with bounds obtained by a novel Lagrangian relaxation. The efficiency of the approach was demonstrated on known benchmarks on which sets our approach significantly outperforms the best existing exact algorithms [3].

- **Alignment of protein structures** First successes were obtained on provably optimal pairwise alignment of protein inter-residue distance matrices, using the popular Dali scoring function. We proposed the first mathematical model for computing optimal structural alignments based on dense inter-residue distance matrices and present algorithm engineering techniques to handle the huge integer linear programs [22]. In a second paper, a strategy was proposed for sparsifying distance matrices in which only the distances needed for uniquely reconstructing the conformations of the proteins are kept. [31].

- **Protein Family Identification** Identification of protein families is a computational biology challenge that needs efficient and reliable methods. First, we used the comparison tool A_purva, which is based on Contact Map Overlap (CMO), to classify protein structure coming from the CATH database. The obtained results showed that A_purva was able to correctly classify 92% of the structures, and that introducing the notion of dominance drastically reduced the computational time needed for classifying the protein structures [33]. Then, we introduced this concept of dominance in a novel combined approach based on Distance Alignment Search Tool (DAST), which contains an exact algorithm with bounds. Our experiments showed that this method successfully finds the most similar proteins in a set without solving all instances [28].

- **Local Protein Threading, sequence-structure alignment**: A novel approach to PTP has been investigated. It aligns a part of a protein structure onto a protein sequence in order to detect local similarities [8]. [Online publication: http://www.sciencedirect.com/science/article/B6TYW-50G78H4-1/2/947312da7a7bbf175cab7b3288ba4f03]

6.5. Systems Biology

Participants: Anne Siegel [contact], Jérémie Bourdon, Michel Le Borgne, Nathalie Theret, Geoffroy Andrieux, Oumarou Abdou-Arbi, Sylvain Prigent, Pierre Blavy, Andres Aravena, Santiago Videla, Valentin Wucher, Brivael Trelhu.

- **Average-case analysis for quantitative data integration** We proposed a probabilistic modeling framework that integrates heterogeneous data. Average case analysis methods were used in combination with Markov chains to link qualitative information about transcriptional regulations to quantitative information about protein concentrations. The approach was illustrated by modeling the carbon starvation response in Escherichia coli. It accurately predicted the quantitative time-series evolution of several protein concentrations using only knowledge of discrete gene interactions and a small number of quantitative observations on a single protein concentration [5]. [Online publication: http://dx.plos.org/10.1371/journal.pcbi.1002157]
• **Combining genetic and metabolic regulations:** We mixed Gale-Nikaido reduction steps and differential inequalities to understand how genetic regulation modifies the behavior of a very abstracted model of lipid metabolism [18]. [Online publication: http://www.springerlink.com/content/n437048670560782/]

• **Extract relevant information with respect to a cancer phenotype:** We designed dedicated logical rules to model the static response of biomolecular interactions implied in the cancer network. This allowed us to trace back genes implied in the cancer phenotype [12]. [Online publication: http://www.computer.org/portal/web/csdl/doi/10.1109/TCBB.2010.71]

• **Integrative biology for brown algal** We proposed a protocol focusing on integrating heterogeneous knowledge gained on brown algal metabolism. The resulting abstraction of the system helps understanding how brown algae cope with changes in abiotic parameters within their unique habitat [19].

• **Search for key regulators** A method was proposed to model the effects of all transcriptional and metabolic regulations contained in transpath in a single influence network. The network was analyzed to find a set of candidates that explain the variations of a set of targets [34].

• **Identification of co-regulation patterns.** We introduced a new approach based on the compilation of Simple Shared Motifs (SSM), groups of sequences defined by their length and similarity and present in conserved sequences of gene promoters. We proved that Simple Shared Motifs analysis provides a clearer definition of expression networks [10]. [Online publication: http://www.biomedcentral.com/1471-2105/12/365]

• **Probabilistic models for systems biology** We reviewed, in a book chapter, some classical concepts concerning probabilistic models and their applications in systems biology. Probabilistic boolean networks were presented in deep with a focus on the effect of synchronization of genes and on stochastic simulation of such networks [36].
6. New Results

6.1. Image Segmentation, Registration and Analysis

6.1.1. Quantitative Analysis of Open Curves in Brain Imaging: Applications to White Matter Fibers and Sulci

Participants: Meena Mani, Christian Barillot.

Shape, scale, orientation and position, the four physical features associated with open curves, have different properties so the usual approach has been to design different metrics and spaces to treat them individually. We took an alternative approach using a comprehensive Riemannian framework where joint feature spaces allow for analysis of combinations of features. We can compare curves by using geodesic distances, which quantify their differences. We validated the metrics we used, demonstrated practical uses and applied the tools to important clinical problems. To begin, specific tract configurations in the corpus callosum are used to showcase clustering results that depend on the Riemannian distance metric used. This nicely argues for the judicious selection of metrics in various applications, a central premise in our work. The framework also provides tools for computing statistical summaries of curves. We represented fiber bundles with a mean and variance, which describes their essential characteristics. This is both a convenient way to work with a large volume of fibers and is a first step towards statistical analysis. Next, we designed and implemented methods to detect morphological changes, which can potentially track progressive white matter disease. With sulci, we addressed the specific problem of labeling. An evaluation of physical features and methods such as clustering leads us to a pattern matching solution in which the sulcal configuration itself is the best feature.

6.1.2. Trimmed-likelihood estimation for focal lesions and tissue segmentation in multisequence MRI for multiple sclerosis

Participants: Sylvain Prima, Christian Barillot.

Following Daniel Garcia-Lorenzo’s PhD, we proposed a new automatic method for segmentation of multiple sclerosis (MS) lesions in magnetic resonance images. The method performs tissue classification using a model of intensities of the normal appearing brain tissues. In order to estimate the model, a trimmed likelihood estimator is initialized with a hierarchical random approach in order to be robust to MS lesions and other outliers present in real images. The algorithm was first evaluated with simulated images to assess the importance of the robust estimator in presence of outliers. The method was then validated using clinical data in which MS lesions were delineated manually by several experts. Our method obtains an average Dice similarity coefficient (DSC) of 0.65, which is close to the average DSC obtained by raters (0.66) [15].

6.1.3. Segmentation of Multimodal Brain Images using Spectral Gradient and Graph Cut

Participants: Camille Maumet, Jean-Christophe Ferré, Christian Barillot.

Following Jeremy Lecoeur’s PhD, we have introduced a new and original scale-space approach for segmenting normal and pathological tissue from multidimensional images. This method can perform a joint segmentation of three complementary imaging volumes at the same time by embedding a scale-space color invariant edge detector - i.e. the spectral gradient - as the boundary term in a graph cut optimization framework. Finally, we have proposed to extend this new scheme to more than three channels. We focussed the contribution onto the segmentation of tissues or structures of interest from multi-dimensional / multi-sequences brain MRI. This new multidimensional segmentation framework has been validated on simulated data and on clinical data (both pathological and healthy brains). We have exhibited the performances of this new method on various combinations of MRI sequences for the segmentation of normal and pathological tissues and showed how it is able to out perform competitive works. This work is under submission to an international journal.
6.1.4. Adaptive pixon represented segmentation for 3D MR brain images based on mean shift and Markov random fields

**Participant:** Christian Barillot.

Following Lei Lin and Daniel Garcia Lorenzo’s PhDs, we proposed an adaptive pixon represented segmentation (APRS) algorithm for 3D magnetic resonance (MR) brain images. Different from traditional methods, an adaptive mean shift algorithm was adopted to adaptively smooth the query image and create a pixon-based image representation. Then K-means algorithm was employed to provide an initial segmentation by classifying the pixons in the image into a predefined number of tissue classes. By using this segmentation as initialization, expectation-maximization (EM) iterations composed of bias correction, a priori digital brain atlas information, and Markov random field (MRF) segmentation were processed. Pixons were assigned with final labels when the algorithm converges. The adoption of bias correction and brain atlas made the current method more suitable for brain image segmentation than the previous pixon-based segmentation algorithm. The proposed method was validated on both simulated normal brain images from BrainWeb and real brain images from the IBSR public dataset. Compared with some other popular MRI segmentation methods, the proposed method exhibited a higher degree of accuracy in segmenting both simulated and real 3D MRI brain data. The experimental results were numerically assessed using Dice and Tanimoto coefficient [18].

6.1.5. EM-ICP strategies for joint mean shape and correspondences estimation: applications to statistical analysis of shape and of asymmetry

**Participant:** Sylvain Prima.

In collaboration with B. Combès, we proposed a new approach to compute the mean shape of unstructured, unlabelled point sets with an arbitrary number of points. This approach can be seen as an extension of the EM-ICP algorithm, where the fuzzy correspondences between each point set and the mean shape, the optimal nonlinear transformations superposing them, and the mean shape itself, are iteratively estimated. Once the mean shape is computed, one can study the variability around this mean shape (e.g. using PCA) or perform statistical analysis of local anatomical characteristics (e.g. cortical thickness, asymmetry, curvature). To illustrate our method, we performed statistical shape analysis on human osseous labyrinths and statistical analysis of global cortical asymmetry on control subjects and subjects with situs inversus [29]. This work was led within the ARC 3D-MORPHINE (http://3dmorphine.inria.fr).

6.1.6. Surface-based method to evaluate global brain shape asymmetries in human and chimpanzee brains

**Participant:** Sylvain Prima.

Following PhD and PostDoc works from Benoit Combès and Marc Fournier, in this work we used humans and chimpanzees brain MRI databases to develop methods for evaluating global brain asymmetries. We performed brain segmentation and hemispheric surface extraction on both populations. The human brain segmentation pipeline was adapted to chimpanzees in order to obtain results of good quality. To alleviate the problems due to cortical variability we proposed a mesh processing algorithm to compute the brain global shape. Surface-based global brain asymmetries were computed on chimpanzee and human subjects using individual mid-sagittal plane evaluation and population-level mean shape estimation. Asymmetry results were presented in terms of axis-wise components in order to perform more specific evaluation and comparison between the two populations [35]. This work was led within the ARC 3D-MORPHINE (http://3dmorphine.inria.fr).

6.1.7. Computational techniques for the analysis of endocranial cast and endocranial structures

**Participant:** Sylvain Prima.
Following Phd and post-doc works from Benoit Combès and Marc Fournier, a series of studies were led within the ARC 3D-MORPHINE (http://3dmorphine.inria.fr) and were presented at the 1836th Journées de la Société d’Anthropologie de Paris (January 26-28) and at the 80th annual meeting of the American Association of Physical Anthropologists (April 12-16). These include: a method to assess 3D endocranial asymmetries in extant and fossil species: new insights into paleoneurology [48]; a method to map the distance between the brain and the inner surface of the skull [51], [34]; a method to compare bony labyrinths in humans, chimpanzees and baboons [28]; a method to compare the endocranial shape and its relationship with ectocranial structures [41]; a new reconstruction of the frontal lobe and temporal pole of the Taung (Australopithecus africanus) endocast [32].

6.1.8. Evaluation of Registration Methods on Thoracic CT: The EMPIRE10 Challenge

Participant: Olivier Commowick.

We participated, as part of a collaboration with the Asclepios team, to the EMPIRE10 challenge on registration. EMPIRE10 (Evaluation of Methods for Pulmonary Image REgistration 2010) is a public platform for fair and meaningful comparison of registration algorithms which are applied to a database of intra-patient thoracic CT image pairs. Evaluation of non-rigid registration techniques is a non trivial task. This is compounded by the fact that researchers typically test only on their own data, which varies widely. For this reason, reliable assessment and comparison of different registration algorithms has been virtually impossible in the past. In this work we present the results of the launch phase of EMPIRE10, which comprised the comprehensive evaluation and comparison of 20 individual algorithms from leading academic and industrial research groups. All algorithms are applied to the same set of 30 thoracic CT pairs. Algorithm settings and parameters are chosen by researchers expert in the configuration of their own method and the evaluation is independent, using the same criteria for all participants. All results are published on the EMPIRE10 website (http://empire10.isi.uu.nl). The challenge remains ongoing and open to new participants. Full results from 24 algorithms have been published at the time of writing. This article details the organisation of the challenge, the data and evaluation methods and the outcome of the initial launch with 20 algorithms. More details are available in [20].

6.2. Image processing on Diffusion Weighted Magnetic Resonance Imaging

6.2.1. Diffusion Directions Imaging (DDI)

Participants: Aymeric Stamm, Christian Barillot.

Diffusion magnetic resonance imaging (dMRI) is the reference in vivo modality to study the connectivity of the brain white matter. Images obtained through dMRI are indeed related to the probability density function (pdf) of displacement of water molecules subject to restricted diffusion in the brain white matter. The knowledge of this diffusion pdf is therefore of primary importance. Several methods have been devised to provide an estimate of it from noisy dMRI signal intensities. They include popular diffusion tensor imaging (DTI) as well as higher-order methods. These approaches suffer from important drawbacks. Standard DTI cannot directly cope with multiple fiber orientations. Higher-order approaches can alleviate these limitations but at the cost of increased acquisition time. We have proposed, in the same vein as DTI, a new parametric model of the diffusion pdf with a reasonably low number of parameters, the estimation of which does not require acquisitions longer than those used in clinics for DTI. This model also accounts for multiple fiber orientations. It is based on the assumption that, in a voxel, diffusing water molecules are divided into compartments. Each compartment is representative of a specific fiber orientation (which defines two opposite directions). In a given compartment, we further assume that water molecules that diffuse along each direction are in equal proportions. We then focus on modeling the pdf of the displacements of water molecules that diffuse only along one of the two directions. Under this model, we derive an analytical relation between the dMRI signal intensities and the parameters of the diffusion pdf. We exploit it to estimate these parameters from noisy signal intensities. We carry out a cone-of-uncertainty analysis to evaluate the accuracy of the estimation of the fiber orientations and we evaluate the angular resolution of our method. Finally, we show promising results on real data and propose a visualization of the diffusion parameters which is very informative to the neurologist. This work was conducted in collaboration with Patrick Perez from Technicolor [56].
6.2.2. **Anatomy of the corticospinal tracts: evaluation of a deterministic tractography method**  
**Participants:** Romuald Seizeur, Nicolas Wiest-Daesslé, Sylvain Prima, Camille Maumet, Jean-Christophe Ferré, Xavier Morandi.

In this work, anatomical, diffusion-weighted and functional 3T MRI were acquired on 15 right-handed healthy subjects to analyse the portions of the corticospinal tract (CST) dedicated to hand motor and sensory functions. The three MR images were then registered and regions of interest were delineated i) in the mid-brain using 3D T1-weighted MRI, and ii) in the cortex using fMRI using hand motor and sensory tasks. Deterministic tractography was then performed using these two ROIs from diffusion-weighted MRI after the diffusion tensors were computed. The ventrolateral tract fibers of the CST were generally not properly identified, due to fiber crossing in the corona radiata [55].

6.2.3. **Tracking of the Hand Motor Fibers within the Corticospinal Tract Using Functional, Anatomical and Diffusion MRI**  
**Participants:** Romuald Seizeur, Nicolas Wiest-Daesslé, Olivier Commowick, Sylvain Prima, Aymeric Stamm, Christian Barillot.

In this work, we proposed to compare three diffusion models to track the portion of the corticospinal tract dedicated to the hand motor function (called hand motor fibers hereafter), using diffusion, functional and anatomical MRI. The clinical diffusion data have few gradient directions and low b-values. In this context, we show that a newly introduced model, called diffusion directions imaging (DDI) outperforms both the DTI and the ODF models. This new model allows to capture several diffusion directions within a voxel, with only a low number of parameters. Two important results are that i) the DDI model is the only one allowing consistent tracking from the mesencephalon to the most lateral part of the cortical motor hand area, and that ii) the DDI model is the only model able to show that the number of hand motor fibers in the left hemisphere is larger than in the contralateral hemisphere for right-handed subjects; the DDI model, as the other two models, fails to find such a difference for left-handed subjects. To the best of our knowledge, this is the first time such results are reported, at least on clinical data. [44].

6.2.4. **Multifiber Deterministic Streamline Tractography Based on a New Diffusion Model**  
**Participants:** Olivier Commowick, Romuald Seizeur, Nicolas Wiest-Daesslé, Sylvain Prima, Aymeric Stamm, Christian Barillot.

In this work, we have built upon a new model, describing the random motion of water molecules in fibrous tissues, to develop a multifiber deterministic tractography algorithm. We apply this algorithm to track the corticospinal tract of the human brain, in both controls and patients with tumors. [31].

6.2.5. **Automated detection of white matter fiber bundles**  
**Participant:** Olivier Commowick.

This work is part of a collaboration with the Computational Radiology Laboratory headed by Simon Warfield in Boston, USA. For this topic, we have studied how white matter fiber bundles can be extracted in a reproducible way from diffusion tensor MRI. Usually, white matter (WM) fiber bundles of the brain can be delineated by diffusion tractography utilizing anatomical regions-of-interest (ROI). These ROIs can specify seed regions in which tract generation algorithms are initiated. Interactive identification of such anatomical ROIs enables the detection of the major WM fiber tracts, but suffers from inter-rater and intra-rater variability, and is time consuming. We developed and compared three techniques for automated delineation of ROIs for the detection of two major WM fiber tracts in 12 healthy subjects. Tracts identified automatically were compared quantitatively to reference standard tracts derived from carefully hand-drawn ROIs. Based on comparative performance of the experimental techniques, a multi-template label fusion algorithm was found to generate tracts most consistent with the reference standard. More details on this work are available in [43].
6.3. Management of Information and Semantic Processing

6.3.1. NeuroLOG project: Sharing of data and sharing of processing tools in neuroimaging

Participants: Bernard Gibaud, Bacem Wali.

The NeuroLOG project (ANR ANR-06-TLOG-024) came to its end in December 2010. However, we managed to maintain the NeuroLOG platform in operation, which is important with regards to publication. Several papers are in preparation. A lot of efforts were devoted in 2011 to submit a new proposal to ANR, building on NeuroLOG’s achievements. A NeuroLOG2 project was submitted in March to the ANR TECSAN program (health technology). This new project aimed both at going on developing the technology for sharing data and processing tools, while being more involved in neuroimaging applications. Two application fields were proposed, concerning research on Alzheimer Disease, on the one hand, and epilepsy, on the other hand. The consortium was enlarged accordingly, with the integration of new partners such as the EDELWEISS project (INRIA, Sophia), the U642 LTSI (INSERM, Rennes) and U1028 CNRL (INSERM, Lyon). A new submission is envisaged in 2012, taking into account the recommendations of ANR.

6.3.2. Semantic annotation of anatomic images in neuroimaging

Participants: Bernard Gibaud, Tristan Moreau, Xavier Morandi.

This project aims at exploring the feasibility of relying on symbolic knowledge provided by ontologies to assist the annotation of anatomical images. The basic assumption underlying this work is that ontologies not only can provide a reference vocabulary to annotate images, but they can also provide useful prior knowledge that may help the annotation process itself, an assumption supported by the interesting results obtained by Ammar Mechouche in his PhD work. The current study, initiated in 2010 in the context of the Master student work of Elsa Magro (analysis of intra-precentral connections and of the U-fibers connecting the precentral and postcentral gyri) was pursued in 2011 (PhD work of Tristan Moreau). Our most recent works try to establish a parcellation of the grey-matter white matter surface based on the connectivity profiles of individual points of this surface, valid for a population of subjects. This is a prerequisite before identifying the more salient fiber bundles to be modelled in our ontology.

6.3.3. Semantic annotation of models and simulated medical images

Participants: Bernard Gibaud, Germain Forestier.

This project is carried out in the context of the Virtual Imaging Platform (VIP) project, an ANR project aiming at setting up a platform for facilitating the use of image simulation software in medical imaging, and coordinated by Creatis (Lyon). The platform will integrate simulation software to generate image of different modalities (i.e. MR, CT, PET, US). In this project, VISAGES is in charge of coordinating the development of an application ontology to support the annotation of the data shared in this platform (simulated images, anatomical models and physiological models used in simulations), as well as the annotation of simulation software components, in order to facilitate their interoperability within the platform. The work completed in 2011 is a continuation of what was started in 2010. Our major result in 2011 is an ontology allowing to annotate the models used for medical image simulations. Actually models are composed of files containing images (3D voxel maps) or surfaces (meshes). Our ontology includes entities called model layers associated with those files and depicting the model contents in terms of: anatomical structures, pathological structures, foreign bodies, contrast agents etc. Each individual object present in the model is referred to by an object layer part to which physical parameter distributions can be associated, that are used by simulation software. The ontology was modelled as OntoSpec documents (a methodology defined by Gilles Kassel in Amiens), then implemented in OWL. A preliminary version of this model was presented at a workshop organized by EBI in Cambridge in March 2011 (in the context of the VPH/RICORDO project). A more complete version was presented at the CBMS’2011 Conference in Bristol. VIP is a collaborative project, supported by ANR (Agence National de la Recherche), through grant ANR-AA-PPPP-000. The partners with whom we have the tightest relations are: Creatis (Lyon), I3S (Sophia), CEA-LETI (Grenoble).
6.4. Image Guided Intervention

6.4.1. Classification of Surgical Process using Dynamic Time Warping

**Participants:** Pierre Jannin, Germain Forestier, Florent Lalys, Brivael Trelhu.

Toward the creation of new computer-assisted intervention systems, Surgical Process Models (SPMs) are more and more used as a tool for analyzing and assessing surgical interventions. SPMs represent Surgical Process (SPs) which are defined as symbolic structured descriptions of surgical interventions, using a pre-defined level of granularity and a dedicated terminology. In this context, an important challenge is the creation of new metrics for the comparison and the evaluation of SPs. Thus, correlations between these metrics and pre-operative data allow to classify surgeries and highlight specific information on the surgery itself and on the surgeon, such as its level of expertise. In this study, we explored the automatic classification of a set of SPs based on the Dynamic Time Warping (DTW) algorithm. DTW allows to compute a distance between two SPs that focuses on the different types of activities performed during the surgery and their sequencing, by minimizing the time differences. Indeed, it turns out to be a complementary approach to classical methods focusing only on the time and the number of activities differences. Experiments were carried out on 24 lumbar disc herniation surgeries to discriminate the level of expertise of surgeons according to prior classification of SPs. Supervised and unsupervised classification experiments have shown that this approach was able to automatically identify groups of surgeons according to their level of expertise (senior and junior), and opens many perspectives for the creation of new metrics for surgeries comparison and evaluation. This work was performed in collaboration with Dr. Laurent Riffaud, and was published in the International Journal of Biomedical Informatics [14].

6.4.2. Surgical phases detection from microscope videos by machine learning

**Participants:** Pierre Jannin, Florent Lalys, Xavier Morandi.

Surgical process analysis and modeling is a recent and important topic aiming at introducing a new generation of computer-assisted surgical systems. Among all of the techniques already in use for extracting data from the Operating Room, the use of image videos allows automating the surgeons’ assistance without altering the surgical routine. In collaboration with Carl Zeiss Medical Systems (Oberkochen, Germany), we proposed an application-dependent framework able to automatically extract the phases of the surgery only by using microscope videos as input data and that can be adaptable to different surgical specialties. First, four distinct types of classifiers based on image processing were implemented to extract visual cues from video frames. Each of these classifiers was related to one kind of visual cue: visual cues recognizable through color were detected with a color histogram approach, for shape-oriented visual cues we trained a Haar classifier, for texture-oriented visual cues we used a bag-of-word approach with SIFT descriptors, and for all other visual cues we used a classical image classification approach including a feature extraction, selection, and a supervised classification. The extraction of this semantic vector for each video frame then permitted to classify time series using either Hidden Markov Model or Dynamic Time Warping algorithms. The framework was validated on cataract surgeries, obtaining accuracies of 95%. This work was performed in collaboration with Laurent Riffaud and was published at the ORASIS and MICCAI conferences.

6.4.3. Surgical tools recognition and pupil segmentation for cataract surgery modeling

**Participants:** Pierre Jannin, Florent Lalys.

In the above project work performed through the MS intership of David Bouget, we focus on developing an application-dependent framework able to extract surgical phases from microscope videos. The aim of this study was to enhance results of this framework by adding new visual cues extraction modules. We studied two modules: one to segment the pupil and one to extract and recognize surgical tools. Validation studies, performed with cataract surgery videos, show an increase of the framework accuracy to detect eight surgical phases. This work has been accepted at the MMVR 2012 international conference.

6.4.4. Automatic computation of electrode trajectories for Deep Brain Stimulation: a hybrid symbolic and numerical approach

**Participants:** Pierre Jannin, Florent Lalys, Camille Maumet, Claire Haegelen.
The optimal electrode trajectory is needed to assist surgeons in planning Deep Brain Stimulation (DBS). We developed and tested a method for image-based trajectory planning. Rules governing the DBS surgical procedure were defined with geometric constraints. A formal geometric solver using multimodal brain images and a template built from 15 brain MRI scans were used to identify a space of possible solutions and select the optimal one. For validation, a retrospective study of 30 DBS electrode implantations from 18 patients was performed. A trajectory was computed in each case and compared with the trajectories of the electrodes that were actually implanted. Computed trajectories had an average difference of 6.45 degrees compared with reference trajectories and achieved a better overall score based on satisfaction of geometric constraints. Trajectories were computed in 2min for each case. We demonstrated that a rule-based solver using pre-operative MR brain images can automatically compute relevant and accurate patient-specific DBS electrode trajectories. This work was published in the International Journal of Computer Assisted Radiology and Surgery.

6.4.5. Analysis of electrodes’ placement and deformation in deep brain stimulation from medical images

Participants: Pierre Jannin, Florent Lalys, Alexandre Abadie, Xavier Morandi, Claire Haegelen.

This work was performed during the internshiof Maroua Mehri. Deep brain stimulation (DBS) is used to reduce the motor symptoms such as rigidity or bradykinesia, in patients with Parkinson’s disease (PD). The Subthalamic Nucleus (STN) has emerged as prime target of DBS in idiopathic PD. However, DBS surgery is a difficult procedure requiring the exact positioning of electrodes in the pre-operative selected targets. This positioning is usually planned using patients’ pre-operative images, along with digital atlases, assuming that electrode’s trajectory is linear. However, it has been demonstrated that anatomical brain deformations induce electrode’s deformations resulting in errors in the intra-operative targeting stage. In order to meet the need of a higher degree of placement accuracy and to help constructing a computer-aided-placement tool, we studied the electrodes’ deformation in regards to patients’ clinical data (i.e., sex, mean PD duration and brain atrophy index). Firstly, we presented an automatic algorithm for the segmentation of electrode’s axis from post-operative CT images, which aims to localize the electrodes’ stimulated contacts. To assess our method, we applied our algorithm on 25 patients who had undergone bilateral STNDBS. We found a placement error of 0.91±0.38 mm. Then, from the segmented axis, we quantitatively analyzed the electrodes’ curvature and correlated it with patients’ clinical data. We found a positive significant correlation between mean curvature index of the electrode and brain atrophy index for male patients and between mean curvature index of the electrode and mean PD duration for female patients. These results help understanding DBS electrode’ deformations and would help ensuring better anticipation of electrodes’ placement. This work has been accepted at the SPIE Medical Imaging 2012 conference.

6.5. Medical Image Computing in Brain Pathologies

6.5.1. Detection of cortical abnormalities in drug resistant epilepsy

Participants: Elise Bannier, Camille Maumet, Jean-Christophe Ferré, Jean-Yves Gauvrit, Christian Barillot.

Focal cortical dysplasia and heterotopias are a recognized cause of epilepsy. Indication for drug resistant epilepsy surgery relies on precise localization and delineation of the epileptogenic zone and lesion identification is an important issue. Visual detection and delineation of small or occult focal cortical dysplasia and heterotopias on MR images are sometimes difficult. The Double Inversion Recovery (DIR) imaging, by nulling white matter and cerebrospinal fluid signal, seems particularly appropriate to detect intracortical lesions in MS and Epilepsy. In this work we evaluated at 3T and using voxel-based morphometry (VBM) the ability of a 9-minute 3D DIR sequence to detect cortical and juxtacortical lesions in drug resistant epileptic patients. Results on 21 patients and 20 healthy volunteers show the potential of 3D DIR VBM to detect cortical abnormalities. Further work will investigate the use of alternate registration frameworks (e.g. DARTEL), improved intensity normalization of 3D DIR images and joint 3D T1-w/DIR analysis to improve detection sensitivity and specificity.
6.5.2. Multi-modal NMR cartography of USPIO positive and negative tissues in MS human models

Participants: Olivier Luong, Olivier Commowick, Christian Barillot.

The main objective of this work was to build an input object for an MRI simulator. Each voxel of the object is defined by its three physical entities which are $T_1$, $T_2$ and $\rho$ MR relaxation parameters. In our case, this object comes from Multiple Sclerosis brains. We initially defined a simplified model with respect to pathological regions, based on a combination of the Brainweb template and the lesion manually delineated from pathological images. From this, we allocated relaxation parameters for each voxels of these ROI based on fixed values of $T_1$, $T_2$ and $\rho$ (initialized from in vivo relaxometry acquisitions). This model model does not allow to obtain a fine description of the pathological regions, as potentially defined by differential contrats between USPIO and Gd enhanced images. In order to obtain this finer description, we used an MRI simulator based on the Bloch’s equations, in order to estimated the $T_1$, $T_2$ and $\rho$ parameters on each voxel from initial conditions coming from in-vivo images acquired in Rennes by using the USPIO-6 protocol.

This work is part of the VIP collaborative project, supported by ANR (Agence National de la Recherche), through grant ANR-AA-PPPP-000. The partners with whom we have the tightest relations are: Creatis (Lyon), I3S (Sophia), CEA-LETI (Grenoble).

6.6. Vascular Imaging and Arterial Spin Labelling

6.6.1. Arterial spin labeling for motor activation mapping at 3T

Participants: Jan Petr, Aymeric Stamm, Elise Bannier, Jean-Christophe Ferré, Jean-Yves Gauvrit, Christian Barillot.

Functional arterial spin labeling (fASL) is an innovative biomarker of neuronal activation that allows direct and absolute quantification of activation-related CBF and is less sensitive to venous contamination than BOLD fMRI. This study evaluated fASL for motor activation mapping in comparison with BOLD fMRI in terms of involved anatomical area localization, intra-individual reproducibility of location, quantification of neuronal activation, and spatial accuracy. Imaging was performed at 3T with a 32-channel coil and dedicated post-processing tools were used. Twelve healthy right-handed subjects underwent fASL and BOLD fMRI while performing a right hand motor activation task. Three sessions were performed 7 days apart in similar physiological conditions. Our results showed an activation in the left primary hand motor area for all 36 sessions in both fASL and BOLD fMRI. The individual functional maps for fASL demonstrated activation in ipsilateral secondary motor areas more often than the BOLD fMRI maps. This finding was corroborated by the group maps. In terms of activation location, fASL reproducibility was comparable to BOLD fMRI, with a distance between activated volumes of 2.1 mm and an overlap ratio for activated volumes of 0.76, over the 3 sessions. In terms of activation quantification, fASL reproducibility was higher, although not significantly, with a CVintra of 11.6% and an ICC value of 0.75. Functional ASL detected smaller activation volumes than BOLD fMRI but the areas had a high degree of co-localization. In terms of spatial accuracy in detecting activation in the hand motor area, fASL had a higher specificity (43.5%) and a higher positive predictive value (69.8%) than BOLD fMRI while maintaining high sensitivity (90.7%). The high intra-individual reproducibility and spatial accuracy of fASL revealed in the present study will subsequently be applied to pathological subjects [25].

6.6.2. Construction and evaluation of a quantitative ASL brain perfusion template at 3T

Participants: Jan Petr, Elise Bannier, Jean-Christophe Ferré, Jean-Yves Gauvrit, Christian Barillot.
Arterial spin labeling (ASL) allows non-invasive imaging and quantification of brain perfusion by magnetically labeling blood in the brain-feeding arteries. ASL has been used to study cerebrovascular diseases, brain tumors and neurodegenerative disorders as well as for functional imaging. The use of a perfusion template could be of great interest to study inter-subject regional variation of perfusion and to perform automatic detection of individual perfusion abnormalities. However, low spatial resolution and partial volume effects (PVE) issues inherent to ASL acquisitions remain to be solved. The purpose of this study is to enhance the template quality by using DARTEL non-rigid registration and by correcting for PVE. PICORE-Q2TIPS ASL datasets were acquired on 25 healthy volunteers at 3T. Four methods of creating the template were evaluated using leave-one-out cross correlation. Subsequently, these methods were applied to hyper-perfusion detection on functional ASL data of 8 healthy volunteers and compared with the standard generalized linear model (GLM) activation detection [40].

6.6.3. Evaluation of functional arterial spin labeling data using a perfusion template

Participants: Jan Petr, Elise Bannier, Jean-Christophe Ferré, Jean-Yves Gauvrit, Christian Barillot.

ASL allows non-invasive imaging and quantification of brain perfusion by magnetically labeling blood in the brain-feeding arteries. In this study, a template created from perfusion images of 25 resting healthy subjects was used to automatically detect hyper perfusion patterns of 8 other subjects. DARTEL registration was used to improve the precision of the template and partial volume correction to prevent interpolation artifacts. This study showed that a perfusion template can be used to assess task-related activation zones in functional ASL data while using only activated phase. Two assumptions can be made to explain why standard functional analysis yields slightly larger activation regions. First, the use of FWHM 6mm Gaussian kernel possibly enlarges the detected zones. Second, the data analyzed using SPM contains both resting and activated phases whereas only the activated phase was compared to the template. Future work will focus on detection of hyperperfusion in different neurodegenerative diseases taking into account registration issues of pathological T1 images. [24].

6.6.4. A contrario detection of focal brain perfusion abnormalities based on an ASL template

Participants: Camille Maumet, Elise Bannier, Jean-Christophe Ferré, Pierre Maurel, Christian Barillot.

Arterial Spin Labeling (ASL) is a recent MRI perfusion technique which enables quantification of cerebral blood flow (CBF). The presence of regions with atypical CBF can characterize a pathology. In brain tumors for instance, perfusion increase can be directly related to the grading of the malignant tissues. It is therefore of great interest to identify these regions in order to provide the patients with the most appropriate therapy. In this work, we proposed to detect abnormal brain perfusion by using an a contrario framework and an ASL template as a model of normal perfusion. Validation was undertaken by qualitative comparison with CBF extracted from dynamic susceptibility weighted contrast enhanced (DSC) sequence. We experimented this framework on four patients presenting brain tumors. Results show that high perfusion regions found in DSC CBF maps are correctly identified as hyperperfusions with a contrario detection. Automatic detection has clear advantages over manual delineation since it is less time-consuming, does not depend on medical expertise and allows quantification of perfusion abnormalities within the detected regions.

6.6.5. Peripheral angiography using non-contrast enhanced imaging

Participants: Elise Bannier, Isabelle Corouge, Nicolas Wiest-Daesslé.

Arteriography, CT and MR angiography are routinely performed in patients presenting peripheral arteriopathy. Yet, contrast agent injection is contraindicated in patients with renal insufficiency and the underlying risk of developing nephrogenic systemic fibrosis further encourages research on non-contrast enhanced MR angiography techniques (NCE MRA). In this context, we evaluated at 3T the ability of a 2 NCE MRA new sequences to reliably detect peripheral vascular abnormalities from the abdominal aorta to the calf in comparison with CE MRA.
A first study including 20 healthy volunteers and 4 patients evaluated the NCE ECG-gated T2 TSE NATIVE SPACE MRI sequence. It demonstrated its potential in noninvasively imaging peripheral vasculature, from the abdominal aorta to the calf, within a clinically acceptable acquisition duration. Although signal inhomogeneity and peristalsis artifacts were sometimes observed in the abdominal aortic station, very good image quality was obtained on all subjects on lower stations, with no venous contamination.

A second study evaluated the NCE ECG-gated Quiescent Interval Single Shot (QISS) MRA sequence. Preliminary results obtained on 11 patients show that several lesions were not detected with QISS MRA especially on the thigh station. Ongoing patient inclusions are required to confirm these results. Finally, a concomitant NCE and CE MRA reading will be performed to compare stenosis grading, stenosis-thrombosis mismatch and lesions not detected with NCE MRA.

6.7. Abnormal functional lateralization and activity of language brain areas in developmental dysphasia

Participants: Clément De Guibert, Camille Maumet, Jean-Christophe Ferré, Pierre Jannin, Christian Barillot.

Atypical functional lateralization and specialization for language have been proposed to account for developmental language disorders, yet results from functional neuroimaging studies are sparse and inconsistent. This functional magnetic resonance imaging study compared children with a specific subtype of specific language impairment affecting structural language, to a matched group of typically developing children using a panel of four language tasks neither requiring reading nor metalinguistic skills, including two auditory lexico-semantic tasks (category fluency and responsive naming) and two visual phonological tasks based on picture naming. Data processing involved normalizing the data with respect to a matched pairs paediatric template, groups and between-groups analysis, and laterality indices assessment within regions of interest using single and combined task analysis. Children with specific language impairment exhibited a significant lack of left lateralization in all core language regions (inferior frontal gyrus-opercularis, inferior frontal gyrus-triangularis, supramarginal gyrus and superior temporal gyrus), across single or combined task analysis, but no difference of lateralization for the rest of the brain. Between-group comparisons revealed a left hypoactivation of Wernicke’s area at the posterior superior temporal/supramarginal junction during the responsive naming task, and a right hyperactivation encompassing the anterior insula with adjacent inferior frontal gyrus and the head of the caudate nucleus during the first phonological task. This study thus provides evidence that this subtype of specific language impairment is associated with atypical lateralization and functioning of core language areas [12].
4. New Results

4.1. Spatial Computing approach and RFIDs

Participants: Michel Banâtre, Paul Couderc [contact], Yann Glouche, Arnab Sinha.

In the line of our previous research in pervasive computing, we are working on spatial computing approaches in the context of RFID. Spatial computing consists in data structures and computing processes directly supported by physical objects. RFID is an attractive technology for supporting spatial computing, enabling any object to interact in a smart environment. Traditional RFID solutions use a logical model, where the RFID tags are simple identifiers referring to data in a remote information system. In our approach, we use the memory of the tags to build self-contained data structures and self-describing objects. While featuring interesting properties, such as autonomous operation and high scalability, this approach also raises difficult challenges: the memory capacity of the tags is very limited, requiring compact and efficient data structures. Some results have been achieved for security applications, where we contributed to efficient integrity checking solution for coupled objects. Integrity checking based on objects group can also be used to provide reliable inventory protocol for RFID, which current readers are lacking. A paper has just been submitted on this aspect.

An applicative project (see 5.2.1 ) in the context of domestic waste management is broadly investigating the use of RFID at item level to provide early waste sorting, to avoid incompatible mix of waste and to prevent hazards. An ontology based system has been proposed to determine the possible interactions of tagged products based on their properties and the external conditions.

4.2. Integrity checking with coupled objects

Participants: Michel Banâtre [contact], Paul Couderc, Jean-Francois Verdonck.

Integrity checking is an important concern in many activities, such as logistic, telecommunication or even day to day tasks such as checking for someone missing in a group. While the computing and telecommunication worlds commonly use digital integrity checking, many activities from the real world do not benefit from automatic integrity control mechanisms. RFID technology offers promising perspectives for facing this problem, but also raises strong privacy concerns as most of the RFID-based systems rely on global identification and tracking. Previously we have already designed Ubi-Check to provide an approach aiming at coupling physical objects and enabling integrity control built on local interactions, without the support of a global information system. Ubi-Check led to the development of various novel applications running quite on the same technology. Most of the partners showed a high interest in defining coupled object-based security solutions, but they were lacking the possibility of defining hierarchical couplings. This is that we have studied and implemented this year. We have designed the Ubi-Tree software which strives to deal with those new requirements.

4.2.1. Hierarchical physical object coupling

Ubi-Tree is a new solution/software designed at INRIA aiming at setting-up and reading hierarchical couplings. It relies on a structure in which physical objects (also called fragments) are seen as external nodes of a tree that we call coupling tree. External nodes of a tree are called leaves. In the system, internal nodes are called coupling nodes. Each fragment embeds an RFID tag supporting coupling data. Coupling data stores the coupling tree. Each internal node can be checked, which means a lacking, illegally forged or corrupted node can be detected at any depth of a coupling. Ubi-Tree proposes a new API to create and check hierarchical couplings and an interactive editing GUI is under development.
4.2.2. Coupling tree structure

New algorithms and structure to store and read hierarchical couplings trees in its leaves (RFID tags) have been developed, making it possible to use multiple coupling levels. Let’s take an example: given three physical objects $o_1$, $o_2$, $o_3$. A user can couple $o_1$ and $o_2$ together. Let $N_1$ be the parent coupling of $o_1$ and $o_2$: $N_1 = o_1$, $o_2$. Then he couples this coupling with $o_3$ to create the $N_2 = o_1$, $o_2$, $o_3$ coupling node. Figure 2 gives an illustration of the described coupling.

The key idea of the coupling structure is that coupling data are spread in a way that only descendant leaves of a coupling node are required to read it and process its integrity control. This way, $N_1$ only needs $o_1$ and $o_2$ to be read and / or checked as $o_1$ and $o_2$ are descendants from $N_1$ whereas $o_3$ is not. This choice enables to process integrity controls at multiple coupling levels. It is very convenient, as an example, if $o_3$ is physically separated from $o_1$ and $o_2$. So if $o_1$, $o_2$ and $o_3$ are brought into the field of the RFID reader, $N_1$ and $N_2$ can be written, read and checked. If $o_3$ is not present in its field, $N_2$ will not be recovered from the read data but $N_1$ can still be read and checked. If checked, $N_1$ will notify it was not the root of the coupling tree when the coupling tree was written. This way, the user knows it did not read the whole structure $o_1$ and $o_2$ are part of. Indeed, coupling nodes can have the following status:

- Valid: the set of detected tags enabled to decode a structure in which the node has the children and the parent it is supposed to have.
- Partial: same as the valid status except the node should have a parent that could not be read from detected tags.
- Incomplete: the node is missing some of its children.

4.2.3. Ongoing work

Today the management algorithms of graphs are located on the memory available RFID reading and writing. However, in the various applications envisaged, only a subset of RFID memory are read / write, others are only accessible in read-only. Currently we are working on the development of our algorithms that take into account this kind of configuration. The other problem we are working on today is the interface provided to users to be sure that the association between RFID tag and physical object is the one that is perceived by our coupling software. The idea is to be able to identify in the right way the RFID tag associated to a physical object when we place one physical object $O$ onto the support of the antenna linked to the RFID reader. The position of $O$, and the tag associated to $O$, in the physical space is determined using a camera coupled with an image recognition algorithm. The result is displayed onto a touch screen. In that way, when we want to couple a set of physical objects $o_1$, $o_2$, ... We place sequentially all these objects onto the support of the antenna, and
from the image of these objects displayed onto the touch screen we touch those we want to couple and activate the coupling operation. This solution is under development.

4.3. Pervasive support for Smart Homes  
**Participants:** Minh Tuan Ho, Michele Dominici, Bastien Pietropaoli, Frédéric Weis [contact].

Pervasive computing involves tight links between real world activities and computing process. While the perception of the real world events can be handled entirely by the application, we think that ad hoc approaches have limitations, in particular the complexity and the difficulty to re-use the code between applications. Instead, we promote the use of system level abstraction that leverage on tangible structures and processes. Important properties of this approach is that applications are, by design, operating in an implicit way ("in the background" of physical processes). They also often exhibits simpler architectures, and "natural" scalability in the sense that being build upon existing real-world process, they are strongly distributed design that relies essentially on local interactions between physical entities. We are applying this approach to "Smart Homes". A Smart Home is a residence equipped with computing and information technology devices conceived to collaborate in order to anticipate and respond to the needs of the occupants, working to promote their comfort, convenience, security and entertainment while preserving their natural interaction with the environment.

### 4.3.1. Definition of a system architecture

In a classical "logical" approach, all the intelligence of the Smart Home is condensed in a single entity that takes care of every device in the house. The sensors distributed in the environment have to send back all the gathered data to the central entity, that takes all the burden of parsing the sensitive information and infer the policies to be implemented. Our architecture is instead focused on a physical approach, where every device carries a part of the global intelligence: every single entity can analyze the part of information sensitive for its goal, derive useful data, and communicate meaningful information to the other devices.

![Four-layer model](Image)

**Figure 3. Four-layer model**

Our work is based on a four-layer model [8], as showed in Figure 3. The first layer of our system should be simply composed of sensors, but some constraints have to be fitted. In order to reduce the global system cost and to protect the inhabitants’ privacy, the number of sensors dispatched in the environment has to be reduced as much as possible. However, a huge number of different sensors are required to sense context pieces and redundancy can significantly increase the reliability of the sources. With this idea in mind, the sensors are grouped in nodes. These nodes are able to preprocess the data with simple computation such as minimum, maximum and average. They also enable the sensors to communicate, using, for instance, 6LowPAN (IPv6 over LoW Power wireless Area Networks).
In the second layer, the raw data are processed to obtain more abstract data about context and occurring situations. It could be, for instance, a presence in a room, the number of people in this room or the posture of someone. The aggregation of raw data is realized thanks to a data fusion algorithm. The one we adopted is called the belief functions theory or theory of evidence [6].

The bridge between the second and the third layer is realized integrating the results of sensor data fusion into a context model called *Context Spaces*. This model uses geometrical metaphors to describe context and situations, relying on the following concepts: the context attributes, the application space, the situation spaces and the context state. The context attributes are information types that are relevant and obtainable by the system; in our case, the context attribute values are provided by the perception layer, together with a degree of confidence on them, needed to cope with the intrinsic uncertainty of sensing systems in real world scenarios. In the situation and context identification layer, the context state provided by the perception layer is analyzed to infer the ongoing situation spaces (representing real-life situations) and also produce a measure of confidence in their occurrence. As the same context state can correspond to several different situation spaces (and vice versa), reasoning techniques are needed to discern the actual ongoing real-life situations in spite of uncertainty [5].

![Figure 4. System architecture](image)

### 4.3.2. Experimentation

The computations required by the second and the third layers to obtain abstract data and to analyze context and situations are too heavy for our nodes to be processed on. To remedy to this problem, more powerful nodes acting like sinks are used. These nodes are small “plug and play” computers called plug computers. Their role is to gather data from sensor nodes and to perform data fusion, required to produce the context attributes, and context space reasoning, used to identify ongoing situations.

The figure 4 gives an overview of our system architecture. The latter has been demonstrated by ACES team at EDF R&D in November.
6. New Results

6.1. Models and abstractions for distributed systems

This section summarizes the major results obtained by the ASAP team that relate to the foundations of distributed systems.

6.1.1. The weakest failure detector to implement a register in asynchronous systems with hybrid communication

Participants: Damien Imbs, Michel Raynal.

This work introduces an asynchronous crash-prone hybrid system model. The system is hybrid in the way the processes can communicate. On the one side, a process can send messages to any other process. On another side, the processes are partitioned into clusters and each cluster has its own read/write shared memory. In addition to the model, a main contribution of the work concerns the implementation of an atomic register in this system model. More precisely, a new failure detector (denoted $M\Sigma$) is introduced and it is shown that, when considering the information on failures needed to implement a register, this failure detector is the weakest. To that end, the work presents an $M\Sigma$-based algorithm that builds a register in the considered hybrid system model and shows that it is possible to extract $M\Sigma$ from any failure detector-based algorithm that implements a register in this model. The work also (a) shows that $M\Sigma$ is strictly weaker than $\Sigma$ (which is the weakest failure detector to implement a register in a classical message-passing system) and (b) presents a necessary and sufficient condition to implement $M\Sigma$ in a hybrid communication system.

This work has been published in SSS 2011 [38].

6.1.2. The universe of symmetry breaking tasks

Participants: Damien Imbs, Michel Raynal.

Processes in a concurrent system need to coordinate using a shared memory or a message-passing subsystem in order to solve agreement tasks such as, for example, consensus or set agreement. However, coordination is often needed to “break the symmetry” of processes that are initially in the same state, for example, to get exclusive access to a shared resource, to get distinct names or to elect a leader.

This work introduces and studies the family of generalized symmetry breaking (GSB) tasks, that includes election, renaming and many other symmetry breaking tasks. Differently from agreement tasks, a GSB task is “inputless”, in the sense that processes do not propose values; the task only specifies the symmetry breaking requirement, independently of the system’s initial state (where processes differ only on their identifiers). Among various results characterizing the family of GSB tasks, it is shown that (non adaptive) perfect renaming is universal for all GSB tasks.

This work was done in collaboration with Sergio Rajsbaum from the Universidad Nacional Autonoma de Mexico and was published in SIROCCO 2011 [36].

6.1.3. Read invisibility, virtual world consistency and probabilistic permissiveness are compatible

Participants: Tyler Crain, Damien Imbs, Michel Raynal.
The aim of a Software Transactional Memory (STM) is to discharge the programmers from the management of synchronization in multiprocess programs that access concurrent objects. To that end, an STM system provides the programmer with the concept of a transaction. The job of the programmer is to design each process the application is made up of as a sequence of transactions. A transaction is a piece of code that accesses concurrent objects, but contains no explicit synchronization statement. It is the job of the underlying STM system to provide the illusion that each transaction appears as being executed atomically. Of course, for efficiency, an STM system has to allow transactions to execute concurrently. Consequently, due to the underlying STM concurrency management, a transaction commits or aborts.

This work studies the relation between two STM properties (read invisibility and permissiveness) and two consistency conditions for STM systems, namely, opacity and virtual world consistency. Both conditions ensure that any transaction (be it a committed or an aborted transaction) reads values from a consistent global state, a noteworthy property if one wants to prevent abnormal behavior from concurrent transactions that behave correctly when executed alone. A read operation issued by a transaction is invisible if it does not entail shared memory modifications. This is an important property that favors efficiency and privacy. An STM system is permissive (respectively probabilistically permissive) with respect to a consistency condition if it accepts (respectively accepts with positive probability) every history that satisfies the condition. This is a crucial property as a permissive STM system that implements opacity while ensuring read invisibility. It then shows that read invisibility, probabilistic permissiveness and virtual world consistency are compatible. To that end the work describes a new STM protocol called IR_VWC_P. This protocol presents additional noteworthy features: it uses only base read/write objects and locks which are used only at commit time; and, in favorable circumstances, the cost of a read operation is $O(1)$. This work has been published in ICA3PP 2011 [29].

6.1.4. Towards a universal construction for transaction-based multiprocess programs

Participants: Tyler Crain, Damien Imbs, Michel Raynal.

The aim of a Software Transactional Memory (STM) system is to discharge the programmer from the explicit management of synchronization issues. The programmer’s job resides in the design of multiprocess programs in which processes are made up of transactions, each transaction being an atomic execution unit that accesses concurrent objects. The important point is that the programmer has to focus her/his efforts only on the parts of code which have to be atomic execution units without worrying on the way the corresponding synchronization has to be realized.

Non-trivial STM systems allow transactions to execute concurrently and rely on the notion of commit/abort of a transaction in order to solve their conflicts on the objects they access simultaneously. In some cases, the management of aborted transactions is left to the programmer. In other cases, the underlying system scheduler is appropriately modified or an underlying contention manager is used in order that each transaction be (“practically always” or with high probability) eventually committed.

This work paper proposed a deterministic STM system in which (1) every invocation of a transaction is executed exactly once and (2) the notion of commit/abort of a transaction remains unknown to the programmer. This system, which imposes restriction neither on the design of processes nor or their concurrency pattern, can be seen as a step towards the design of a deterministic universal construction to execute transaction-based multiprocess programs on top of a multiprocessor. Interestingly, the proposed construction is lock-free (in the sense that it uses no lock).

This work has been published in ICDCN 2012 [30].

6.1.5. A transaction friendly binary search tree

Participants: Tyler Crain, Michel Raynal.
Transactions, which provide optimistic synchronization by avoiding the use of blocking, greatly simplify multithreaded programming. In fact, the programmer has to encapsulate sequential operations or existing critical sections into transactions to obtain a safe concurrent program. Programmers have thus started evaluating transactional memory using data structures originally designed for pessimistic (i.e., non-optimistic) synchronization, whose prominent example is the red-black tree library developed by Oracle Labs that is part of STAMP and microbenchmark distributions. Unfortunately, existing data structures are badly suited for optimistic synchronization as they rely on strong structural invariants, like logarithmic tree depth, to bound the step complexity of pessimistically synchronized accesses. By contrast, this complexity does not apply to optimistically synchronized accesses thus making the invariants overly conservative. More dramatically, guaranteeing such invariants tends to increase the probability of aborting and restarting the same access before it completes. We introduced a concurrent binary search tree that breaks transiently its balance structural invariants for efficiency, a property we call transaction-friendly. This new tree outperforms the existing transaction-based version of the AVL and the red-black trees. Its key novelty stems from the decoupling of update operations: they are split into one transaction that modifies the abstraction state and multiple ones that restructure its tree implementation. The resulting transaction-friendly library trades aborts for few additional access steps and, in particular, it speeds up a transaction-based travel reservation application by up to 3.5X. This work was done in collaboration with Vincent Gramoli from EPFL Lausanne, and is described in [52].

### 6.1.6. Relations linking failure detectors associated with k-set agreement in message-passing systems

**Participants:** Achour Mostefaoui, Michel Raynal, Julien Stainer.

The k-set agreement problem is a coordination problem where each process is assumed to propose a value and each process that does not crash has to decide a value such that each decided value is a proposed value and at most k different values are decided. While it can always be solved in synchronous systems, k-set agreement has no solution in asynchronous send/receive message-passing systems where up to t ≥ k processes may crash.

A failure detector is a distributed oracle that provides processes with additional information related to failed processes and can consequently be used to enrich the computability power of asynchronous send/receive message-passing systems. Several failure detectors have been proposed to circumvent the impossibility of k-set agreement in pure asynchronous send/receive message-passing systems. Considering three of them (namely, the generalized quorum failure detector Σk, the generalized loneliness failure detector Lk, and the generalized eventual leader failure detector Ωk) this work investigates their computability power and the relations that link them. There are three main contributions: (a) it shows that the failure detector Ωk and the eventual version of Lk have the same computational power; (b) it shows that Lk is realistic if and only if k ≥ n/2; and (c) it gives an exact characterization of the difference between Lk (that is too strong for k-set agreement) and Ωk (that is too weak for k-set agreement). This work was published at SSS 2011 [45].

### 6.1.7. The price of anonymity: optimal consensus despite asynchrony, crash and anonymity

**Participant:** Michel Raynal.

This work [23], done in collaboration with François Bonnet, from JAIST, Japan, addresses the consensus problem in asynchronous systems prone to process crashes, where additionally the processes are anonymous (they cannot be distinguished one from the other: they have no name and execute the same code). To circumvent the three computational adversaries (asynchrony, failures and anonymity) each process is provided with a failure detector of a class denoted ψ, that gives it an upper bound on the number of processes that are currently alive (in a non-anonymous system, the classes ψ and P-the class of perfect failure detectors- are equivalent).

The first part presents a simple ψ-based consensus algorithm where the processes decide in 2t + 1 asynchronous rounds (where t is an upper bound on the number of faulty processes). It then shows one of its main results, namely, 2t + 1 is a lower bound for consensus in the anonymous systems equipped with ψ. The second contribution addresses early-decision. The paper presents and proves correct an early-deciding algorithm where the processes decide in min(2f + 2, 2t + 1) asynchronous rounds (where f is the actual number of faulty processes).

...
of process failures). This leads to think that anonymity doubles the cost (wrt synchronous systems) and it is conjectured that min(2f + 2, 2t + 1) is the corresponding lower bound.

The work finally considers the k-set agreement problem in anonymous systems. It first shows that the previous ψ-based consensus algorithm solves the k-set agreement problem in \( R_t = 2 \left\lceil \frac{t}{k} \right\rceil + 1 \) asynchronous rounds. Then, considering a family of failure detector classes \( \{\psi_t\}_{0 \leq t < k} \) that generalizes the class \( \psi(= \psi_0) \), the paper presents an algorithm that solves the k-set agreement in \( R_{t, k} = 2 \left\lceil \frac{1}{k-t} \right\rceil + 1 \) asynchronous rounds. This last formula relates the cost \( (R_{t, k}) \), the coordination degree of the problem \( (k) \), the maximum number of failures \( (t) \) and the strength \( (\ell) \) of the underlying failure detector.

### 6.1.8. On the road to the Weakest Failure Detector for k-Set Agreement in Message-passing Systems

**Participant:** Michel Raynal.

In the k-set agreement problem, each process (in a set of \( n \) processes) proposes a value and has to decide a proposed value in such a way that at most \( k \) different values are decided. While this problem can easily be solved in asynchronous systems prone to \( t \) process crashes when \( k > t \), it cannot be solved when \( k \leq t \). Since several years, the failure detector-based approach has been investigated to circumvent this impossibility. While the weakest failure detector class to solve the k-set agreement problem is unsolvable in read/write shared-memory systems has recently been discovered (PODC 2009), the situation is different in message-passing systems where the weakest failure detector classes are known only for the extreme cases \( k = 1 \) (consensus) and \( k = n - 1 \) (set agreement).

This work \([22]\), done in collaboration with François Bonnet, from JAIST, Japan, has four contributions whose aim is to help pave the way to discover the weakest failure detector class for k-set agreement in message-passing systems. These contributions are the following. (a) The first is a new failure detector class, denoted \( \Pi_k \), that is such that \( \Pi_1 = \Sigma \times \Omega \) (the weakest class for \( k = 1 \)), and \( \Pi_{n-1} = \mathcal{L} \) (the weakest class for \( k = n - 1 \)). (b) The second is an investigation of the structure of \( \Pi_k \) that shows that \( \Pi_k \) is the combination of two failures detector classes \( \Sigma_k \) (that is new) and \( \Omega_k \) (they generalize the previous “quorums” and “eventual leaders” failure detectors classes, respectively). (c) The third contribution concerns \( \Sigma_k \) that is shown to be necessary requirement (as far as information on failure is concerned) to solve the k-set agreement problem in message-passing systems. (d) Finally, the last contribution is a \( \Pi_{n-1} \)-based algorithm that solves the \((n - 1)\)-set agreement problem. This algorithm provides us with a new algorithmic insight on the way the \((n - 1)\)-set agreement problem can be solved in asynchronous message-passing systems. It is hoped that these contributions will help discover the weakest failure detector class for k-set agreement in message-passing systems.

### 6.1.9. A non-topological proof for the impossibility of k-set agreement.

**Participant:** Armando Castañeda.

This work was done in collaboration with Hagit Attiya, from Technion, Haifa, Israel. In the k-set agreement task each process proposes a value, and each correct process has to decide a value which was proposed, so that at most \( k \) distinct values are decided. Using topological arguments it has been proved that k-set agreement is unsolvable in the asynchronous wait-free read/write shared memory model, when \( k < n \), the number of processes.

This work \([34]\) focuses on a simple, non-topological impossibility proof of k-set agreement. The proof depends on two simple properties of the immediate snapshot executions, a subset of all possible executions, and on the well known handshaking lemma stating that every graph has an even number of vertices with odd degree.

The paper was presented in the 13th Int’l Symposium on Stabilization, Safety, and Security of Distributed Systems (SSS’11) in Grenoble, France. The journal version of the paper was submitted to Theoretical Computer Science.
6.1.10. Enriching the reduction map of sub-consensus tasks

Participants: Armando Castañeda, Damien Imbs, Michel Raynal.

This work [51] was done in collaboration with Sergio Rajsbaum from the Universidad Nacional Autonoma de Mexico.

Understanding the relative computability power of tasks, in the presence of asynchrony and failures, is a central concern of distributed computing theory. In the wait-free case, where the system consists of n processes and any of them can fail by crashing, substantial attention has been devoted to understanding the relative power of the subconsensus family of tasks, which are too weak to solve consensus for two processes. The first major results showed that set agreement and renaming (except for some particular values of n) cannot be solved wait-free in read/write memory. Then it was proved that renaming is strictly weaker than set agreement (when n is odd).

This work considers a natural family of subconsensus tasks that includes set agreement, renaming and other generalized symmetry breaking (GSB) tasks. It extends previous results, and proves various new results about when there is a reduction and when not, among these tasks. Among other results, the work shows that there are incomparable subconsensus tasks.

6.1.11. Byzantine Consensus Decidability

Participants: Achour Mostefaoui, Michel Raynal.

Solving the consensus problem requires in one way or another that the underlying system satisfies synchrony assumptions. Considering a system of n processes where up to t < n/3 may commit Byzantine failures, we proposed in [26] a necessary and sufficient synchrony assumption to solve consensus.

Such a condition is formulated with the notions of a symmetric synchrony property and property ambiguity. A symmetric synchrony property is a set of graphs, where each graph corresponds to a set of bi-directional eventually synchronous links among correct processes. Intuitively, a property is ambiguous if it contains a graph whose connected components are such that it is impossible to distinguish a connected component that contains correct processes only from a connected component that contains faulty processes only. The paper connects then the notion of a symmetric synchrony property with the notion of eventual bi-source, and shows that the existence of a virtual $\Diamond[t+1]$ bi-source is a necessary and sufficient condition to solve consensus in presence of up to t Byzantine processes in systems with bi-directional links and message authentication. Finding necessary and sufficient synchrony conditions when links are timely in one direction only, or when processes cannot sign messages, still remains open (and very challenging) problems.

6.1.12. Solving $k$-set agreement in message-passing systems

Participants: Achour Mostefaoui, Michel Raynal, Julien Stainer.

The $k$-set agreement problem is a coordination problem where each process is assumed to propose a value and each process that does not crash has to decide a value such that each decided value is a proposed value and at most $k$ different values are decided. While it can always be solved in synchronous systems, $k$-set agreement has no solution in asynchronous send/receive message-passing systems where up to $t \geq k$ processes may crash.

A failure detector is a distributed oracle that provides processes with additional information related to failed processes and can consequently be used to enrich the computability power of asynchronous send/receive message-passing systems. Several failure detectors have been proposed to circumvent the impossibility of $k$-set agreement in pure asynchronous send/receive message-passing systems. Considering three of them (namely, the generalized quorum failure detector $\Sigma_k$, the generalized loneliness failure detector $L_k$ and the generalized eventual leader failure detector $\Omega_k$), we investigated their computability power and the relations that link them in [45]. It has three mains contributions: (a) it shows that the failure detector $\Omega_k$ and the eventual version of $L_k$ have the same computational power; (b) it shows that $L_k$ is realistic if and only if $k \geq n/2$; and (c) it gives an exact characterization of the difference between $L_k$ (that is too strong for $k$-set agreement) and $\Sigma_k$ (that is too weak for $k$-set agreement).
6.1.13. Efficient Implementations of Concurrent Objects

Participants: Achour Mostefaoui, Michel Raynal.

As introduced by Taubenfeld, a contention-sensitive implementation of a concurrent object is an implementation such that the overhead introduced by locking is eliminated in the common cases, i.e., when there is no contention or when the operations accessing concurrently the object are non-interfering. In [44], we present a methodological construction of a contention-sensitive implementation of a concurrent stack. In a contention-free context a push or pop operation does not rest on a lock mechanism and needs only six accesses to the shared memory. In case of concurrency a single lock is required. Moreover, the implementation is starvation-free (any operation is eventually executed). The paper, that presents the algorithms in an incremental way, visits also a family of liveness conditions and important concurrency-related concepts such as the notion of an abortable object.

6.2. Large-scale and user-centric distributed system

This section summarizes the major results obtained by the team in 2011 in the context of large-scale distributed systems and social networks. This includes the results obtained within the GOSSPLE ERC project, which encompass two types of social networks: explicit and implicit.

**Explicit networks** connect users based on explicit social connections. In FACEBOOK or MYSPACE, users issue and accept friendship requests. In TWITTER, they decide that they wish to follow the tweets of specific users. In all cases, the topology of the resulting network reflects the choices of users and often consists of links that already exist between real people. Explicit networks are therefore very useful in reinforcing and exploiting existing connections but provide little support for discovering new content.

**Implicit networks** complement explicit ones by providing each user with a set of anonymous acquaintances that share similar interests, that visit similar websites or that have otherwise similar profiles. Different from explicit networks, implicit ones are naturally suited to support the discovery of new content. In previous work [1], we exploited this network to improve web navigation. In the following, we consider additional applications encompassing news dissemination, online transactions, and recommendation.

6.2.1. WhatsUp: P2P news recommender

Participants: Antoine Boutet, Davide Frey, Anne-Marie Kermarrec.

The main application in the context of GOSSPLE is WhatsUp, an instant news system designed for a large-scale network with no central authority. WhatsUp builds an implicit social network based on the opinions users express about the news items they receive (like-dislike). This is achieved through an obfuscation mechanism that does not require users to ever reveal their exact profiles. WhatsUp disseminates news items through a novel heterogeneous gossip protocol that biases the choice of its targets towards those with similar interests and amplifies dissemination based on the level of interest in every news item. WhatsUp outperforms various alternatives in terms of accurate and complete delivery of relevant news items while preserving the fundamental advantages of standard gossip: namely simplicity of deployment and robustness. This work has been carried out in collaboration with Rachid Guerraoui from EPFL and was demonstrated during the different local events.

6.2.2. Personalized top-k processing

Participant: Anne-Marie Kermarrec.

Another way to improve the experience of users on the web is to personalize top-k queries. In collaboration with Xiao Bai and Vincent Leroy from Yahoo! Research in Barcelona and Rachid Guerraoui from EPFL Lausanne we, therefore, introduced P4Q, a fully decentralized gossip-based protocol to personalize query processing in social tagging systems. P4Q dynamically associates each user with social acquaintances sharing similar tagging behaviors. Queries are gossiped among such acquaintances, computed on the fly in a collaborative, yet partitioned manner, and results are iteratively refined and returned to the querier. Analytical and experimental evaluations convey the scalability of P4Q for top-k query processing, as well its inherent ability to cope with users updating profiles and departing. The work appeared in the ACM transactions of database systems [12].
6.2.3. Social Market

Participants: Davide Frey, Arnaud Jegou, Anne-Marie Kermarrec.

The ability to identify people that share one’s own interests is one of the most interesting promises of the Web 2.0 driving user-centric applications such as recommendation systems or collaborative marketplaces. To be truly useful, however, information about other users also needs to be associated with some notion of trust. Consider a user wishing to sell a concert ticket. Not only must she find someone who is interested in the concert, but she must also make sure she can trust this person to pay for it. Social Market (SM) solve this problem by allowing users to identify and build connections to other users that can provide interesting goods or information and that are also reachable through a trusted path on an explicit social network like Facebook. This convergence of implicit and explicit networks yields TAPS, a novel gossip protocol that can be applied in applications devoted to commercial transactions, or to add robustness to standard gossip applications like dissemination or recommendation systems.

This work has been published at SSS 2011 [ 33 ], and an extended version bringing better performances and strong privacy guarantees have recently been submitted for publication.

6.2.4. Member classification and party characteristics in Twitter

Participant: Antoine Boutet.

In modern politics, parties and individual candidates must have an online presence and usually have dedicated social media coordinators. In this context, real time member classification and party characterization, taking into account the dynamic nature of social media, are essential to highlight the main differences between parties and to monitor their activities, influences, structures, contents and mood. This work [ 53 ] was done in collaboration with E. Yoneki from Computer Lab, Cambridge, UK.

6.2.5. Graph Drawing and Visual Recommendations

Participants: Anne-Marie Kermarrec, Afshin Moin.

An important aspect of social network is their graph structure. In a collaboration with Vincent Leroy (Yahoo! Research) and Gilles Tredan (TU Berlin) [ 41 ], we started from this structure to propose a decentralized gossip-based algorithm called SoCS (Social Coordinate Systems). SoCS achieves efficient distributed social graph embedding using a force-based graph embedding technique to extract communities from a graph. SoCS (i) scales to large dynamic graph, aggregating the computing power of individual nodes and, (ii) avoids a central entity controlling users sensitive data such as relations and preferences. We evaluated SoCS using two different force-based models and compare them in the context of a generated Kleinberg small-world topology. More specifically, we showed that the SoCS graph embedding enables to clearly distinguish between short and long-range links. We also evaluate SoCS against a real DBLP data set, showing that removed links are correctly predicted.

Graph structures are also at the basis of our work on energy/force-based models for graph visualization. We applied visualization both to social network and in the context of recommendation systems. In particular we are working on an SVD-like algorithm for drawing precise 2-dimensional visual recommendations based on Principal Component Analysis (PCA) and Curvilinear Component Analysis (CCA).

6.2.6. Private Similarity Computation in Distributed Systems: from Cryptography to Differential Privacy

Participants: Mohammad Alaggan, Anne-Marie Kermarrec.

The use of personal data in the context of social networks raises important concerns about privacy. In a collaboration [ 24 ] with Sébastien Gambs from the CIDre team, we addressed the problem of computing the similarity between two users (a key operation in an implicit social network [ 1 ]) while preserving their privacy in a fully decentralized system and for the passive adversary model. First, we introduced a two-party protocol for privately computing a threshold version of the similarity and applied it to well-known similarity measures such as the scalar product and the cosine similarity. The output of this protocol is only one bit
of information telling whether or not two users are similar beyond a predetermined threshold. Afterwards, we explored the computation of the exact and threshold similarity within the context of differential privacy. Differential privacy is a recent notion developed within the field of private data analysis guaranteeing that an adversary that observes the output of the differentially private mechanism, will only gain a negligible advantage (up to a privacy parameter) from the presence (or absence) of a particular item in the profile of a user. This provides a strong privacy guarantee that holds independently of the auxiliary knowledge that the adversary might have. More specifically, we designed several differentially private variants of the exact and threshold protocols that rely on the addition of random noise tailored to the sensitivity of the considered similarity measure. We also analyzed their complexity as well as their impact on the utility of the resulting similarity measure.

6.2.7. Constellation: Programming decentralized social networks

**Participants:** François Taiani, Anne-Marie Kermarrec.

As they continue to grow, social and collaborative applications (e.g. twitter, Facebook, digg) are increasingly calling for disruptive distributed solutions than can cater for the millions of users these applications serve daily, in hundreds of countries, over a wide variety of devices. To address these challenges, fully decentralized versions of social and collaborative applications are progressively emerging that seek to provide naturally scalable solutions to deliver their services. Gossip protocols in particular appear as a natural solution to implement these decentralized versions, as they intrinsically tend to be highly resilient, efficient, and scalable.

Social applications based on gossip have however been limited so far to relatively homogeneous systems: They typically rely on one similarity measure to self-organize large amount of distributed users in implicit communities, and thus offer powerful means to search, mine, and serve personalized data in a distributed manner.

We posit in this work [54] that we now need to move to more complex gossip-based social applications that can cater for different types of data and similarity, organized in multiple levels of abstraction. Exploring, designing, and evaluating such novel approaches is unfortunately time-consuming and error-prone. To help in this task, we have started to design a new programming language, Constellation, that seeks to simplify the realization and experimentation with social gossip-based applications. Constellation is based on two central observations: (i) future decentralized social applications will need to handle heterogeneous forms of data and self-organization, and (ii) to offer more powerful services, these applications will need to move beyond physical nodes to encompass richer data structures organized in virtualized levels of abstractions.

6.2.8. Leveraging content interconnections for efficient data storage.

**Participants:** Anne-Marie Kermarrec, Konstantinos Kloudas, François Taiani.

Traffic generated by User Generated Content (UGC) sharing sites, such as **YOUTUBE**, accounts for a substantial fraction of today’s global Internet load. This success has however brought a number of key technical challenges, crucial for system sustainability and user experience. One of them is the need to place content close to consumers, so that user perceived latency is reduced and bandwidth utilization is minimized. In a joint work with Kevin Huguenin, we try to tackle this problem by leveraging the fact that content hosted by these sites is interconnected, forming a content graph that as shown by former works, has an important impact on a file’s view pattern. In our work titled “Recommended nearby in UGC delivery networks: leveraging geographical and content locality”, we focused on **YOUTUBE** and we studied how two types of locality previously analyzed in isolation in UGC systems, namely content locality (a.k.a graph locality, induced by the related video feature) and geographic locality, are in fact correlated. Leveraging the above finding, we proposed a novel algorithm for replica placement that tries to predict where future views for a video will come from based on the video’s related videos and places its replicas accordingly. This work has been submitted for publication.

6.2.9. Transparent Componentization: High-level (Re)configurable Programming for Evolving Distributed Systems

**Participants:** François Taiani, Marin Bertier, Anne-Marie Kermarrec.
This work was done in collaboration Component frameworks and high-level distributed languages have been widely used to develop distributed systems, and provide complementary advantages: Whereas component frameworks foster composability, reusability, and (re)configurability; distributed languages focus on behavior, simplicity and programmability. We argue that both types of approach should be brought together to help develop complex adaptive systems, and we propose an approach to combines both technologies without compromising on any of their benefits. Our approach, termed Transparent Componentization [43], automatically maps a high-level distributed specification onto a underlying component framework. It thus allows developers to focus on the programmatic description of a distributed system’s behavior, while retaining the benefits of a component architecture. As a proof of concept, we present WhispersKit, a programming environment for gossip-based distributed systems. Our evaluation shows that WhispersKit successfully retains the simplicity and understandability of high-level distributed language while providing efficient and transparent reconfigurability thanks to its component underpinnings.

6.2.10. Efficient peer-to-peer backup services through buffering at the edge  
**Participants:** Anne-Marie Kermarrec, Alexandre Van Kempen.

The availability of end devices of peer-to-peer storage and backup systems has been shown critical for usability and for system reliability in practice. This has led to the adoption of hybrid architectures composed of both peers and servers. Such architectures mask the instability of peers thus approaching the performances of client-server systems while providing scalability at a low cost. In this work [31] - done in collaboration with Erwan Le Merrer, Serge Defrance, Nicolas Le Scouarne and Gilles Straub from Technicolor, Rennes, France - we advocate the replacement of such servers by a cloud of residential gateways, as they are already present in users’ homes, thus pushing the required stable components at the edge of the network. In our gateway-assisted system, gateways act as buffers between peers, compensating for their intrinsic instability. This enables to offload backup tasks quickly from the user’s machine to the gateway, while significantly lowering the retrieval time of backed up data. We evaluate our proposal using real world traces including existing traces from Skype and Jabber as well as a trace of residential gateways for bandwidth. Results show that the time required to backup data in the network is comparable to a server-assisted approach, while substantially improving the time to restore data, which drops from a few days to a few hours. As gateways are becoming increasingly powerful in order to enable new services, we expect such a proposal to be leveraged on a short term basis.

6.2.11. Commutative Replicated Data Type for Semantic Stores  
**Participant:** Stéphane Weiss.

This work has been done in collaboration with Khaled Aslan (Université de Nantes - Lina), Pascal Molli (Université de Nantes - Lina) and Hala Skaf-Molli (Université de Nantes - Lina).

Web 2.0 tools are currently evolving to embrace semantic web technologies. Blogs, CMS, Wikis, social networks and real-time notifications, integrate ways to provide semantic annotations and therefore contribute to the linked data and more generally to the semantic web vision. This evolution generates a lot of semantic datasets of different qualities, different trust levels and partially replicated. This raises the issue of managing the consistency among these replicas. This issue is challenging because semantic data-spaces can be very large, they can be managed by autonomous participants and the number of replicas is unknown. A new class of algorithms called Commutative Replicated Data Type are emerging for ensuring eventual consistency of highly dynamic content on P2P networks. We define C-Set [25] a CRDT specifically designed to be integrated in Triple-stores. C-Set allows efficient P2P synchronization of an arbitrary number of autonomous semantic stores.

6.2.12. Building large scale platform for chemical program  
**Participants:** Marin Bertier, Achour Mostefaoui.

This work [28] was done in collaboration with the Myriads project team.
Chemical programming is a promising paradigm to design autonomic systems. Within such a paradigm, computations can be seen as chemical reactions controlled by a set of chemical rules. In other words, data are molecules of a chemical solutions, reacting together to produce new data. Reactions take place in an implicitly parallel, and autonomic fashion.

Our objective was to design a distributed chemical platform bringing such concepts. This platform should be adapted to large scale distributed system to benefit at his best the inherent distribution of chemical program.
6. New Results

6.1. Aspects

Participants: Rémi Douence, Abdelhakim Hannousse, Ismael Mejía, Jacques Noyé, Angel Núñez, Nicolas Tabareau, Mario Südholt.

We have provided results on three subject matters in the domain of aspect-oriented software development: the foundations of aspects, aspect languages and their applications, as well as the use of aspects for the manipulation of distributed systems.

As a side note on the form of this document, the reader should be aware that much of our work reported in this section is not exclusive to AOSD but also contributes to software composition issues in a larger sense. This applies, in particular, to the results cited in the subsections on aspect languages and distributed aspects.

6.1.1. Foundations of Aspects

In the domain of foundations of AOSD, we have mainly provided new results on the preservation of formal correctness properties in the presence of aspects and how AOP can be modeled using category theory.

- **Property preservation in the presence of aspects.** In general aspects can arbitrarily change the semantics of the base program. We have identified categories of aspects that preserve class of properties of the base programs [17]. This approach makes it possible to prove once and for all that a category of aspects preserve a class of properties. The categories are defined with respect to the semantics of the woven program as well as with restricted aspect languages. In this latter case, languages are defined by grammars hence checking for property preservation boils down to a syntactic check for aspects. Classes of properties are defined using a subset of temporal logic both for sequential and concurrent programs. Our framework is abstract in that it does not depend on the actual programming language but only on conditions on its small step semantics.

- **AOP and category theory.** Aspect-Oriented Programming (AOP) started ten years ago with the remark that modularization of so-called crosscutting functionalities is a fundamental problem for the engineering of large-scale applications. Originating at Xerox PARC, this observation has sparked the development of a new style of programming that is gradually gaining traction. However, AOP lacks theoretical foundations to clarify new ideas showing up in its wake. We have proposed to put a bridge between AOP and the notion of 2-category to enhance the conceptual understanding of AOP [36], [46]. Starting from the connection between the \(\lambda\)-calculus and the theory of categories, we have provided an internal language for 2-categories and shown how it can be used to define the first categorical semantics for a realistic functional AOP language, called MinAML.

  We have later taken advantage of this new categorical framework to introduce the notion of computational 2-monads for AOP [37]. We have illustrated their conceptual power by defining a 2-monad for Éric Tanter’s execution levels—which constitutes the first algebraic semantics for execution levels—and then introducing the first exception monad transformer specific to AOP that gives rise to a non-flat semantics for exceptions by taking levels into account.

- **Membranes for AOP.** AOP aims to enhance modularity and reusability in software systems offering an abstraction mechanism to deal with crosscutting concerns. However, in most aspect-oriented languages aspects have a global view of computation that actually introduces a strong coupling between advised code and aspect code, hampering modularity. Proposals that address this problem can be classified in two categories: the first one focuses on controlling aspect scoping, i.e. the visibility of join points to aspects, while the second one focuses on protecting software units from aspects. As a new approach, we have proposed programmable membranes (inspired by the work of Boudol for distributed processes) to control aspect influence over software systems as a uniform framework that unifies and extends previous approaches [49].
• **Aspects and invertible program restructurations.** When one chooses a main axis of structural decomposition of a problem, the other axes become secondary. This is known as the tyranny of the dominant decomposition and this hinders program maintenance of secondary concerns. We propose to use automatic program transformations built on top of refactorers in order to solve this tyranny issue [47]. This offers a new approach to the expression problem by always providing the right view to the programmers.

### 6.1.2. Aspect Languages

This year we have provided major results on the integration of programming features for objects, events and aspects, and aspect execution infrastructures. Furthermore, we have investigated the application of aspect languages to the Java security model, component-based software development and software product lines.

• **Integrating OOP, AOP and EBP:** Object-Oriented Programming (OOP) has become the de facto programming paradigm. Event-Based Programming (EBP) and Aspect-Oriented Programming (AOP) complement OOP, covering some of its deficiencies when building complex software. Today’s applications combine the three paradigms. However, OOP, EBP and AOP have not yet been properly integrated. Their underlying concepts are in general provided as distinct language constructs, whereas they are not completely orthogonal. This lack of integration and orthogonality complicates the development of software as it reduces its understandability, its composability and increases the required glue code. [16] proposes an integration of OOP, EBP and AOP leading to a simple and regular programming model. This model integrates the notions of class and aspect, the notions of event and join point, and the notions of piece of advice, method and event handler. It reduces the number of language constructs while keeping expressiveness and offering additional programming options. ECaesarJ had previously been developed based on these ideas. [16] introduces a simpler variant of it, EJava, a plain extension of Java. [24] shows that these ideas can also be easily applied to Scala, leading to EScala, and focuses on the idea that the declarative events provided by the model can be seen as object-oriented events, which, unlike global typed events, obey to the basic OO principles (OO modular reasoning, encapsulation and late-binding). An efficient implementation of EScala is described, based on the idea that an event should not be produced in the absence of at least a consumer. This is equivalent to what could be programmed by hand using variants of the observer pattern, except that all the necessary scaffolding is provided by the language compiler and runtime rather than by the programmer.

• **Prototyping and composing Aspect Languages:** [11] presents CALI (Common Aspect Languages Interpreter), a framework for rapid prototyping and composition of aspect languages based on interpreters. In practice the common interpreter is actually a thin interpretative layer built on top of Java and implemented as an AspectJ aspect. This interpreter implements common aspect mechanisms and leaves holes to be defined when developing concrete languages. The approach has been validated by implementing prototypes of significant subsets of well-known general-purpose and domain-specific aspect languages (AspectJ, EAOP, COOL and a couple of other small domain-specific aspect languages) and exploring variants of AspectJ. Languages implemented with CALI, for instance AspectJ and COOL, can then be easily composed.

• **Application to security:** It is inevitable that some concerns crosscut a sizeable application, resulting in code scattering and tangling. This issue is particularly severe for security-related concerns: it is difficult to be confident about the security of an application when the implementation of its security-related concerns is scattered all over the code and tangled with other concerns, making global reasoning about security precarious. In [21], we consider the case of access control in Java, which turns out to be a crosscutting concern with a non-modular implementation based on runtime stack inspection. We describe the process of modularizing access control in Java by means of AOP. We first show a solution based on AspectJ that must rely on a separate automata infrastructure. We then put forward a novel solution via dynamic deployment of aspects and scoping strategies. Both solutions, apart from providing a modular specification of access control, make it possible to easily express other useful policies such as the Chinese wall policy. However, relying on expressive scope
control results in a compact implementation, which, at the same time, permits the straightforward expression of even more interesting policies.

- **Aspects and software components**: The relationship of aspects and components, as well as their integration as part of real-world infrastructures and application is still subject to many open issues.

  We have studied, in particular, how crosscutting concerns naturally appear when several architectural views must be considered at the same time. In this context, we have proposed a domain-specific language to specify these architectures. We have proposed an implementation of composable controllers in Fractal as well as composition operators that makes it possible to solve aspects interferences [25]. We have also shown how to formally model the implementation in Uppaal in order to statically check aspects interference [26]. This work has principally been undertaken in the context of the PhD thesis of Abdelhakim Hannouse [12] that has been defended in Sep. 2011.

  Another fundamental issue that relates components and aspects is the question whether behavioral protocols can be used to concisely define crosscutting concerns of component-based systems. Furthermore, the protocols should then be instrumental to enable (automatic) reasoning about composition properties of component systems. As part of her PhD thesis (defended in Oct. 2011) [14], Dong Ha Nguyen has defined a notion of aspects that allow crosscutting concerns to be expressed in terms of a class of non-regular behavioral protocols, so-called visibly-pushdown languages. She has developed, most notably, a constructive way of building correct component-based systems that respect such aspect-modified behavioral protocols. Furthermore, she has shown this year how composition properties can generally be verified using model checking techniques.

- **Aspects and software product lines**: In [41], we take a closer look at the difficulties of feature-oriented modularisation of product lines and demonstrate, using a case study in the domain of home automation, how a better modularisation can be achieved with the ECaesarJ programming language, through a type-safe and stable decomposition of a broad spectrum of software abstractions: classes, methods, events, and state machines, based on late binding and mixin composition. A nice property of this modularisation is that it directly captures, at the implementation level, the high-level decomposition in features, without requiring the user to resort on complex transformation technology (the technology is embedded in the compiler).

  Furthermore, ASCOLA members have co-edited and contributed to a comprehensive book on aspects, model-driven engineering and product lines [42] (see Sec. 6.2 for details).

### 6.1.3. Aspects for Distributed Systems

In the field of aspects for distributed systems, we have mainly extended the AWED model for distributed aspects (see Sec. 5.1) in order to define and implement a gray-box distributed composition model. We have also worked on a notion of distributed crosscuts that enable causality relationships to be captured using logical clocks.

- **Gray-box distributed composition.** Existing composition and coordination techniques for distributed applications typically support only interface-level (black-box) composition or arbitrary access to the implementation (gray-box or white-box composition). In [20], we have presented a structured approach to the composition of complex distributed software systems that require invasive modifications. Concretely, we have provided three contributions:
  
  - We have introduced a small kernel composition language for structured gray-box composition using invasive distributed patterns.
  
  - We have motivated that gray-box composition approaches should be defined and evaluated in terms of the flexibility and control they provide; a notion of degrees of invasiveness has been introduced to help assess this trade-off.
  
  - We have applied and evaluated it in the context of several studies involving two real-world software systems: benchmarking of grid algorithms with NASGrid and transactional replication with JBoss Cache. As a main result, we have shown that gray-box composition using
invasive distributed patterns allows the declarative and modular definition of evolutions of real-world applications that need moderate to high degrees of invasive modifications. We have then provided a first framework-based implementation based on our AWED tool and applied it to evolution tasks in OpenMRS, a health information system [31]. The composition framework supports the definition of expressive pattern-based invasive compositions based on the skeleton paradigm for distributed patterns. Concretely, we have shown that the composition framework enables the concise definition of an evolution scenario of OpenMRS that supports the consolidation of patient data from different distributed instances.

Finally, we have studied a model that integrates distributed aspects and actors and thus improves on the robustness properties of existing models for distributed aspects [32].

- Causal aspects for distributed applications. In [45], we have applied logical clocks à la Lamport in order to define causality between distributed join points. This enables us to declaratively define pointcuts in a distributed context such as web-based applications in JavaScript. The definition of advice can then be simplified, because they are executed only when necessary and do not have to maintain and check crosscutting information.

### 6.2. Software Composition

**Participants:** Jean-Claude Royer, Hervé Grall, Christine Louberry, Mario Südholt.

ASCOLA’s work on software composition addresses the foundations of software composition methods, the definition and implementation of concrete composition techniques and their application to different functionalities and application domains. This year we have contributed to the questions of how to identify components in legacy software, how to effectively deploy and reconfigure component-based software in pervasive environments, and how to apply software composition techniques to Cloud security as well as software product lines.

- **Component identification.** The communication integrity property is one of the major principles to implement software architectures, however, there is a lack of tooling for assessing the quality of components codes. To cope with this issue, we defined, in [23], a Java component model and a tool for identifying component types based on the communication integrity property. We apply it to several case studies and compare the result with the SOMOX component recovery tool.

- **QoS-driven deployment and reconfiguration of pervasive applications.** [28] presents the Kalimucho platform, a platform for the dynamic reconfiguration of applications on mobiles and constraints devices. First this article focuses on the heuristics implemented by Kalimucho. They support finding a configuration and a deployment matching two criteria: utility and durability. Moreover, we present a case study to experiment the approach of Kalimucho. It confirms that Kalimucho provides a satisfactory execution time for mobile devices.

A second result we have contributed in this context is a two-dimensional QoS model for Kalimucho that supports the QoS-driven deployment of mobile applications [29]. After presenting the definitions of the context and the quality of service considered in this work, this article describes the QoS model and the process allowing finding the best configuration and deployment. Finally, we present several results concerning the execution time of the deployment process of Kalimucho with different devices.

- **Securing the Cloud: cross domain and multi-level concerns.** The evolution of new deployment architectures, as illustrated by the move towards mobile platforms and the Internet Of Services, and the introduction of new security regulations (imposed by national and international regulatory bodies, such as SOX4 or BASEL5) are important constraints in the design and development of service composition. In such a context, it is not sufficient to apply the corresponding adaptations only at the service orchestration or at the choreography level; there is also the need for controlling the impact of new security requirements on several architectural layers. In [35] we have presented a new service model that supports the clean modularization of such crosscutting security concerns.
- **Software composition and product lines.** As part of the AMPLE project (see [http://www.ample-project.net](http://www.ample-project.net)) we have co-edited a book [42] covering the main scientific solutions and techniques on new methods and techniques for software product lines that have been developed in the project. This project aims at improving traditional software product line engineering using advanced software engineering namely model-driven and aspect-oriented engineering approaches. Software composition takes an important part in this book and it appears in several chapters with models, events, aspects and components.

Chapter [43] provides an overview of software product lines, model-driven engineering and aspect-oriented software development. The challenges to address and the expected benefits are drawn, this is concluded by an overview of the AMPLE approach and its tool chain.

Chapter [40] reviews the specificities of traceability in a product line context starting by identifying the challenges of maintaining traceability for traditional system development and for software product lines. This work defines the concepts that should guide the adoption of a traceability environment for product lines and illustrates these specifications with a concrete example of a traceability repository. It also provides examples of scenarios that use this traceability environment to solve concrete problems.

### 6.3. Cloud Computing, Virtualization and Data Centers

**Participants:** Frederico Alvares, Gustavo Bervian Brand, Thomas Chavrier, Fabien Hermenier, Adrien Lèbre, Thomas Ledoux, Guillaume Le Louët, Jean-Marc Menaud, Hien Nguyen Van, Rémy Pottier, Flavien Quesnel.

In the context of Cloud computing ASCOLA members have principally worked this year on capacity planning solutions for large scale distributed system. Capacity planning is the process of planning, analyzing, sizing, managing and optimizing capacity to satisfy demand in a timely manner and at a reasonable cost.

Applied to distributed systems like the Cloud, a capacity planning solution must mainly provide necessary resources for the proper execution of applications and respect service-level agreements in a large distributed environment.

The main challenges in this context are: scalability, fault tolerance and reactivity of the solution in a large-scale distributed system; analyzing, sizing, and optimizing resources to minimize the cost (energy, human, hardware etc.); and profiling, adapting application to ensure the quality of service (throughput, response time, availability etc.).

Our solutions are mainly based on virtualized infrastructures. Our main results concern the management and the execution of applications by leveraging virtualization capabilities on cloud infrastructures and the investigation of solutions that aim at optimizing the trade-off between performance and energy costs of both applications and Cloud resources.

#### 6.3.1. Virtualization and Job Management

This year, in cooperation with the Myriads project-team from INRIA Rennes-Bretagne Atlantique, we have continued to address resource-management issues concerning the federation of very large scale platforms. We have completed our approach aiming at the automatic adaptation of both hardware and software resources to the needs of the applications through a unique method. For each application, scientists describe the requirements in terms of both hardware and software expectations through the definition of a *Virtual Platform (VP)* and a *Virtual System Environment (VSE)* [18].

In addition, we started to address the design and the implementation of a fully distributed cloud OS. Designing and implementing such models is a tedious and complex task. However, as well as research studies on OSes and hypervisors are complementary at the node level, we showed that virtualization frameworks can benefit from lessons learned from distributed operating system proposals [34]. Leveraging this preliminary result, we designed and developed a first proof-of-concept of a fully distributed scheduler [33]. This system makes it possible to schedule VMs cooperatively and dynamically in large scale distributed systems. Preliminary results showed that our scheduler was more reactive. This building block is a first element of a more complete
cloud OS, entitled DISCOVERY (DIStributed and COoperative mechanisms to manage Virtual EnViRonments autonomically) [30]. The ultimate goal of this system is to overcome the main limitations of the traditional server-centric solutions. The system, currently under development, relies on a peer-to-peer model where each agent can efficiently deploy, dynamically schedule and periodically checkpoint the virtual environments s/he manage.

We have contributed new results on the Entropy software and extended our solution VM that features dynamic consolidation. In [44] and [38] we extended our dynamic consolidation manager to take into account not only resource constraints but also the placement constraints of highly available (HA) applications. In fact, most previous dynamic consolidation systems optimize the placement of the VMs according to their resource usage but do not consider the application placement constraints that are required to achieve both high availability and scalability. Our approach provides efficient dynamic consolidation while guaranteeing to the application administrator that placement requirements will be satisfied and relieving the data center administrator of the burden of considering the constraints of the applications when performing maintenance.

In the same domain, Jean-Marc Menaud has defended his habilitation (HdR - Habilitation à diriger des recherches) [13] on Jun. 22, 2011. One part of this HDR focuses on dynamic adaptation strategies for cluster administration. We have proposed a dedicated language for virtual machines management and one particular feature of our solution is to use a constraint solver to provide an appropriate placement. Based on these results, our recent contributions address the problem of data center energy consumption.

Moreover, we have continued to analyze how energy concerns can be addressed in large scale distributed infrastructures.

6.3.2. Optimization of Energy Consumption in Data Centers

As a direct consequence of the increasing popularity of Cloud Computing solutions, data centers are growing at a fast rate and hence have to face difficult energy consumption issue. Available solutions rely on Cloud Computing models and virtualization techniques to scale up/down applications based on their performance metrics. Although those proposals can reduce the energy footprint of applications and, by transitivity, of cloud infrastructures, they do not consider the internal characteristics of applications to finely define a trade-off between QoS properties of applications and their energy footprint. In [22], we propose a self-adaptation approach that considers both application internals and system properties to reduce the energy footprint in cloud infrastructures. Each application and the infrastructure are equipped with their own control loop, which allows them to autonomously optimize their executions. Simulations show that the approach may lead to appreciable energy savings without interfering on application provider revenues.

6.3.3. Cloud Computing and SLA Management

Cloud computing is a paradigm for enabling remote, on-demand access to a set of configurable computing resources as a service. The pay-per-use model enables service providers to offer their services to customers at different Quality-of-Service (QoS) levels. These QoS parameters are used to compose service-level agreements (SLAs) between a service provider and a service consumer. A main challenge for a service provider is to manage SLAs for its service consumers, i.e., automatically determine the appropriate resources required from the lower layer in order to respect the QoS requirements of the consumers. In [27], we have proposed an optimization framework driven by consumer preferences to address the SLA dependencies problem across the different cloud layers as well as the need of flexibility and dynamicity required by the domain of Cloud computing. Our approach aims at selecting the optimal vertical business process designed using cross-layer cloud services and enforcing SLA dependencies between layers. Experimental results demonstrate the flexibility and effectiveness of our approach.

6.4. Foundations of program semantics

Participants: Nicolas Tabareau, Guilhem Jaber.

ASCOLA team members have contributed several results to the foundations of program semantics.
6.4.1. Program Equivalences

Reasoning about program equivalence is a major problem in semantics. This very old topic has many applications, e.g., program verification, compiler correctness or representation independence. It has been understood since the late 1960s that tools and structures arising in mathematical logic and proof theory can usefully be applied to the development of reasoning principles for program equivalence. In recent years, based on the seminal work of Pitts and Stark, the notion of logical relation appeared to be very fruitful for proving the equivalence of programs in the presence of features like general recursive types, general (higher-order) mutable references, and first-class continuations. We have studied the notion of logical relations for proving program equivalences of low level machine codes [39].

We have also developed a forcing theory inspired by Cohen’s forcing to increase the power of a semantics model just as the latter makes it possible to enrich a set theoretical universe. In this way, we can define a new generation of logical relations—that can be introduced modularly using forcing theory—to be used for proving program equivalence for low level languages, concurrent languages or domain specific languages. [48]
6. New Results

6.1. Model Transformation

Model transformation and in particular our ATL model transformation language continues playing a key role in our MDE strategy. During 2011 the new results in this area have been:

- The development of an execution algorithm for the lazy execution of ATL transformations [31]. The increasing adoption of Model-Driven Engineering in industrial contexts highlights scalability as a critical limitation of several MDE tools. When these tools are built around model-to-model (M2M) transformations, the efficiency of the transformation engine risks to become a performance bottleneck for the whole MDE environment. This new execution mode solves this problem by providing on-demand execution of model transformations. The computation required to generate a data element of the target model is only triggered once the user requests to read that data element.

- Towards a general semantics for transformation languages. In the mid-term, we would like to be able to achieve interoperability among existing transformation languages (e.g. to create mixed transformations or simply to compare alternative transformation solutions for the same scenario). In this direction, we have worked on a general composition semantics for rule-based transformation languages [32] and an extensive survey of inheritance support in existing languages [33]. Moreover, to make ATL closer to other transformation languages, in special graph transformations, we have developed a new ATL refining mode [43] that allows executing in-place transformations.

- Also relevant, this year AtlanMod has coorganized the main international conference on model transformations [41].

6.2. Model Representation

As part of our work on core techniques for the specification and internal representation of models we would like to highlight:

- Improved model typing. In order to represent metadata more accurately, we have worked on a functional typing system for megamodeling [18]. The basic idea is to consider transformations as functions, and to give them functional types that can facilitate the reasoning on them.

- Virtual EMF [21] for the transparent Composition, Weaving and Linking of Models. Virtual EMF facilitates the global manipulation of an heterogeneous but related set of models by providing the illusion of having a single and unified view of the modeled system.

- EMF Profiles [25], a mechanism to enable users annotate existing models without polluting them (i.e. without directly changing their content).

- Batch scripting support for the retrieval and manipulation of models from model repositories thanks to our new language MoScript [24].

6.3. Model Quality

Our work on model quality defends the idea that there is not a silver bullet technique to verify the quality of models and that a combined approach offers the best trade-off. Lightweight techniques can provide a quick feedback even if it is partial and, when necessary, more complete ones can be utilized to complement the results. In this sense, this year we have developed:

- Lightweight techniques for the verification of ATL transformations [27] and UML Executable models [26].
• Proposed a method for the automatic generation of correct dynamic models for a given domain [11]. Basic operations to modify the information the software needs to store and manipulate about the domain are generated. The generation process ensures that all generated operations are strongly executable.

• Created a brand new Eclipse Lab project aimed at facilitating the verification of any kind of (EMF-based) model with the EMFtoCSP tool (see the tool description in the tools section of this report)

• And applied some of these ideas to the verification and testing of web applications [15]

6.4. Domain Specific Language

There is a growing interest in the research community on the definition of a new generation of language workbenches to facilitate the definition of non-trivial domain specific languages. In this area, we are working on the definition of a set of quality properties that will help language designers to validate if their language is well-adapted to the needs of the language users. This is done by mining repositories containing a corpus of examples of the language.

To begin with we have been reusing the notion of software clones [16] [30], a well-known property in the programming domain and see if it is also meaningful for languages at the model level [29].

6.5. Reverse Engineering

We have continued our work on MoDisco, specially wrt the dissemination activities around the platform [19] [40] and the extension of the tooling support mainly through the spin-off of the EMFFacet Eclipse project, explained in the tools section of this report.

6.6. Systems interoperability

MDE can be used as an intermediate representation between two different technical spaces / platforms / tools to facilitate their interoperability. During this year, we have followed this approach in the following results:

• Automation of the interactions with APIs by automatically discovering and expressing them as models thanks to our API2MOL approach [14]

• Bridging the business and the technical domains, developing among others the Portolan tool [42] for the cartography of Information Systems.

• MDE itself can benefit from the work and techniques available in a different technical space. In this context we have combined MDE and constraint programming to see how the combination improves the solution of classical problems like the configuration of a set of components/plug-ins [22]. We have even organized a workshop on the topic of merging MDE and logic programming [20] to better understand how they can benefit each other.
6. New Results

6.1. Intrusion Detection

Metamorphic codes:

In [12], we have proposed an advance code obfuscation technique for metamorphic codes (i.e., codes that automatically recode themselves each time they propagate or are distributed). We have shown that the detection of such codes was a problem for classical nowadays static detection tools. In [25], we focus on a new dynamic detection approach which allows to detect variants produced by our metamorphic engine. In addition, our approach can detect unknown malware as long as their behavior approaches that of a known malware. For this, we propose to use a measure of similarity between program behaviors. This measure is obtained by lossless compression of execution traces in terms of system calls.

Intrusion Detection based on an Analysis of the Flow Control:

In [13], intrusion detection mechanisms based on the construct a model of normal behavior of the supervised entity are studied. Such a model is used during the detection phase to raise an alarm when a deviation is observed. This approach allows to detect unknown attacks.

The most common anomaly detection mechanisms at application level consist in detecting a deviation of the control-flow of a program. A popular method to detect such anomaly is the use of application sequences of system calls. However, such methods do not detect mimicry attacks or attacks against the integrity of the system call parameters. To enhance such detection mechanisms, we propose in [27] an approach to detect in the application the corruption of data items that have an influence on the system calls. This approach consists in building automatically a data-oriented behavior model of an application by static analysis of its source code. The proposed approach is illustrated on various examples, and an injection method is experimented to obtain an approximation of the detection coverage of the generated mechanisms.

Most of today’s MAC implementations can be turned into permissive mode, where no enforcement is performed but alerts are raised instead. This behavior is very close to an anomaly IDS except that the system is configured through a MAC policy. MAC implementations such as SELinux and AppArmor come with a default policy including real life and practical rules ready to be used as is or as a basis for a custom policy. In [30], we first propose an extension of an IDS based on information flow control. We address issues concerning programs execution and improve its expressiveness in terms of security policy. This extended model can be configured to reach a wide variety of different security goals. Particularly, it allows for information flow checking based on users and/or programs dependent policy rules. Furthermore, suspicious modification of binary programs can be detected to avoid malware execution. We also propose an algorithm for deriving an AppArmor MAC policy into an information flow policy, and thus get the advantage of having a ready to use policy offering good security. An integration within Android is described in [37].

Flow based Interpretation of Access Control Policies:

In [32], we introduce a formal property characterizing access control policies for which the interpretations of access control as mechanism over objects and as mechanism over information contained into objects are similar. This leads us to define both a flow based interpretation of access control policies and the information flows generated during the executions of a system implementing an access control mechanism. When these two interpretations are not equivalent, we propose to add a mechanism dedicated to illegal information flow detection to the mechanism of access control over objects. Such a mechanism is parameterized by the access control policy and is proved sound and complete. We also briefly describe two real implementations, at two levels of granularity, of our illegal flow detection mechanism: one for the Linux operating system and one for the Java Virtual Machine.

Intrusion Detection based on Invariants:
RRABIDS (Ruby on Rails Anomaly Based Intrusion Detection System) [40] is an application level intrusion detection system for applications implemented with the Ruby on Rails framework. The goal of this intrusion detection system is to detect attacks against data in the context of web applications. This anomaly based IDS focuses on the modeling of the normal application profile using invariants. These invariants are discovered during a learning phase. Then, they are used to instrument the web application at source code level, so that a deviation from the normal profile can be detected at run-time. On simple examples we show how the approach detects well known categories of web attacks that involve a state violation of the application, such as SQL injections. Finally, an assessment phase is performed to evaluate the accuracy of the detection provided by the proposed approach.

Alert Correlation:
Alert correlation is a crucial problem for monitoring and securing computer networks. It consists in analyzing the alerts triggered by intrusion detection systems (IDSs) and other security related tools in order to detect complex attack plans, discover false alerts, etc. The huge amounts of alerts raised continuously by IDSs and the impossibility for security operators to efficiently analyze them requires tools for eliminating false and redundant alerts on the one hand and prioritize them according the detected activities? dangerousness and preferences of the analysts on the other hand. In [35], we describe an architecture that combines AI-based approaches for representing and reasoning with security operators? knowledge and preferences. Moreover, this architecture allows to combines experts’ knowledge with machine learning and classifier based tools. This prototype collects the alerts raised by security related tools and analyzes them automatically.

Trust-Based IDS for the AODV Protocol:
Routing in ad hoc networks is based on mutual trust between collaborating nodes. Security problems arise when supposedly honest nodes lie deliberately to maximize their profit. In [11], we are interested in detecting misbehaving nodes within the ad hoc routing protocol AODV. We propose and implement a real-time intrusion detection system based on implicit trust relations: a node implementing this system collects its neighbors’ routing messages and reasons on them to decide on their honesty. We also evaluate our implementation, and, based on simulations, show that the system we have developed to detect dishonest behavior is efficient.

Modelization and Simulation of Zombies Behaviours:
In [26], we study the modelization and simulation of zombie machines for the evaluation of Network Intrusion Detection Systems (NIDS), used to detect botnets. We propose an automatic method to infer zombies behaviours through the analysis of messages exchanged with their masters. Once computed, a model provides a way to generate realistic and manageable traffic, which is mandatory for an NIDS evaluation. We propose to use a Stochastic Mealy Machine to model zombies behaviours, and an active inference algorithm to learn it. With our approach, it is possible to generate a realistic traffic corresponding to the communications of botnets while ensuring its controllability in the context of an NIDS evaluation.

6.2. Privacy
Computer privacy is a domain where the education and information of the general public is paramount. In this perspective, through [44] we have participated to the popularization effort in the area, by exposing a survey of accessible computing tools allowing users to better protect their online privacy.

Formal Privacy Policies and Logical Tools:
One of the obstacles to the improvement of the privacy level in distributed applications is the lack of expressiveness, usability and enforceability of the associated policies. This new research track aims at designing better privacy policies for complex systems, more adapted to the specific needs of personal data protection regulations and easier to enforce in a distributed fashion. Logical languages, in particular, are considered as interesting candidates because of the reasoning capabilities attached to the formalisms, allowing autonomous peers to perform efficient, privacy-aware planning. [18] is a contribution to the modal logics used to model formal norms, focusing on specific deadline-related temporal notions often encountered in privacy policies. In [39], we propose an ambitious, collaborative research project based on an epistemic view of
the privacy laws and regulations, which should lead to the design of several tools, including policy writing assistants and validation software. [24] is a generic work in the domain of formal policies, where we propose a logical model of various concepts of responsibility in an organizational framework featuring obligation delegation. This kind of framework is intended to model the handling of complex policies in real-life human institutions.

**Privacy in Social Networking Sites:**

Social Networking Sites (SNS), such as Facebook and LinkedIn, have become the established place for keeping contact with old friends and meeting new acquaintances. As a result, a user leaves a big trail of personal information about him and his friends on the SNS, sometimes even without being aware of it. This information can lead to privacy drifts such as damaging his reputation and credibility, security risks (for instance identity theft) and profiling risks. Another research challenge stems from the fact that in the digital world where it is possible to copy the information as often as desired, it is not easy to control how information is disseminated once it is out on the Internet. In an ongoing collaboration [23] with Ai Thanh Ho and Esma Aïmeur (Université de Montréal), we investigate tools that can help user to maintain the sovereignty of their data on the World Wide Web. We also introduce PrivacyMarker, an approach drawing on the concept of provenance and accountability to protect user privacy on SNS. More precisely, it is possible to imagine that by a combination of logs and techniques such as watermarking and traitor-tracing schemes, the dissemination of information can be (at least partially) controlled and that in case of a privacy breach, it is possible to identify which persons are potentially suspect because they have previously accessed this information.

**Geo-privacy:**

A geolocated system generally belongs to an individual and as such knowing its location reveals the location of its owner, which is a direct threat against his privacy. To protect the privacy of users, a sanitization process, which adds uncertainty to the data and removes some sensible information, can be performed but at the cost of a decrease of utility due to the quality degradation of the data. In a joint work [16] with Marc-Olivier Killijian and Miguel Nunez del Prado (LAAS-CNRS), we describe GEPETO (for GEoPrivacy-Enhancing TOolkit), a flexible open source software which can be used to visualize, sanitize, perform inference attacks and measure the utility of a particular geolocated dataset. We also introduce a mobility model that we coin as mobility Markov Chain, which can represent in a compact yet precise way the mobility behaviour of an individual. Finally, we describe an algorithm for learning such a structure from the mobility traces of an individual.

Geosocial networks are relatively new compared to the more “traditional” (i.e. non-geolocated) social networking sites such as Facebook or LinkedIn that have been around since more than 6 years, but they are currently growing relatively fast along with the widespread development of other geolocated applications and technologies. In a study [29] done in cooperation with Olivier Heen (Technicolor) and Christophe Potin, we provide a comparative analysis of some existing geosocial networks with respect to privacy in order to (1) highlight some of privacy issues that are raised by the fast development of these system and (2) propose recommendations that could be integrated in the design of these systems to enhance the privacy of their users based on this analysis.

**Privacy in Distributed Systems:**

In a joint work [19] with Anne-Marie Kermarrec and Mohammad Alaggan (team INRIA ASAP), we address the problem of computing the similarity between two users (according to their profiles) while preserving their privacy in a fully decentralized system and for the passive adversary model. First, we introduce a two-party protocol for privately computing a threshold version of the similarity and apply it to well-known similarity measures such as the scalar product and the cosine similarity. The output of this protocol is only one bit of information telling whether or not two users are similar beyond a predetermined threshold. Afterwards, we explore the computation of the exact and threshold similarity within the context of differential privacy, a recent notion developed that provides a strong privacy guarantee that holds independently of the auxiliary knowledge that the adversary might have. More specifically, we design several differentially private variants of the exact and threshold protocols and we also analyze their complexity as well as their impact on the utility
of the resulting similarity measure. Finally, we provide experimental results validating the effectiveness of the proposed approach on real datasets.

Other ongoing work tackles the problem of computing an aggregation function in a secure and scalable way in a distributed network [42] (joint work with Rachid Guerraoui, Hamza Harkous, Florian Huc and Anne-Marie Kermarrec).

6.3. Accidental and Malicious Faults in Distributed Systems

Induced Churn to Face Malicious Behaviors:

In reputation mechanisms, ensuring durable access to feedbacks is a first barrier against simple attacks. To bias the reputation mechanism, an adversary can create and use several distinct identities. In that case, if the reputation mechanism is solely based on statistical measurements, the trustworthiness can be violated. Our contribution is centered around the study of robust mechanisms that can resist such attacks.

Toward this goal, we have first investigating the problem of uniform sampling in large scale open systems in presence of adversarial nodes. Uniform sampling ensures that any individual in a population has the same probability to be selected as sample. Uniform sampling finds its root in many problems such as data collection, dissemination, load balancing, and data-caching.

By relying on the topological properties of structured peer-to-peer systems, it has been shown that it is possible to guarantee with high probability that any node is equally likely to appear in the local view of each other honest node in a number of rounds polynomial in the size of the system. This is achieved by imposing nodes to frequently depart from their position and move to another random position in the system. Indeed, in [15], we have shown that an adversary can very quickly subvert overlays based on distributed hash tables by simply never triggering leave operations. We have also demonstrated that when all nodes (honest and malicious ones) are imposed on a limited lifetime, the system eventually reaches a stationary regime where the ratio of polluted clusters is bounded, independently from the initial amount of corruption in the system.

In unstructured peer-to-peer systems, nodes cannot rely on the topological nature of structured graphs to detect undesirable behaviors. The sampling has to be uniform and ergodic. Informally, this second property guarantees that each received node id infinitely often has a non-null probability to locally appear as a sample. In [21], we determine necessary and sufficient conditions under which uniform and ergodic sampling is achievable in unstructured peer-to-peer systems potentially populated with a large proportion of Byzantine nodes. Strict restrictions are imposed on the number of messages gossiped by malicious nodes during a given period of time and providing each honest node with a very large memory (in the size of the system).

In [38], we consider the problem of targeted attacks in large scale peer-to-peer overlays. These attacks aimed at exhausting key resources of targeted hosts to diminish their capacity to provide or receive services. To defend the system against such attacks, we rely on clustering and implement induced churn to preserve randomness of nodes identifiers so that adversarial predictions are impossible. We propose robust join, leave, merge and split operations to discourage brute force denial of services and pollution attacks.

Sequence of Consensus Instances:

To be able to coordinate efficiently the activities of replicas, a significant body of work on replication techniques, group communication services and agreement problems has been done. The Consensus service has been recognized as a fundamental building block for fault-tolerant distributed systems. Many different protocols to implement such a service have been proposed, however, little effort has been placed in evaluating their performance. During her PhD thesis [14], Izabela Moise has presented a protocol designed to solve several consecutive consensus instances in an asynchronous distributed system prone to crash failures and message omissions. The protocol [31] follows the Paxos approach [49], [47] and integrates two different optimizations to reduce the latency of learning a decision value. As one optimization is risky [48], dynamics triggering criterion are defined to check at runtime if the context seems to be favorable or not. The proposed protocol [34] is adaptive as it tries to obtain the best performance gain depending on the current context. Moreover, it guarantees the persistence of all decision values. Our experimentation results [
focus on the impact of the prediction of collisions (i.e., the cases where the use of the risky optimization is counterproductive).

Transactional Mobile Agent:

Mobile devices are now equipped with multiple sensors and networking capabilities. They can gather information about their surrounding environment and interact both with nearby nodes, using a dynamic and self-configurable ad-hoc network, and with distant nodes via the Internet. While the concept of mobile agent is appropriate to explore the ad-hoc network and autonomously discover service providers, it is not suitable for the implementation of strong distributed synchronization mechanisms. Moreover, the termination of a task assigned to an agent may be compromised if the persistence of the agent itself is not ensured. In the case of a transactional mobile agent, we identify two services, Availability of the Sources and Atomic Commit, that can be supplied by more powerful entities located in a cloud. In [33], we propose a solution where these two services are provided in a reliable and homogeneous way. To guarantee reliability, the proposed solution relies on a single agreement protocol that orders continuously all the new actions whatever the related transaction and service.
6. New Results

6.1. Network Economics

Participants: Pierre Coucheney, Hai Tran Hoang, Bruno Tuffin, Jean-Marc Vigne.

While pricing telecommunication networks was one of our main activities for the past few years, we are now dealing with the more general topic of network economics. We have tackled it from different sides: i) investigating how QoS or QoE can be related to users’ willingness to pay, ii) investigating the consequences and equilibrium due to competition among providers in different contexts, iii) studying the economic aspect of interdomain relationships, iv) looking at the economics of applications, for example adword auctions for search engines, v) investigating the economics of security in telecommunications, vi) studying the network neutrality issue.

On the first item, in [70], [29], we have studied how utility functions can be related to QoE recent research. Indeed, a logarithmic version of utility usually serves as the standard example due to its simplicity and mathematical tractability. We argue that there are much more (and better) reasons to consider logarithmic utilities as really paradigmatic, at least when it comes to characterizing user experience with specific telecommunication services. We justify this claim and demonstrate that, especially for Voice-over-IP and mobile broadband scenarios, there is increasing evidence that user experience and satisfaction follows logarithmic laws. Finally, we go even one step further and put these results into the broader context of the Weber-Fechner Law, a key principle in psychophysics describing the general relationship between the magnitude of a physical stimulus and its perceived intensity within the human sensory system.

A notable part of our activity has been related to competition among telecommunication providers, mainly within the framework of the ANR CAPTURES project. The goal is to improve most of the pricing models analysis which only deal with a single provider while competition (that is observed in the telecommunication industry) can drive to totally different outcomes. A general view of some of our results is summarized in [67]. A general model of competition in loss networks is described and analyzed in [22] as a two-levels game: at the smallest time scale, users’ demand is split among providers according to the Wardrop principle, depending on the access price and available QoS (depending itself on the level of demand at the provider), and at the largest time scale, providers play a pricing game, trying non-cooperatively to maximize their revenue. A striking result is that this game leads to the same outcome than if providers were cooperatively trying to maximize social welfare: the so-called price of anarchy is equal to one. An additional (higher) level of game is analyzed in [23] (but using another type of negative externality for users, based here on delay), at which providers play on the technologies to implement, based on the infrastructure and licence (if any) costs, anticipating what would be the resulting price war outcome and revenue for given profiles of sets of technologies. This type of study may help a regulator to decide a licence cost, in order to drive the resulting Nash equilibrium to a better point in terms of social or user welfare. A specific situation we have analyzed is the case for a competitive market operated by a Mobile Network Operator (MNO) and a Mobile Virtual Network Operator (MVNO) [46]. The resource that is leased by the MNO to the MVNO is spectrum. MNO and MVNO compete posting subscription prices and the mobile users may choose to subscribe to one operator, or not to subscribe. The scenario is modeled by a three-level game comprising: a bargaining game, which models the spectrum leasing by the MNO; a competition game, which models the price competition between the MNO and the MVNO, and a subscription game, which models the subscription choice by the mobile users, and the outcome of which may be either not to subscribe, to subscribe to the MNO or to subscribe to the MVNO. We assess which conditions lead to an equilibrium where the competition does take place and the amount of the spectrum that should be leased to maximize user or social welfare.
Another important activity is around interdomain issues, with a network like the Internet being made of thousands of autonomous systems. Intermediate domains need some (economic in our case) incentives for forwarding the traffic of other domains. In [33], we have described the problem, provided a state of the art and highlighted the difficulties that must be solved. In [32], we have designed a decentralized algorithm based on double-sided auctions to allocate (and charge) the resource usage.

But network economics is not only about ISPs, it also deals with the application side. In order to make money many service providers base their revenue on advertisement. Search engines for example get revenue thanks to adword auctions, where commercial links are proposed and charged to advertisers as soon as the link is clicked through. Most search engines have chosen (or switched to) a revenue-based ranking and charging scheme instead of a bid-based one. In [53] we investigate the relevance of that scheme when advertisers’ valuation comes from a random distribution, showing that depending on the search engine’s click-through-rate, revenue-based does not always outperform bid-based in terms of revenue to the search engine. But in this adword auction context too, there exist very few works dealing with search engines in competition for advertisers. We have developed a two-level game where at the largest time scale search engines decide which allocation rule to implement, between revenue-based and bid-based; and at the lowest time-scale advertisers decide how to split their advertising budget between the two search engines, depending on the benefits this will bring to them. The game at the largest time scale is solved using backward induction, the search engines anticipating the reactions of advertisers [54], [52]. We describe the advertisers best strategies and show how to determine, depending on parameters, an equilibrium on the ranking rule strategy for search engines; this may explain Yahoo!’s move to switch from bid-based to revenue-based ranking to follow Google’s strategy.

We similarly have looked at the competition aspects linked to security. We have reviewed the interactions and strategies of attackers and defenders [68]. But we have also looked at the economics of network security, when network users can choose among different security solutions to protect their data, offered by competitive security providers [51]. The interactions among users are modeled as a noncooperative game, with a negative externality coming from the fact that attackers target popular systems to maximize their expected gain.

A new issue we are investigating is the network neutrality debate coming from the increasing asymmetry between Internet Service Providers (ISPs), mainly due to some prominent and resource consuming content providers which are usually connected to a single ISP. We have described and analyzed in [69] the respective arguments of neutrality proponents and opponents, and are currently completing the analysis of several promising game-theoretic models on this issue.

### 6.2. Dependability and extensions

**Participants:** Raymond Marie, Gerardo Rubino, Samira Saggadi, Bruno Tuffin.

We maintain a permanent research activity in different domains related to dependability, performability and vulnerability analysis of communication systems. Our focus is on evaluation techniques using both the Monte Carlo and the Quasi-Monte Carlo approaches. Monte Carlo (and Quasi-Monte Carlo) methods often represent the only available tool to solve complex problems in the area, and rare event simulation requires a special attention, in order to be able to efficiently analyze the model, that is, to be able to use good estimators having, in particular, a sufficiently small relative variance. Novel results in simulation can be decomposed into two subsets: results on rare event simulation, and those on Randomized Quasi-Monte Carlo methods.

The effectiveness of randomized quasi-Monte Carlo (RQMC) techniques is examined in [26] to estimate the integrals that express the discrete choice probabilities in a mixed logit model, for which no closed form formula is available. These models are used extensively in travel behavior research. We consider popular RQMC constructions, but our main emphasis is on randomly-shifted lattice rules, for which we study how to select the parameters as a function of the considered class of integrands. We compare the effectiveness of all these methods and of standard Monte Carlo (MC) to reduce both the variance and the bias when estimating the log-likelihood function at a given parameter value.
The main part of our activity in this simulation area in 2011 has been on rare event simulation though. The two major simulation families or rare event estimations are importance sampling and splitting. In [63], we have provided a recent view of those methods, while in [64] we have overviewed how the zero-variance importance sampling can be approximated in classical reliability problems.

The problem of estimating the probability that a given set of nodes is connected in a graph (or network) where each link is failed with a given probability has received a lot of attention from us in 2011. We have proposed in [21] a new Monte Carlo method, based on dynamic Importance Sampling. The method generates the link states one by one, using a sampling strategy that approximates an ideal zero-variance importance sampling scheme. The approximation is based on minimal cuts in subgraphs. In an asymptotic rare-event regime where failure probability becomes very small, we prove that the relative error of our estimator remains bounded, and even converges to 0 under additional conditions, when the unreliability of individual links converges to 0. The empirical performance of the new sampling scheme is illustrated by examples. The method is even sped up in [50] by applying series-parallel reductions at each step of the algorithm.

The same problem is also analyzed in [15] by a novel method that exploits a generalized splitting (GS) algorithm. We show that the proposed GS algorithm can accurately estimate extremely small unreliabilities and we exhibit large examples where it performs much better than existing approaches. Remarkably, it is also flexible enough to dispense with the frequently made assumption of independent edge failures. In [17], another splitting approach is explored for the same problem, with very good results. It consists of a standard splitting procedure applied to the so-called Creation Process that can be associated with the initial static model. The paper discusses both a method for splitting this process, and an experimental analysis of the covering of the resulting estimator, showing its good behavior on different classes of test problems. Last, in [16], for the same static reliability problem, we proposed a new procedure belonging to the RVR family (Recursive Variance Reduction) where a new estimator based both in computed minpaths and mincuts of the graph, together with series-parallel reductions, allows to obtain very good accuracy in many rare events situations.

Finally, a versatile Monte Carlo method for estimating multidimensional integrals, with applications to rare-event probability estimation is presented in [39], [75]. The method uses two distinct and popular Monte Carlo simulation techniques, namely Markov chain Monte Carlo (MCMC) and Importance Sampling, combined into a single algorithm. We show that for some illustrative and applied numerical examples the proposed Markov Chain Importance Sampling algorithm performs better than methods based solely on Importance Sampling or solely on MCMC.

Concerning the risk on spares for life-time maintenance purposes which is due to uncertainties on the mean up time, an extended version of a presentation made in 2010 has been published in [24].

6.3. Performance evaluation

Participants: Laura Aspirot, Raymond Marie, Gerardo Rubino, Bruno Sericola.

An important problem arising when dimensioning a P2P system is to understand the evolution of the peers’ population with time. The number of units being usually large, the standard stochastic models used to represent this kind of system (e.g. a Markovian stochastic process) are difficult to use in practice. Instead, it is popular today to move to deterministic continuous-state (fluid) models whose dynamics is governed by differential equations. It is then of interest to analyze the conditions under which the latter are the limit, in some sense, of the former. We started to develop this program in [36] by focusing on some popular models of P2P systems, and analyzed when and how the deterministic model is the limit of the stochastic one when the number of peers goes to infinity.

In [60], we continued to explore the concept of power of a queueing model proposed by Kleinrock in the 80s. Kleinrock’s idea was to build a metric combining two “competing” ones, the mean throughput and the mean response time, for the system in equilibrium. The power is defined as the ratio of normalized versions of those metrics. We discuss different ways of adapting this concept to more general queueing systems such as queueing networks. In this research line, [60] opens the way for a definition of efficiency, which is currently analyzed in the team.
In [30], we expose a clear methodology to analyze maximum level and hitting probabilities in a Markov driven fluid queue for various initial condition scenarios and in both cases of infinite and finite buffers. Step by step we build up our argument that finally leads to matrix differential Riccati equations for which there exists a unique solution. The power of the methodology resides in the simple probabilistic argument used that permits to obtain analytic solutions. We illustrate our results by a comprehensive fluid model that we solve exactly.

In [65], we analyze the transient behavior of a fluid queue driven by a general ergodic birth and death process using spectral theory in the Laplace transform domain. These results are applied to the stationary regime and to the busy period analysis of that fluid queue.

Finally, in [71] we present a global view of the performance evaluation area in computer and communication systems, an extended and reviewed version of a talk given in 2010.

6.4. Quantitative aspects of distributed systems

Participants: Bruno Sericola, Romaric Ludinard.

This work is a collaboration with the Inria team-project Asap. We proposed in [20] a fully decentralized algorithm to provide each of the nodes of a distributed system with a value reflecting its connectivity quality. Comparing these values between nodes, enables to have a local approximation of a global characteristic of the graph. Our algorithm relies on an anonymous probe visiting the network in an unbiased random fashion. Each node records the time elapsed between visits of the probe which is called the return time of the random walk. Computing the standard deviation of such return times enables to approximate the conductance of the graph. Typically, this information may be used by nodes to assess their position, and therefore the fact that they are critical, in a graph exhibiting low conductance.

We continue our collaboration with the Inria team-projects Adept and Ipso. It is well-known that peer-to-peer overlays networks can only survive Byzantine attacks if malicious nodes are not able to predict what will be the topology of the network for a given sequence of join and leave operations. In [13] and in [35], we investigate adversarial strategies by following specific games. Our analysis demonstrates first that an adversary can very quickly subvert DHT-based overlays by simply never triggering leave operations. We then show that when all nodes (honest and malicious ones) are imposed on a limited lifetime, the system eventually reaches a stationary regime where the ratio of polluted clusters is bounded, independently from the initial amount of corruption in the system. These results have been obtained using Markov models. In [14] and [34], we consider the behavior of a stochastic system composed of several identically distributed, but non independent, discrete-time absorbing Markov chains competing at each instant for a transition. The competition consists in determining at each instant, using a given probability distribution, the only Markov chain allowed to make a transition. We analyze the first time at which one of the Markov chains reaches its absorbing state. We obtain its distribution and its expectation and we propose an algorithm to compute these quantities. We also exhibit the asymptotic behavior of the system when the number of Markov chains goes to infinity. Actually, this problem comes from the analysis of large-scale distributed systems and we show how our results apply to this domain.

6.5. QoE (Quality of Experience)

Participants: Sebastián Basterrech, Yassine Hadjadj-Aoul, Sofiene Jelassi, Adlen Ksentini, Gerardo Rubino, Kamal Singh, César Viho.

We continue the development of the PSQA technology (Pseudo-Subjective Quality Assessment) in the area of Quality of Experience (QoE). PSQA is today a stable technology allowing to build measuring modules capable of quantifying the quality of a video or an audio sequence, as perceived by the user, when received through an IP network. It provides an accurate and efficiently computed evaluation of quality. Accuracy means that PSQA gives values close to those than can be obtained from a panel of human observers, under a controlled subjective testing experiment, following an appropriate standard (which depends on the type of sequence or application). Efficiency means that our measuring tool can work in real time, if necessary. Observe that perceived quality is the main component of QoE. PSQA works by analyzing the networking environment of the communication
and some the technical characteristics of the latter. It works without any need to the original sequence (as such, it belongs to the family of no-reference techniques).

It must be pointed out that a PSQA measuring or monitoring module is network dependent and application dependent. Basically, for each specific networking technology, application, service, the module must be built from scratch. But once built, it works automatically and very efficiently, allowing if necessary to use it in real time.

At the heart of the PSQA approach there is the statistical learning process necessary to develop measuring modules. So far we have been using Random Neural Networks (RNNs) as our learning tool (see [82] for a general description), but recently, we have started to explore other approaches. For instance, in the last ten years a new computational paradigm was presented under the name of Reservoir Computing (RC) [78] covering the main limitations in training time for recurrent neural networks while introducing no significant disadvantages. Two RC models have been developed independently and simultaneously under the name of Liquid State Machine (LSM) [81] and Echo State Networks (ESN) [78] and constitute today one of the basic paradigms for Recurrent Neural Networks modeling [79]. The main characteristic of the RC model is that it separates two parts: a static sub-structure called reservoir which involves the use of cycles in order to provide dynamic memory in the network, and a parametric part composed of a function such as a multiple linear regression or a classical single layer network. The reservoir can be seen as a dynamical system that expand the input stream in a space of states. The learning part of the model is the parametric one. In a recent collaboration with the Applied Computational Intelligence Research Unit, Artificial Neural Networks Group of the University of the West of Scotland during the first half of the year, we developed an algorithm based on a combination of topology preserving maps such as the Self-Organising Map [80] and the Scale Invariant Map [77] to improve the performance of RC models. The obtained results are presented in two papers: [37] and [38].

In [42] we developed a PSQA version for evaluating the perceived quality in the context of SVC video coding. The tool is based on the use of the RNN model. The main difficulties in defining this tool is regarding the relation between the SVC layers, since the enhanced layers require the information of the base layer in order to be decoded.

In [61], we developed a tool for evaluating the perceived quality of an application distributing streamed video using HTTP (and thus, TCP). The difficulties here are focused around the possible playout interruptions and the quality variations due to the use of adaptive bitrate techniques. Our procedure belongs to the no-reference family of learning ones, and it is also based on the use of the RNN tool.

In [41] we compared PSQA used for the video evaluation to other no-reference tools as well as two objective evaluation tools. We showed that PSQA outperforms the majority of the other tools, in terms of high correlation with human evaluation. This version will be used as the main metric for evaluating the QoE in the future internet architecture proposed by the FP7 Alicante project.

We have also being developing single-ended parametric-model speech-quality assessors of VoIP conversations over future networks. To do that, a careful identification and accurate characterization of quality-degrading factors over next-generation networks has been done. The recent progress and challenges for accurate assessment of voice quality over evolving VoIP systems has been detailed in the survey paper [19]. In [18], we study the perceived effects of packet loss processes, which are the principal source of quality degradation over IP networks. In reality, the perceived effect of a given packet loss process is highly related to the distribution of missing packets. Basically, the higher the burstiness of packet loss processes, the greater the perceived quality degradation. Recently, several assessors of speech quality sensitive to packet loss burstiness have been proposed in the literature. A comprehensive comparison study of bursty-packet-loss-aware artificial assessors has been conducted in [18]. An extended and more elaborated version has been published in [47]. Moreover, novel artificial quality assessors that consider transient loss of connectivity incurred by mobile users over mobile transport system have been developed. A paper describing our developed tools and performance results is under preparation. Recently, we started to work on new analytical models of packet losses and delays of packet-based voice conversations over wireless ad-hoc networks. The developed models will be used to design specialized artificial quality assessors of multimedia services over wireless ad-hoc networks. Moreover,
we are working on the enhancement of a voice quality assessor version of PSQA, by considering the features of removed speech signals.

6.6. Wireless networks

Participants: Nizar Bouabdallah, Yassine Hadjadj-Aoul, Adlen Ksentini, Raymond Marie, Bruno Sericola, César Viho.

We continue working on wireless networking. The focus mainly concerns wireless distribution of audio and video, which require strict Quality of Service (QoS) support.

In [27], we investigated the main challenges when the goal is to constitute an efficient Radio Resource Management (RRM) framework. The existing solutions of RRM were classified based on the considered decision-making technique. Moreover, we investigated in [28] how QoE can help for designing efficient RRM for wireless networks. A resource allocation mechanism is proposed in [62]. In [59] we proposed a novel network selection mechanism for heterogeneous wireless networks that take QoE into consideration for decision-making. The main idea is to use QoE of ongoing users in candidate networks as an indicator to select the best network for connection. Besides, in order to provide efficient interworking between the different access players, we first defined some issues related to the interworking operation between the satellite and terrestrial domains. We suggested some solutions and discussed their potential in [31].

We also investigated in [74] solutions that ensure the scalability of mobile networks, which are facing a rapid increase of data traffic. We devise methods that enable User Equipments, both in idle and active mode and while being on the move, to always have optimal Packet Data Network (PDN) connections (i.e., IP addresses) in such decentralized networks. We demonstrated the effectiveness of such approach in current mobile and wireless networks. In these systems, minimizing energy consumption is becoming more and more crucial. In [66], we devised a PID (Proportional Integral Derivative)-based controller permitting to reduce the amount of wasted energy by determining an optimal schedule between the sleep and wakeup periods of the wireless interface during the VoIP communication while keeping the perceived quality at the desired level.

Based on our previous research on proactive routing for wireless ad-hoc networks, we have published a book chapter in [73], focusing on modeling the resilience of routing information for ad hoc networks where topology information is uncertain.

We continue our collaboration with the Inria team-projects Pops (Lille), D-Net (Lyon), Reso (Lyon) and the NPA (Networks and Performance Analysis) research group of LIP6 (Paris) on fast self-stabilization in large scale wireless networks. In these systems, distributed self-organization is more convenient than centralized planification. Self-stabilization protocols are a useful technique to provide self-organization but their stabilizing time is related to the size of the network. In [25], we show that a clustering algorithm, known for its good robustness properties, is actually self-stabilizing. We propose several enhancements to the scheme in order to reduce the stabilization time and thus improved the stability in a dynamic environment. The key technique to these enhancements is a localized self-stabilizing algorithm for directed acyclic graph construction. We provide extensive studies (both theoretical and experimental) that show that our approach enables efficient yet adaptive clustering in wireless multihop networks.

6.7. Sensor networks

Participants: Nizar Bouabdallah, Sofiane Moad.

Wireless Sensor Networks (WSNs) are composed of tiny sensor nodes, which are capable of sensing and processing data from inaccessible environments and communicating them to the end-user for further analysis. WSNs are characterized by the limited capacity of their sensor node batteries, making energy efficiency a critical issue. Once a WSN is deployed, sensor nodes must self-organize and live as long as possible, based only on their initial energy stores. Consequently, techniques minimizing energy consumption are required to improve network lifetime. Our research on WSNs [72] revolves around two main directions: 1) clustering, and 2) radio diversity. Regarding clustering, we first developed a Connectivity Degree-Based Energy Efficient
Clustering Protocol for WSNs (CDEEC) [55], resulting in better topology management and decreased energy consumption compared to the well-known clustering protocol HEED. Then, we integrated a compression mechanism within a cluster-based architecture to develop a Compression Cluster-based scheme in a Spatial Correlated Region protocol (CC_SCR) [56], with the goal of further decreasing the energy consumption. In the direction of radio diversity, we first proposed the WETX metric [58] which uses a minimum-energy radio while routing, then we proposed the BL metric [57], on top of WETX, which allows energy-balancing inside a network in order to further extend its lifetime. The validation of our contributions was carried out with analytical analysis, and simulation using TOSSIM.

6.8. Scalable Video Coding (SVC) transmission over IP and Broadcast networks

Participants: Majd Ghareeb, Adlen Ksentini, César Viho, Yassine Hadjadj-Aoul.

One of the multimedia market trends is audiovisual service (TV or VoD) anywhere, at any time. To support such service, a Video Service Provider has to manage, store, and distribute content towards multiple kinds and scales of terminals, and over different and transient access technologies to reach the end user. To solve such issues, video scalability seems to be the most relevant solution. It encodes the video in multiple separated layers, which enable a large number of users with heterogeneous capability to view any desired video stream, at anytime, and from anywhere. One of the most well known scalable standards is the Scalable Video Coding (SVC) extension of H.264/MPEG-4 AVC video compression. Our researches in this topic are related to how to optimize and enhance SVC transmission over IP and broadcast networks.

With the aim at keeping a high perceived video quality using SVC, MultiPath Video Streaming (MPVS) over Video Distribution Network (VDN) comes as a promising solution to overcome the limitations of the classical single path and IP-level video streaming approaches. In [45] and [43] we proposed different approaches that couple the three SVC scalability modes (Spatial, Temporal, SNR), with the path diversity provided by VDN. Our method adapts to both the heterogeneity of end-users using the scalable video coding as well as to network bandwidth fluctuations by observing the changes of the available bandwidth over the multiple overlay paths, and updating the streaming strategy accordingly. In [44] we enhanced the precedent solutions by using the PSQA-SVC version [42] in order to get the end-user feedback in terms of QoE, which helps adapting the streaming strategy. In [48] we designed a new protocol optimizing the energy consumption when transmitting video streams. We propose to exploit the SVC coding to adapt dynamically the received video quality to the instantaneous wireless nodes’ characteristics. This is achieved through determining the number of the transmitted/received enhancements layers of an SVC video based on the wireless node context.

In [49] we proposed to support SVC over DVBT2 networks, by associating the layering architecture of both technologies in order to tackle users mobility. This association allows mobile receivers with good physical channels to decode all the SVC layers and benefit from high video quality. Meanwhile, users with poor channel conditions can at least decode the base layer and benefit from acceptable video quality. Further, we introduced a novel QoE-based adaptive mechanism for SVC layers decoding. The proposed approach selects dynamically the number of layers to decode, at the receiver side, so as to maximize the users’ perceived quality. Thus, no feedbacks or signaling messages are needed to implement the proposed algorithm. This makes it compliant with unidirectional technologies such as DVB-T2.
6. New Results

6.1. Fundamental results and algorithms: distributed planning

Participants: Eric Fabre, Loig Jézéquel.

A planning problem consists in organizing some actions in order to reach an objective. Formally, this is equivalent to finding a path from an initial state to a goal/marked state in a huge automaton. The latter is specified by a collection of resources, that may be available or not (which defines a state), and actions that consume and produce resources (which defines a transition). In the case of optimal planning, actions have a cost, and the objective is to find a path of minimal cost to the goal.

Our interest in this problem is threefold. First, it is naturally an instance of a concurrent system, given that actions have local effects on resources. Secondly, it is a weak form of an optimal control problem for a concurrent/distributed system. Finally, we are interested in distributed solutions to such problems, which is an active topic in the planning community under the name of “factored planning.”

Our previous contributions to the domain was the first optimal factored planning algorithm [61]. The main idea is to represent a planning problem as a network of interacting weighted automata, the objective being to jointly drive all of them to a target state, while minimizing the cost of their joint trajectory. We have developed and tested [68] a distributed algorithm to solve this problem, based on a weighted automata calculus, and that takes the shape of a message passing procedure. Components perform local computations, exchange messages with their neighbors, in an asynchronous manner, and the procedure converges to the path that each component should follow. The optimal global plan is thus given as a tuple of (compatible) local plans, i.e. a partial order of actions.

In 2011, we have extended this framework in two directions. In terms of modelling, first. In most planning problems, some actions consume/produce resources, but also are enabled by the presence of other resources, that they only read but not consume. We have proposed to model this feature under the form of networks of automata with read arcs. Interactions then take the form of synchronous actions, as previously, but also the form of readings: a component may only be allowed to fire some local transition if another component is in a specific state. Our distributed planning approach has then been extended to this new model of distributed systems [38].

The second improvement is algorithmic. So far, our distributed optimal planning algorithm computes all possible distributed plans, in a factored form. This contrasts with the philosophy of planning algorithms, that look for one plan only, and organize the computations to quickly reach the best plan. In other words, most planning algorithms are based on a common ground known as the A-star algorithm, a depth-first search procedure in a graph, guided by a heuristic function that estimates the remaining cost to reach the goal. We have developed a truly distributed version of this algorithm, to perform a search on a product graph. Each component runs an A-star procedure to find a path to its goal, taking into account the costs of its neighbors in order to guarantee that all components converge to local plans that are compatible and jointly optimal.

6.2. Fundamental results and algorithms: communication with messages and scenarios

Participants: Loïc Hélouët, Rouwaida Abdallah, Claude Jard, Blaise Genest.

In this paragraph, we collect our fundamental results regarding the models and algorithms we use for communicating systems, and in particular, scenarios.
A major challenge with models communicating with messages (e.g.: scenarios) is to exhibit good classes of models allowing users to specify easily complex distributed systems while preserving the decidability of some key problems, such as diagnosis, equality and intersection. Furthermore, when these problems are decidable for the designed models, the second challenge is to design algorithms to keep the complexity low enough to allow implementation in real cases.

This year, we have considered analysis for a timed extension of scenarios called Time-constrained MSCs and implementation techniques that take scenarios as an input model and output an equivalent distributed implementation.

The first part of our work is the study of Time-Constrained MSC graphs (TC-MSGS for short). Time-constrained MSCs (TC-MSCs) are simply MSCs decorated with constraints on the respective occurrence dates of events. The semantics of a TC-MSC $T$ is a dated MSC, that is a MSC where events are associated with an occurrence date. For a given TC-MSC, there can be an infinite set $L(T)$ of dated MSCs satisfying its constraints. Note however that some time-constraints in a TC-MSC may not be satisfiable, and hence $L(T)$ can simply be empty. TC-MSCs can be extended by composition mechanisms such as TC-MSC graphs. TC-MSC graphs are simply automata labeled by TC-MSC. Each path $\rho$ of a TC-MSC $G$ is associated with a TC-MSC $T_{\rho}$ obtained by concatenation of TC-MSC along $\rho$. The language $L(G) = \bigcup_{\rho \text{ path of } G} L(T_{\rho})$ of a TC-MSC Graph is then the union of all dated MSCs associated to paths of $G$. Because of inconsistent timing constraints, some path may have no possible realization (i.e $L(T_{\rho} = \emptyset$). One can even design a MSC Graph $G$ such that $L(G) = \emptyset$ - such TC-MSC graph is clearly inconsistent-. It has been shown [64] that checking whether $L(G) = \emptyset$ is an undecidable problem in general, but can be decided for the restricted subclass of regular TC-MSC graphs (that have the expressive power of event-count timed automata). We have proposed two restrictions allowing for the decision of emptiness. The first one is $K$-drift boundedness, which imposes for a fixed integer $K$ that for every $T_\rho$ there exists one dated realization such that for every pair of events $e$, $f$ appearing in the same transition of $G$, the dates of $e$ and $f$ differ by at most $K$. We have shown that $K$-drift boundedness is decidable in a symbolic and efficient way, and that for $K$−drift bounded TC-MSC graphs, emptiness is decidable [52]. This extends decidability results beyond regular specifications. The second restriction is $K$-non-zenoness, which imposes that for a fixed $K$, for every path $\rho$ of $G$, there exists one realization such that at every date $d$, at most $K$ events occur between date $d$ and $d+1$. When a TC-MSC graph is $A$-drift-bounded and $B$-non-zeno, then $L(G)$ has a regular set of representants, which opens the way for more involved model-checking applications [51].

The second part of our work is the study of realistic implementation of scenarios. The main idea is to propose distributed implementation (communicating state machines) of High-level MSCs that do not contain deadlocks, and behave exactly as the original specification. It is well known that a simple projection of a HMSC on each of its process to obtain communicating finite state machines results in an implementation with more behaviors than the original specification. An implementation of a HMSC $H$ is considered as consistent if and only if it exhibits the same prefix closed set of behaviors as $H$. We have studied how such projection with additional local controllers allows the distributed synthesized behavior to remain consistent with the original specification. This work has been implemented in our scenario prototype (see the Software section). As usually for scenarios, the synthesis algorithm works for a particular syntactic class of scenarios, namely the class of local HMSCs. Roughly speaking, in local HMSC, a decision to behave according to a scenario or another is always taken by a single participant. The deciding process need not be the same at each choice. This class is a sensible restriction of HMSCs, as distributed choices can not be implemented without additional synchronization among processes [53].

Last, we have extended existing results on diagnosis from scenarios [15]. We have shown that when a distributed implementation is instrumented with software probes that publish their observations while the system is running, and when the system is modeled as a High-level MSC, then diagnosis can be expressed as a new HMSC the executions of which are all explanations of the observation. The construction of diagnosis can be performed offline or online, and we have considered the conditions under which online diagnosis can run with finite memory.
6.3. Fundamental results and algorithms: timed models

Participants: Claude Jard, Aurore Junier, Akshay Sundararaman.

Our works in that subject concern Time Petri Nets (TPNs) and Network Calculus. With TPNs, we are particularly interested in symbolic unfoldings (extended with parameters). Possible applications are supervision of distributed timed systems [8] and testing of concurrent systems (work done in collaboration with Stefan Haar, INRIA-LSV in Cachan).

The article [26] was made during the internship of the master degree of Aurore Junier under the supervision of Anne Bouillard (ENS Paris). It uses a (min, plus)-algebra to define a worst-case delay bound for networks where flows have fixed priorities.

After that, we studied a well-known problem: detection of congestion and failure in networks. The idea was to find an efficient and deterministic method that is very reactive and takes little memory space. Such a method does not exists for now and is an important issue for Alcatel-Lucent. We achieved a solution that solves this problem based on the analysis of flows behaviour. This work is part of the work done within the Alcatel-Lucent-Inria joint lab and a patent is being established.

We are also studying the way buffers of routers can increase. The objective is to find a method that can detect if sizes of buffers can dangerously increase on a defined topology. We start by looking at Link State Advertisements (LSA) in the OSPF protocol. We represent the topology and a part of the protocol by a Time Petri Net and try to infer parameters ensuring stability.

6.4. Fundamental results and algorithms: dynamic epistemic logic

Participants: Guillaume Aucher, François Schwarzentruber.

Dynamic Epistemic Logic (DEL) deals with the representation and the study of knowledge and belief change in a multi-agent setting. The core representative task of this logical framework can be split up in three parts: 1/ the initial global state of the distributed system, 2/ an event occurring in this system, 3/ a product update taking as argument these two representations and yielding a new representation of the new global state of the distributed system. Therefore, we can express uniformly within the DEL framework epistemic statements about:

(i) what is true about an initial state
(ii) what is true about an event occurring in this initial state
(iii) what is true about the resulting state after the event has occurred.

We axiomatized within the DEL framework what we can infer about (iii) given (i) and (ii), what we can infer about (ii) given (i) and (iii), and what we can infer about (i) given (ii) and (iii). Given three logical formulas \( \phi, \phi', \phi'' \) describing respectively (i), (ii) and (iii), we also showed how to build three formulas that capture respectively all the information which can be inferred about (iii) from \( \phi \) and \( \phi' \), all the information which can be inferred about (ii) from \( \phi \) and \( \phi'' \), and all the information which can be inferred about (i) from \( \phi' \) and \( \phi'' \).

We showed how our results extend to other modal logics than the minimal modal logic \( K \). These results are to appear in [9] and [10]. In [19], we also provided a tableau method deciding whether such inferences are valid. We implemented it in LOTRECScheme and showed that this decision problem is NEXPTIME-complete. This work contributes to the proof theory and the study of the computational complexity of DEL which have rather been neglected so far.

Application to fault localization in IMS network (see the UNIVERSELF project) has started. The various agents involved in an IMS network (clients, assistance, administrators...) have a partial view of the network and so need to communicate their partial knowledge of the network to each other in order to localize the fault in the network (each communication having possibly a different cost). One of the main problems is to determine which communication should occur and which agent should be queried so that the fault is eventually localized. This problem can naturally be expressed in the DEL framework. We have shown how the initial state of an IMS network representing the knowledge of each agent can be represented by a particular kind of epistemic
model (i) and how the desired state where the fault is localized can be expressed by a logical formula (iii). The problem amounts to determining which communication or sequence of communications should occur (ii) so that one passes from the initial epistemic model (i) to another epistemic model where the fault is localized (iii), and also to determine if such a communication or sequence of communications is possible. We have focused so far on the case of a single communication, but we plan to extend it to a sequence of communications. Further theoretical work still needs to be done to address the issue of sequential communication.

In parallel to this work, we also axiomatized different notions of knowledge and belief which are defined by means of a 'sphere' semantics. This work is the result of an invited contribution and is to appear in [48].

6.5. Fundamental results and algorithms: statistical model checking

**Participants:** Sean Sedwards, Cyrille Jégourel, Axel Legay.

Our work on statistical model checking (SMC) avoids an explicit representation of the state space by building a statistical model of the executions of a system and giving results within confidence bounds. The key challenges of this approach are to reduce the length (simulation steps and cpu time) and number of simulation traces necessary to achieve a result with given confidence. Rare properties pose a particular problem in this respect, since they are not only difficult to observe but their probability is difficult to bound. A further goal is to make a tool where the choice of modelling language and logic are flexible.

We have developed the prototype of a compact, modular and efficient SMC platform which we have named PLASMA (PLatform for Statistical Model checking Algorithms). PLASMA incorporates an efficient discrete event simulation algorithm and features an importance sampling engine that can reduce the necessary number of simulation runs when properties are rare. We have found that PLASMA performs significantly better than PRISM (the de facto reference probabilistic model checker) when used in a similar mode: PLASMA's simulation algorithm scales with a lower order and can handle much larger models. When using importance sampling, PLASMA's performance with rare properties is even better.

6.6. Fundamental results and algorithms: quantitative model checking and quantitative specification Theories

**Participants:** Uli Fahrenberg, Axel Legay.

Model checking of systems deals with the question whether a given model of a computer system satisfies the properties one might want to require of it. This is a well-established and successful approach to formal verification of safety-critical computer systems.

When the models of the systems contain quantitative information, the model checking problem becomes complicated by the fact that in most cases, quantitative properties of the systems do not need to be satisfied exactly. Indeed, the model or the properties might be subject to measurement error, or probabilistic information might only be an approximation. In this case, it is of little use to know whether or not a model satisfies a specification precisely; what is needed instead is a notion of satisfaction distance: a measure which can assess to which extent a quantitative model satisfies a quantitative specification.

In other words, what is needed is a notion of satisfaction which is robust in the sense that small deviations in the model or the specification only lead to small changes in the outcome of the model checking question. We have published work on such distances in the papers [37], [34].

For more elaborate reasoning about distributed systems or systems-of-systems, an important role is played by specification theories. Such systems are often far too complex to reason about, or model-check, as a whole, and additionally they might be composed of a large number of components which are implemented by different vendors. Hence one needs methods for compositional reasoning, which allow to infer properties of a system from properties of its components, and for incremental design, which allow to synthesize and refine specifications in a step-wise manner.
Such specification theories are by now well-established e.g. in the incarnations of interface theories and modal transition systems. Additionally to defining a formalism for describing and model-checking specifications, they provide notions of refinement of specifications, logical conjunction of specifications, and structural composition and quotient.

When the models and specifications contain quantitative information, all the above notions need to be made robust. One needs to introduce a quantitative version of refinement, and the operations on specifications need to be continuous with respect to refinement distance: compositions of specifications with small refinement distance need themselves to have small refinement distance. We have published work on these issues in the papers [21], [35]; additionally, two other papers within this research area are currently under submission.

6.7. Specific studies: Web services orchestrations

Participants: Ajay Kattepur, Albert Benveniste, Claude Jard.

Web services orchestrations and choreographies refer to the composition of several Web services to perform a co-ordinated, typically more complex task. We decided to base our study on a simple and clean formalism for WS orchestrations, namely the Orc formalism proposed by Jayadev Misra and William Cook [71].

Main challenges related to Web services QoS (Quality of Service) include: 1/ To model and quantify the QoS of a service. 2/ To establish a relation between the QoS of queried Web services and that of the orchestration (contract composition); 3/ To monitor and detect the breaching of a QoS contract, possibly leading to a reconfiguration of the orchestration. Typically, the QoS of a service is modeled by a contract (or Service Level Agreement, SLA) between the provider and consumer of a given service. To account for variability. In previous years, we proposed soft probabilistic contracts specified as probabilistic distributions involving the different QoS parameters; we studied contract composition for such contracts; we developed probabilistic QoS contract monitoring; and we studied the monotonicity of orchestrations; an orchestration is monotonic if a called service improves its performance, then so does the overall orchestration.

This year, in the framework of the Associated Team FOSSA with the University of Texas at Austin (John Thywissen (PhD), Jayadev Misra and William Cook), we have extended our approach to general QoS parameters, i.e., beyond response time. We now encompass composite parameters, which are thus only partially, not totally, ordered. We have developed a general algebra to capture how QoS parameters are transformed while traversing the orchestration and we have extended our study of monotonicity. Finally, we have developed corresponding contract composition procedures. John Thywissen (from UT Austin) and Ajay Kattepur have started extending the Orc language and execution engine to support QoS according to our theory. This extension mainly consists in 1/ providing a rich type system to declare QoS domains and related algebra, and 2/ providing a new operator for Orc that allows for selecting competing returns from different sites on the basis of their QoS. A journal paper is under revision.

A key task in extending Orc for QoS was to extend the Orc engine so that causalities between the different site calls are made explicit at run time while execution progresses. This benefits from our previous work on Orc semantics, but a new set of rules has been proposed to generate causalities in an efficient way, by covering new features of the language. This is joint work of Claude Jard, Ajay Kattepur and John Thywissen from Austin. A publication is in preparation.

Besides this main line of work, other topics have been addressed by Ajay Kattepur as part of his thesis.

- In [41], we study variability of composite services. We model variability as a feature diagram (FD) that captures all valid configurations of its orchestration. Then, we apply pair-wise testing to sample the set of all possible configurations to obtain a concise subset. Finally, we test the composite service for selected pairwise configurations for a variety of QoS metrics such as response time, data quality, and availability. Using two case studies, Car crash crisis management and e-Health management, we demonstrate that pairwise generation effectively samples the full range of QoS variations in a dynamic orchestration. The pairwise sampling technique eliminates over 99% redundancy in configurations, while still calling all atomic services at least once.
• Web services orchestrations conventionally employ exhaustive comparison of runtime quality of service (QoS) metrics for decision making. The ability to incorporate more complex mathematical packages is needed, especially in case of workflows for resource allocation and queuing systems. By modeling such optimization routines as service calls within orchestration specifications, techniques such as linear programming can be conveniently invoked by non-specialist workflow designers. Leveraging on previously developed QoS theory, we propose the use of a high-level flexible query procedure for embedding optimizations in languages such as Orc. The Optima site provides an extension to the sorting and pruning operations currently employed in Orc. Further, the lack of an objective technique for consolidating QoS metrics is a problem in identifying suitable cost functions. We use the analytical hierarchy process (AHP) to generate a total ordering of QoS metrics across various domains. With constructs for ensuring consistency over subjective judgements, the AHP provides a suitable technique for producing objective cost functions. Using the Dell Supply Chain example, we demonstrate the feasibility of decision making through optimization routines, especially when the control flow is QoS dependent. This work was published in [39].

• With web services quality of service (QoS) modeled as random variables, the accuracy of sampled values for precise service level agreements (SLAs) come into question. Samples with lower spread are more accurate for calculating contractual obligations, which is typically not the case for web services QoS. Moreover, the extreme values in case of heavy-tailed distributions (e.g., 99.99 percentile) are seldom observed through limited sampling schemes. To improve the accuracy of contracts, we propose the use of variance reduction techniques such as importance sampling. We demonstrate this for contracts involving demand and refuel operations within the Dell supply chain example. Using measured values, efficient forecasting of future deviation of contracts may also be performed. A consequence of this is a more precise definition of sampling, measurement and variance tolerance in SLA declarations. This work was published in [40].

6.8. Specific studies: active documents and web services

Participants: Albert Benveniste, Loïc Hélouët, Benoît Masson.

Active Documents have been introduced by the GEMO team at INRIA Futurs, headed by Serge Abiteboul, mainly through the language Active XML (or AXML for short). AXML is an extension of XML which allows to enrich documents with service calls or sc’s for short. These sc’s point to web services that, when triggered, access other documents; this materialization of sc’s produces in turn AXML code that is included in the calling document. One therefore speaks of dynamic or intentional documents. In the past years, we have collaborated with the GEMO team to study a distributed version of their language.

This year, we have addressed the problem of distributed documents from a different point of view. Starting from our knowledge of distributed active XML (DAXML), we have first proposed a Petri Net semantics for a subset of DAXML [43], and then considered compositionality issues [54]. Compositionality in services can be addressed in several ways: first one have to ensure that modules that provide services an modules that use them agree on the data that they exchange. This notion is called composability of modules. However, composability does not ensure that a service always terminates (i.e., it returns a result to the caller) when it is invoked with appropriate data. Composability plus termination of services is called compatibility. We have shown that under some restrictions on the recursion in active documents, on the data, and upon the assumption that services use positive guards, composability is decidable. This work has also helped us isolate the core idea behind active documents, and propose a model for them called Docnets. Docnets are dynamic Petri nets which places are typed, which transitions are guarded computable type transformations, and which can receive new tokens from their environment. Docnet modules compose well, and if their closure by type transformation is finite, their compatibility is decidable. This work led to a publication [43].

Within the context of the DST associated team, we have proposed a new model, that combines arbitrary numbers of finite workflows, hence allowing for the definition of sessions. Sessions is a central paradigm in web-based systems. As messages exchange between two sites need not follow the same route over the net, a site can not rely on the identity of machines to uniquely define a transaction. This unique identification is
essential, as commercial site, for instance, need to manage several interactions at a given time. The current trend, as in BPEL, is to associate a unique identifier to each session. Modeling realistic session hence often forces to include session counters, and hence render most of properties undecidable. The session formalism studied in 2011 can be seen as a mix of BPEL and ORC elements, but was designed to keep several properties decidable. The strength of this formalism is to allow designing systems that use sessions without the obligation to provide identifiers. The formalism has the expressive power of reset Petri nets for which coverability is decidable. This is sufficient to decide whether a set of agent can be found in some bad configuration during the lifetime of a system. This joint work with Ph. Darondeau from the S4 Team, and with M. Mukund from the Chennai Mathematical Institute led to a publication in the ATVA conference [28].

Our last work on Web-services was the development of an experimental platform. During his post-doc, Benoît Masson has designed a distributed Active XML engine, which can be distributed over a network. We have built a lightweight experimentation platform, made of four linux machines, that run DAXML services and communicate with one another. Simultaneously, R. Abdallah has designed a synthesis tool to generate REST services from High-level Message Sequence Charts. These services were successfully tested on the platform.

6.9. Specific studies: security and privacy

Participants: Guillaume Aucher, Blaise Genest.

We have worked on three parallel lines of research related to security and privacy. The first line deals with problems of deleguation and revocation in distributed systems. The second line deals with problems of compliance of a system with respect to a privacy regulation expressed in a language combining epistemic, deontic and dynamic modalities. The third line tackles the minimal information needed at runtime to e.g. break in a (stochastic) system.

6.9.1. Delegation and revocation in distributed systems

Together with Steve Barker from King’s College London, Guido Boella from the University of Torino, Valerio Genovese and Leon van der Torre from the University of Luxembourg, we defined a (sound and complete) propositional dynamic logic to specify and reason about delegation and revocation schemes in distributed systems. This logic describes formally a family of delegation and revocation models that are based on the work of [65]. We extended our logic to accommodate an epistemic interpretation of trust. What emerges from this work is a rich framework of formally well-defined delegation and revocation schemes that accommodates an important trust component. In particular, we showed how to automatically reason about whether an agent is authorized to do an operation on an object and about the authorization policy resulting from the execution of a sequence of actions. We used our logical framework to give a formal account of eight different types of revocation schemes informally introduced in previous literature. This work is published in [18].

6.9.2. Privacy policy with modal logic: the dynamic turn

As explained in Section 6.4, we want to define a logical language to specify privacy policies which is close to the natural language. In general, privacy policies can be defined either in terms of permitted and forbidden knowledge, or in terms of permitted and forbidden actions. For example, it may be forbidden to know the medical data of a person, or it may be forbidden to disclose these data. Implementing a privacy policy based on permitted and forbidden actions is relatively easy, since we can add a filter on the system checking the outgoing messages. Such a filter is an example of a security monitor. If the system attempts to send a forbidden message, then the security monitor blocks the sending of that message. However, the price to pay for this relatively straightforward implementation is that it is difficult to decide which actions are permitted or forbidden so that a piece of information is not disclose. We are therefore interested in privacy policies expressed in terms of permitted and forbidden knowledge. Expressing a privacy policy in terms of permitted and forbidden knowledge is relatively easy, since it lists the situations, where, typically, it may not be permitted to know some sensitive information. Implementing a privacy policy based on permitted and forbidden knowledge is quite difficult, since the system has to reason about the relation between permitted knowledge and actions. The challenge is that the exchange of messages changes the knowledge, and the
security monitor therefore needs to reason about these changes. This inference problem is already non-trivial with a static privacy policy, and becomes challenging when privacy policies can change over time. Together with Guido Boella and Leon van der Torre, we therefore introduced a dynamic modal logic that permits not only to reason about permitted and forbidden knowledge to derive the permitted actions, but also to represent explicitly the declarative privacy policies together with their dynamics. The logic can be used to check both regulatory and behavioral compliance, respectively by checking that the permissions and obligations set up by the security monitor of an organization are not in conflict with the privacy policies, and by checking that these obligations are indeed enforced. We also showed that the complexity of the model checking problem is quadratic in the size of the model and the formula and provided the corresponding model-checking algorithms. This work is published in [11].

6.9.3. Minimal information needed
Together with Nathalie Bertrand from Vertecs, we tackle the problem of the minimal information a user needs at runtime to achieve a simple goal, modeled as reaching an objective with probability one [25]. The natural question is then to minimize the additional information the user needs to fulfill her objective. This optimization question gives rise to two different problems, whether we consider to minimize the worst case cost, or the average cost. On the one hand, concerning the worst case cost, we show that efficient techniques from the model checking community can be adapted to compute the optimal worst case cost and give optimal strategies for the users. On the other hand, we show that the optimal average price (a question typically considered in the AI community) cannot be computed in general, nor can it be approximated in polynomial time even up to a large approximation factor. Following this negative results, we investigate with P.S. Thiagarajan’s group at NUS, Singapore basic algorithms of the AI community to infer the exact probability in (compact) stochastic systems. We proposed in [45] a simple parametrized extension of the usual Factored Frontier algorithm in order to choose the desired accuracy of the algorithm, at the cost of additional but manageable computations. We showed its benefit when dealing with biological pathways.

6.10. Specific studies: network maintenance
Participants: Eric Fabre, Carole Hounkonnou.

This work represents part of our activities within the research group “High Manageability,” supported by the common lab of Alcatel-Lucent Bell Labs (ALBLF) and INRIA. It concerns a methodology for the graceful shut down and restart of routers in OSPF networks, one of the core protocols of IP networks. A methodology has been proposed to safely switch off the software layer of a router while still maintaining this router in the forwarding plane: the router still forwards packets, but is not able to adapt its routing table to changes in network conditions or topology. Nevertheless, it is possible to check whether this frozen router is harmless or can cause packet losses, through a centralized or distributed algorithm. And if ever it puts the network at risk, minimal patches can be set up temporarily until the router comes back to normal a activity. This avoids running twice a global OSPF update at all nodes (once for shutdown of the equipment, one for restart). There is a patent project on this activity, that we don’t detail more here.

6.11. Specific studies: network and service diagnosis
Participants: Eric Fabre, Carole Hounkonnou.

This work represents part of our activities within the research group “High Manageability,” supported by the common lab of Alcatel-Lucent Bell Labs (ALBLF) and INRIA. It is also supported by the UniverSelf EU integrated project, and conducted in relation with Orange Labs. The objective is to develop a framework for the joint diagnosis of networks and of the supported services. We are aiming at a model-based approach, in order to taylor the methods to a given network instance and to follow its evolution. We also aim at active diagnosis methods, that collect and reason on alarms provided by the network, but that can also trigger tests or the collection of new observations in order to refine a current diagnosis.
In 2011, the main effort was dedicated to a key and difficult part of this approach: the definition of a methodology for self-modelling. This consists in automatically building a model of the monitored system, by instantiating generic network elements. There are several difficulties to address:

- The model must capture several layers, from the physical architecture up to the service architecture and its protocols. As a case-study, we have chosen VoIP services on an IMS network, deployed over a wired IP network.
- The model should be hierarchical, to allow for multiscale reasoning, and to reflect the intrinsic hierarchical nature of the managed network.
- The model should be generic, i.e. obtained by assembling component instances coming from a reduced set of patterns, just like a text is obtained by assembling words.
- The model should be adaptive, to capture the evolving part of the network (e.g. introduction of new elements) but also its intrinsically dynamic nature (e.g. opened/closed connections).
- The model should display the hierarchical dependency of resources, specifically the fact that lower-level resources are assembled to provide a support to a higher level resource or functionality.
- The model should allow progressive discovery and refinement: for a matter of size, it is not possible to first build a model of the complete network and then monitor it; one must adopt an approach where the model is build on-line, and where the construction is guided by the progress of the diagnosis algorithms.

The first elements of a methodology achieving these objectives have been designed in 2011. The next efforts will aim at refining the grammar of this model, for our specific case study, and at developing the dedicated diagnosis algorithms. For the latter, we envision a new setting of hierarchical and generic Bayesian networks, in order to capture the dependencies between network elements at different granularities.
6. New Results

6.1. BlobSeer and Map-Reduce programming

6.1.1. BlobSeer-based cloud storage

Participants: Alexandra Carpen-Amarie, Alexandru Costan, Gabriel Antoniu, Luc Bougé.

As data volumes generated and processed by such applications increase, a key requirement that directly impacts the adoption rate of the Cloud paradigm is efficient and reliable data management. In this context, we investigate the requirements of Cloud data services in terms of data-transfer performance and access patterns and we explore the ways to leverage and adapt existing data-management solutions for Cloud workloads. We aim at building a Cloud data service both compatible with state-of-the-art Cloud interfaces and able to deliver high-throughput data storage.

To achieve this goal, we developed a file system layer on top of BlobSeer, which exposes a hierarchical file namespace enhanced with the concurrency-optimized BlobSeer primitives. Furthermore, we integrated the BlobSeer file system as a backend for Cumulus, an efficient open-source Cloud storage service. We validated our approach through extensive evaluations performed on Grid’5000. We devised a set of synthetic benchmarks to measure the performance and scalability of the Cumulus system backed by BlobSeer, showing it can sustain high-throughput data transfers for up to 200 concurrent clients.

Next, we explored the advantages and drawbacks of employing Cloud storage services for distributed applications that manage massive amounts of data. We investigated two types of applications. We relied on an atmospheric phenomena modeling application to conduct a set of evaluations in a Nimbus Cloud environment. This application is representative for a large class of simulators that compute the evolution in time set of parameters corresponding to specific points in a spatial domain. As a consequence, such applications generate important amounts of output data. We evaluated an S3-compliant Cloud storage service as a storage solution for the generated data. To this end, we employed distributed Cumulus services backed by various storage systems. The reason for targeting this approach is that storing output data directly into the Cloud as the application progresses can benefit higher-level applications that further process such simulation data. As an example, visualization tools need to have real-time access to output data for analysis and filtering purposes.

We built an interfacing module to enable the application to run unmodified in a Cloud environment and to send output data to an S3-based Cloud service. Our experiments show that distributed Cumulus backends, such as BlobSeer or PVFS, sustain a constant throughput even when the number of application processes that concurrently generate data becomes 3 times higher than the number of storage nodes.

6.1.2. Optimizing Intermediate Data Management in MapReduce Computations

Participants: Diana Moise, Gabriel Antoniu, Luc Bougé.

MapReduce applications, as well as other cloud data flows, consist of multiple stages of computations that process the input data and output the result. At each stage, the computation produces intermediate data that is to be processed by the next computing stage. We studied the characteristics of intermediate data in general, and we focused on the way it is handled in MapReduce frameworks. Our work addressed intermediate data at two levels: inside the same MapReduce job, and during the execution of pipeline applications.
We focused first on efficiently managing intermediate data generated between the “map” and “reduce” phases of MapReduce computations. In this context, we proposed to store the intermediate data in the distributed file system used as underlying backend. In this direction, we investigated the features of intermediate data in MapReduce computations and we proposed a new approach consisting in storing this kind of data in a DFS. The major benefit of this approach is better illustrated when considering failures. Existing MapReduce frameworks store intermediate data on nodes local disk. In case of failures, intermediate data produced by mappers can no longer be retrieved and processed further by reducers. the solution of most frameworks is to reschedule the failed tasks and to re-generate all the intermediate data that was lost because of failures. This solution is costly in terms of additional execution time. With our approach of storing intermediate data in a DFS, we avoid the re-execution of tasks in case of failures that lead to data loss. As storage for intermediate data, we considered BSFS as being a suitable candidate for providing for the requirements of intermediate data: availability and high I/O access. The tests we performed in this context, measured the impact of using a DFS as storage for intermediate data instead of the local-disk approach. We then assessed the performance of BSFS and HDFS when serving as storage for intermediate data produced by several MapReduce applications.

We then considered another type of intermediate data that appears in the context of pipeline MapReduce applications. In order to speed-up the execution of pipeline MapReduce applications (applications that consist of multiple jobs executed in a pipeline) and also, to improve cluster utilization, we proposed an optimized Hadoop MapReduce framework, in which the scheduling is done in a dynamic manner. We introduced several optimizations in the Hadoop MapReduce framework in order to improve its performance when executing pipelines. Our proposal consisted mainly in a new mechanism for creating tasks along the pipeline, as soon as the tasks’ input data becomes available. As our evaluation showed, this dynamic task scheduling leads to an improved performance of the framework, in terms of job completion time. In addition, our approach ensures a more efficient cluster utilizations, with respect to the amount of resources that are involved in the computation.

We evaluated both approaches for intermediate data through a set of experiments on the Grid’5000 testbed. Preliminary results show the scalability and efficiency of our proposals, as well as additional benefits brought forward by our approach.

6.1.3. A-Brain: Perform genetic and neuroimaging data analysis in Azure clouds

**Participants:** Radu Tudoran, Alexandru Costan, Gabriel Antoniu, Louis-Claude Canon.

Joint genetic and neuroimaging data analysis on large cohorts of subjects is a new approach used to assess and understand the variability that exists between individuals. This approach has remained poorly understood so far and brings forward very significant challenges, as progress in this field can open pioneering directions in biology and medicine. As both neuroimaging- and genetic-domain observations represent a huge amount of variables (of the order of $10^6$), performing statistically rigorous analyses on such amounts of data represents a computational challenge that cannot be addressed with conventional computational techniques.

In order to perform an accurate analysis we need to provide a programming platform and a high throughput storage. The target infrastructure is the Azure clouds. Hence we have adapted the BlobSeer storage approach for Azure, thus providing a new way to store data in clouds, that federates the local storage space from computational nodes into a uniform shared storage, called TomusBlobs. Using this storage system as a storage backend, we have built a MapReduce prototype for Azure clouds. This MapReduce system, called TomusMapReduce -TMR, is used to perform the simulation of the joint genetic and neuroimaging application. For validating the framework, a toy application that simulate the data access and computation patterns of the real application, was used. The next step, after the evaluation of the framework, that has just began, consists in replacing the toy application with the real one and the scaling of the framework in the limit allowed by the cloud provider. In addition a demo for this project is in progress, that will consists in providing a visualization tool for the framework. This will be used to intuitively represent the results for the simulation of the scientific application, this being useful both for better presenting the project to interested parties and for the researchers from bioinformatics.

6.2. Efficient VM management in clouds

**Participants:** Alexandru Costan, Alexandra Carpen-Amarie, Gabriel Antoniu.
Infrastructure as a Service (IaaS) cloud computing allows users to lease computational resources from the cloud provider’s datacenter for a short time by deploying virtual machines (VMs) on these resources. This model raises new challenges in the design and development of IaaS middleware. One of those challenges is the need to deploy a large number (hundreds or even thousands) of VM instances simultaneously. Once the VM instances are deployed, another challenge is to simultaneously take a snapshot of many images and transfer them to persistent storage to support fault tolerance and management tasks, such as suspend-resume and migration. With datacenters growing rapidly and configurations becoming heterogeneous, it is important to enable efficient concurrent deployment and snapshotting that are at the same time hypervisor independent and ensure a maximum compatibility with different configurations.

We addressed these challenges by proposing a virtual file system specifically optimized for virtual machine image storage [19]. It is based on a lazy transfer scheme coupled with object versioning that handles snapshotting transparently in a hypervisor-independent fashion, ensuring high portability for different configurations. Large-scale experiments on hundreds of nodes demonstrate excellent performance results: speedup for concurrent VM deployments ranges from a factor of 2 up to 25, with a reduction in bandwidth utilization of as much as 90% [18]. We implemented this deployment scheme in the Nimbus cloud and presented a demo illustrating it at the Grid’5000 School [26].

Given the dynamic nature of IaaS clouds and the long runtime and resource utilization of scientific applications, an interesting use-case for the multi-snapshotting techniques is for efficient checkpoint-restart. We introduced an approach that leverages VM disk-image multi-snapshotting and multi-deployment inside checkpoint-restart protocols running at guest level in order to efficiently capture and potentially roll back the complete state of the application, including file system modifications. This framework is specifically optimized for tightly-coupled scientific applications that were written using a message passing system (in particular MPI) and need to be ported to IaaS clouds. Our solution introduces a dedicated checkpoint repository that is able to efficiently take incremental snapshots of the whole disk attached to the virtual machine instances, thus offering support to use any checkpointing protocol that can save the state of processes into files, including application level mechanisms, where the process state is managed by the application itself, and process-level mechanisms, where the process state is managed transparently at the level of the message passing library. Experiments on the G5K testbed show substantial improvement for MPI applications over existing approaches, both for the case when customized checkpointing is available at application level and the case when it needs to be handled at process level.

We integrated this checkpointing scheme inside the Nimbus cloud with some promising preliminary results. We plan to complement the existing solution with live incremental snapshotting using asynchronous background transfers for high checkpointing efficiency and with adaptive prefetching to achieve high restart efficiency.

6.3. Cloud data storage management

6.3.1. Autonomic storage for cloud services

Participants: Alexandru Costan, Alexandra Carpen-Amarie, Gabriel Antoniu, Florin Pop, Ciprian Dobre, Elena Apostol.

A means to achieve performance improvement and resource-usage optimization in cloud storage systems consists in enabling an autonomic behavior based on introspection. Self-adaptation incurs a high degree of complexity in the configuration and tuning of the system, with possible repercussions on its availability and reliability. To address these challenges we introduced in BlobSeer in [11] a three-layered architecture designed to identify and generate relevant information related to the state and the behavior of the system, based on the MonALISA monitoring framework. Such information is then expected to serve as an input to a higher-level self adaptation engine. These data are yielded by an (1) introspection layer, which processes the raw data collected by a (2) monitoring layer. The lowest layer is represented by the (3) instrumentation code that enables BlobSeer to send monitoring data to the upper layers.
A first approach to leverage the introspection framework aims at enhancing BlobSeer with self-configuration capabilities, as a means to support storage elasticity through dynamic deployment of data providers. This solution enables the data providers to scale up and down depending on the detected system’s needs. The component we designed adapts the storage system to the environment by contracting and expanding the pool of storage providers based on the system’s load. The key idea of this component is the automatic decision that has to be made on how many resources the system needs to operate normally while keeping the resources utilization down to a minimum. This problem is addressed by using a test-decided heuristic based on the monitoring data. The introspective architecture has been evaluated on the Grid’5000 testbed, with experiments that prove the feasibility of generating relevant information related to the state and the behavior of the system.

We plan to use the introspective BlobSeer to develop a distributed data aggregation service. Its primary goal will be to serve as a repository backend for complex analysis and automatic mining of scientific data. Another direction that will be explored is to use the introspective BlobSeer as a cloud-based storage layer for sensitive context data, collected from a vast amount of sources: from smartphones to sensors located in the environment. Clouds are perfect candidates to handle the storage and aggregation of such data for even larger context-aware applications. Such solutions rely on more relaxed storage capabilities than traditional relational databases (eventual consistency suffices for example). This, combined with the high concurrency support and the flexible storage schema make BlobSeer a suitable candidate for the storage layer. We plan to develop a new layer on top of BlobSeer targeting context aware applications. At the logical level, this layer will provide transparency, mobility, real-time guarantees and access based on meta-information. At the physical layer, the most important capability will rely on BlobSeer’s elasticity to scale up and down according to real-time usage, in order to reduce the costs within the Cloud.

6.3.2. Managing data access on Clouds through security policies

Participants: Alexandru Costan, Alexandra Carpen-Amarie, Gabriel Antoniu.

With the emergence of Cloud computing, there has been a great need to provide an adequate security level in such environments, as they are vulnerable to various attacks. Malicious behaviors such as Denial of Service attacks, especially when targeting large-scale data management systems, cannot be detected by typical authentication mechanisms and are responsible for drastically degrading the overall performance of such systems.

In [14] we proposed a generic security management framework allowing providers of Cloud data management systems to define and enforce complex security policies. The generality of this approach comes from the flexibility both in terms of supporting custom security scenarios and interfacing with different Cloud storage systems. This security framework is designed to detect and stop a large array of attacks defined through an expressive policy description language and to be easily interfaced with various data management systems. We introduced a modular architecture consisting of three components. The Policy Management module represents the core of the framework, where security policies definition and enforcement takes place. This module is completely independent of the Cloud system, as its input only consists in user activity events monitored from the system. The User Activity History module is a container for monitoring information describing users’ actions. It collects data by employing monitoring mechanisms specific to each storage system and makes them available for the Policy Management module. The Trust Management module incorporates data about the state of the Cloud system and provides a trust value for each user based on his past actions. The trust value identifies a user as a fair or a malicious one. Furthermore, the trust values enable the system to take custom actions for each detected policy violation, by taking into account the history of each user.

As a case study, we applied the proposed framework to BlobSeer. We defined a specific policy to detect DoS attacks in BlobSeer and we evaluated the performance of our framework through large scale experiments on the Grid’5000 testbed. The results show that the Policy Management module meets the requirements of a data storage system in a large-scale deployment: it was able to deal with a large number of simultaneous attacks and to restore and preserve the performance of the target system.

As a next step we will focus on more in-depth experiments involving the detection of various types of attacks in the same time. Moreover, we will investigate the limitations of our Security Management framework, with
respect to the accuracy of the detection in the case of more complex policies, as well as the probability and the impact of obtaining false positive or false negative results. Another research direction is to further develop the Trust Management component of the security management framework and study the impact it has on the Policy Enforcement decisions for complex scenarios.

6.4. Storage architecture and adaptive consistency for clouds

**Participants:** Houssem-Eddine Chihoub, Gabriel Antoniu.

As more and more applications are becoming data-intensive, the design of a scalable storage architecture providing a huge file sharing and fine grain access with high throughput under heavy concurrency is a timely and relevant challenge.

In [24], we introduce a storage architecture for Cloud computing. This architecture proposes efficient and scalable storage support for both VM images and application data. Our architecture relies on BlobSeer [12], a data sharing platform optimized for concurrent accesses, as a basic storage backend enhanced in term of quality of service and efficiency by GloBeM tool [61] that rely on behavior modeling and monitoring to avoid bad case scenarios. The architecture uses this approach to have a better and efficient storage VM images, that allow a faster image deployment and efficient versioning and snapshotting. Furthermore, the architecture provides a platform to store, manage, and share cloud application data allowing several key features to clouds such as storage elasticity.

The main aim is to provide high availability and good scalable performance at low cost. This is justified by the growing need of data-intensive applications for managing huge sets of data replicated over several data centers. In order to provide good performance and high availability, data replication is mandatory. But this generates the issue of replicas consistency as shown by the CAP theorem [52]. To achieve the aforementioned goals, relaxing consistency rules is unavoidable. On the other hand, opting for weaker consistency, all the time, can be too costly.

In current work we leverage the trade-off between consistency and availability and performance. We are investigating an approach that changes the consistency level at runtime considering system and application needs. In order to choose the most suitable consistency level, our approach monitor the storage system and collect useful information, such as network load and applications access patterns, that enable the system to estimate the amount of expected stale reads.

6.5. Modelling cloud storage performance

**Participants:** Daniel Higuero, Louis-Claude Canon, Alexandru Costan, Gabriel Antoniu.

The objective of this research direction is to provide comprehensive performance models for storage systems. Their role is to capture how the system components interact for different usage patterns (number of reads or writes). The objective is to determine the incurred costs in terms of storage space and efficiency for a given workload.

One application of this model consists in dynamically adjusting the parameters of the storage system as required in an autonomic approach. For this purpose, it is necessary to identify the characteristics of the storage system for meeting a given level of requirements. Progress has been made on this part during the 3-month visit of Daniel Higuero (University Carlos III, Madrid). A preliminary performance model currently predicts the available bandwidth when multiple concurrent transfers occur. This model serves as a basis for a dimensioning strategy that is formulated through a linear program.

This approach has further been complemented with an offline analysis of several traces of the BlobSeer storage system when it is used as a backend for MapReduce applications. Mining this information in an automated fashion allowed to detect the different trade-offs that influence a BlobSeer deployment: time required to execute the application vs. number of machines used by the storage system, communication costs vs. space usage. The final goal is to tune BlobSeer for specific applications. The proposed strategy is currently being evaluated.
Future directions are directed towards refining the proposed model. Several parameters significantly impact the performance of storage systems such as the redundancy mechanism, the data placement strategy or disk-related effects. As a first step, experiments for assessing the quality of finer models will be designed. Ultimately, we aim to capture the I/O variability of storage systems, in particular in the context of the cloud.

This work will enable new collaborations. It is planned to work on the models mentioned above with the Mescal INRIA team in the context of a collaboration between the MapReduce ANR project and the Songs ANR project. Moreover, in the framework of the MapReduce project, we expect to work on a performance model for designing decision algorithms that are required by the component-based MapReduce framework that is developed in the GRAAL/Avalon INRIA team. Finally, the GRAAL/AVALON team works on scheduling algorithms that could beneficially profit from a storage performance model.

### 6.6. Scalable I/O and visualization for post-petascale HPC simulations

**Participants:** Matthieu Dorier, Gabriel Antoniu.

In the context of the Joint INRIA/UIUC Laboratory for Petascale computing (JLCP), we are addressing the new challenges related to I/O, data analysis and visualization for extreme-scale simulations. As HPC resources approaching millions of cores become a reality, a growing challenge in maintaining high performance is the presence of high variability in the effective throughput of codes performing I/O operations. Since I/O is mainly performed for the purpose of subsequent data analysis and visualization, another way to limit the impact of I/O performance on scientific discovery consists in enabling in-situ visualization. This brings again new challenges such as how to efficiently couple large-scale simulations with visualization software.

We started the development of Damaris, a middleware targeting multicore SMP nodes to efficiently address the problems mentioned above. Damaris has been evaluated with the CM1 atmospheric simulation [43], one of the targeted application for the BlueWaters project. To show the capability of our approach to efficiently hide I/O jitter and related costs, experiments have been carried on the French Grid’5000, on a Power5 cluster at NCSA and on the Kraken Cray XT5 supercomputer (currently 11th in the Top500) with up to 9K cores. By gathering data into large files while avoiding synchronization between processes, our solution brings several benefits:

1. it increases the sustained write throughput of the simulation by a factor of almost 15;
2. it provides almost 70 % overall application speedup on 9,000 cores;
3. it fully hides I/O-related costs;
4. it enables a 600 % compression ratio without any additional overhead, leading to a major reduction of storage requirements.

A poster [16] presenting some of these results has been accepted at ICS’11 and awarded the second price at the ACM Student Research Competition. All these results are presented in a research report [25] pending for publication.

Current work addresses the efficient coupling of large-scale simulations and visualization tools through Damaris. We have been able to get access to the Jaguar supercomputer hosted at ORNL and we are planning very-large scale experiments on up to 100,000 cores to show the benefits of our Damaris approach.

### 6.7. Scalable array-oriented active storage

**Participants:** Viet-Trung Tran, Gabriel Antoniu, Luc Bougé.
The recent explosion in data sizes manipulated by distributed scientific applications has prompted the need to develop specialized storage systems capable to deal with specific access patterns in a scalable fashion. In this context, a large class of applications focuses on parallel array processing: small parts of huge multi-dimensional arrays are concurrently accessed by a large number of clients, both for reading and writing. However, many established storage solutions such as parallel file systems and database management systems expose data access models (e.g., file systems, structured databases) that are too general and do not exactly match the nature requirements of the application. This forces the application developer to either adapt to the exposed data access model or to use an intermediate layer that performs a translation. In either case, the mismatch leads to suboptimal data management: the one-storage-solution-fits-all-needs has reached its limits.

Thus, there is an increasing need to specialize the I/O stack to match the requirements of the application. The objective of this research is to design Pyramid: an array-oriented active storage system optimizing for applications that represent and manipulate data as huge multi-dimensional arrays. However, a specialized storage system that deals with such an access pattern faces several challenges at the level of data/metadata management, we carefully design the system with the following principles: (1) we introduce a dedicated array-oriented data access model that offers support for active storage and versioning; (2) we enrich striping techniques specifically optimized for multi-dimensional arrays with a distributed metadata management scheme that avoids potential I/O bottlenecks observed with centralized approaches.

We evaluated Pyramid through a set of experiments on the Grid’5000 [56] testbed that aims to evaluate both the performance and the scalability of our approach under concurrent accesses. Preliminary evaluation in our recent papers [22], [13] shows promising results: our prototype demonstrates good performance and scalability under concurrency, both for read and write workloads.
6. New Results

6.1. Introduction

Research results are presented according to the research directions of the MYRIADS team.

6.2. Autonomous Management of Virtualized Infrastructures


6.2.1. Cloud Federations


6.2.1.1. Virtual Execution Platforms in Cloud Federations

In the context of the Contrail European project, we have defined the overall architecture of the Contrail software stack for cloud computing on top of cloud federations [51]. We have focused on the design and the implementation of a first basic prototype of the Virtual Execution Platform (VEP) component [52]. It is in charge of provisioning hardware resources from Cloud providers and to deploy and run distributed applications submitted by users under the control of a negotiated Service Level Agreement (SLA) [16]. Within VEP software, REST interface, OVF parsing, SSL security, Authorization modules are under active development and at various levels of integration. A first demo version of VEP running on top of OpenNebula IaaS cloud has been successfully demonstrated at the first annual project review.

6.2.1.2. Efficient virtual cluster migration

We continued our work on Shrinker, a system providing efficient live migration of virtual clusters on wide area networks. The design has been improved to coordinate the deduplication on the source site of the migration. Deduplication is now performed only within an individual virtual cluster, in order to reduce security issues and avoid performance impact on virtual machines of other users. We performed a comprehensive performance evaluation of Shrinker. An article presenting the design, implementation, and performance of Shrinker was published in [28].

6.2.1.3. Elastic Map/Reduce over Cloud Federations

We worked on the development of Resilin, a system to easily create execution platforms over distributed cloud resources for executing MapReduce computations. Resilin implements the Amazon Elastic MapReduce web service API and uses resources from private and community clouds. Resilin takes care of provisioning, configuring and managing cloud-based Hadoop execution platforms, potentially using multiple clouds. An initial implementation of Resilin was presented at the CCA ’11 workshop [36]. Further development was performed in the context of Ancuta Iordache’s master internship. The results of this work were published as a research report [45].

6.2.1.4. Sky Computing Experiments

We continued our collaboration with the University of Florida on sky computing experiments, which led to the publication of a book chapter [38].

6.2.2. Infrastructure as a Service Clouds

Participants: Stefania Costache, Eugen Feller, Yvon Jégou, Christine Morin, Nikos Parlavantzas, Pierre Riteau.
6.2.2.1. Large scale Energy aware self-healing IaaS

The research done in 2011 was two fold. A prototype of the previously proposed scalable, fault-tolerant and energy-aware virtual machine (VM) management framework called Snooze was implemented and evaluated on the Grid5000 testbed [41]. In 2011, we have focused on the implementation of the self-healing mechanisms and protocols, and on integrating in Snooze the system-level mechanisms (e.g. for automatically switching on/off cluster nodes) to support energy aware resource management algorithms. Our experimental results show that the fault-tolerance features of the framework do not impact application performance. Moreover, negligible cost is involved in performing distributed VM management and the system remains highly scalable with increasing amounts of resources. A nature-inspired VM placement algorithm [24] based on the Ant Colony Optimization (ACO) meta-heuristic was developed and evaluated by means of simulations.

6.2.2.2. Resource Management in Private Clouds

We focused on the design of a resource management system for private clouds that provides support for different application SLAs while maximizing the resource utilization of the infrastructure. As we also considered the need of providing users the incentives to truthfully express their valuation for the performance of their application we investigated existing economic models of allocating resources. As a result, we proposed a novel resource management architecture based on a virtual economy. In this system, independent agents monitor the application’s performance and dynamically provision virtual machines from the infrastructure under user’s budget constraints. To provision virtual machines, a proportional share auction is used, allowing a fine-grain resource sharing at a low complexity cost. This work was done as part of a collaboration with EDF R&D and was published at the VHPC 2011 workshop [22]. We have also implemented a first prototype of this proposal. In collaboration with Vydia Rajagopalan (Master student at VU Amsterdam) we have implemented the proportional-share auction scheduler and integrated it with the OpenNebula Virtual Infrastructure Manager. Then, we have extended this work with the design of agents that execute scientific applications (MPI and Bag-of-Task applications) under deadline and budget constraints. Experimental evaluations are currently performed on Grid5000.

6.2.2.3. Resilience

We initiated a collaboration with Box Leangsuksun’s group on high availability of cloud infrastructures. We carried out a preliminary study on how the HA-OSCAR environment developed at the Louisiana Tech University could be used to ensure the high availability of critical services in Nimbus IaaS clouds [30].

6.2.3. XtreemOS Grid Distributed Operating System

Participants: Amine Belhaj, Jérôme Gallard, Rémy Garrigue, Yvon Jégou, Christine Morin, Yann Radenac.

6.2.3.1. Facilitating Experiments with XtreemOS Grid System

XtreemOS Grid system that has been developed in the framework of the XtreemOS European project is now evolving as an open source software in a community driven by INRIA in the framework of the XtreemOS Easy ADT. We have provided first level support to users and maintained XtreemOS website, wiki and mailing-lists. We have updated XtreemOS documentation to reflect the evolution of XtreemOS system. We facilitated the access to XtreemOS in three different ways: maintaining an open public testbed running XtreemOS, providing ready-to-use XtreemOS virtual machines and developing tools to automatically deploy XtreemOS Grid system on the Grid’5000 large-scale experimentation platform. In 2011, we have finalized a new 3.0 version of XtreemOS system and ported it on top of the OpenSuse 11.4 Linux distribution. We performed a number of tests to validate the installation, configuration and execution of the new XtreemOS version based on openSuSE Linux distribution. An incremental integration process has been set up to facilitate the integration of patches and bug fixes. We have run a number of experiments with XtreemOS 3.0 based on Mandriva Linux distribution: MPI programs, Salomé numerical simulation platform, bio-informatics applications. Yann Radenac, in the framework of the COOP project funded by ANR contributed to XtreemOS’s code by fixing bugs, cleaning the source code to improve maintainability, and adding a few minor features.
6.2.3.2. Resource Management for Dynamic Applications

In the framework of the COOP project funded by ANR, we compared the features offered by the CooRM resource manager for dynamic applications developed by Christian Perez and Cristian Klein at ENS Lyon with those provided by the XtreamOS Grid system. A plan has been set to adapt CooRM to XtreamOS system and to extend XtreamOS’s API to include a CooRM-like interface [53]. We worked on the definition of a variant of CooRM that can work in the context of XtreamOS Grid operating system.

6.3. Dynamic Adaptation of Service-based Applications

Participants: Françoise André, Djawida Dib, Erwan Daubert, Guillaume Gauvrit, André Lage, Christine Morin, Nikos Parlavantzas, Jean-Louis Pazat, Chen Wang, Mohamed Zouari.

6.3.1. Dynamic Adaptation in a Distributed Operating System

Participants: Françoise André, Djawida Dib, Christine Morin, Nikos Parlavantzas.

We have studied the feasibility to dynamic adapt the features of a distributed operating system using a framework for self-adaptation of service oriented distributed applications [46]. We have focused on the consistency protocols for replicated data in distributed shared memory systems. We have considered two strict consistency protocols, one based on invalidation and one based on broadcast on write operations. The adaptation framework selects one of these two algorithms based on the inter-node data transfer delay. We have implemented a prototype based on Kerrighed single system image operating system for clusters and the SAFDIS adaptation framework. We have integrated a broadcast based consistency protocol in Kerrighed that already implements a write invalidation consistency protocol. We have implemented the adaptation policy in the SAFDIS framework and the needed adaptation mechanisms in Kerrighed as well as a component for monitoring data transmission delays. An experimental evaluation is being carried out.

6.3.2. Adaptation for Data Management

Participants: Françoise André, Mohamed Zouari.

The usage of context-aware data management in mobile environments has been previously investigated by Françoise André in collaboration with Mayté Segarra and Jean-Marie Gilliot from Telecom Bretagne Brest (previously known as ENST Bretagne). This work focuses on data management in grid and mobile environments; an ambient assisted living application illustrates the approach. This work was realized in the context of the ALORAD project (Architecture LOgicielle pour la Réplication Adaptative de Données), financed by the Brittany council. Mayté Segarra from Telecom Bretagne Brest was co-adviser for the PhD thesis of M. Zouari [12].

6.3.3. Adaptation for Service-Oriented Architectures

Participants: Françoise André, Erwan Daubert, Guillaume Gauvrit, André Lage, Nikos Parlavantzas, Jean-Louis Pazat, Chen Wang.

Service-Oriented Computing is a paradigm that is rapidly spreading in all application domains and all environments - grids, clusters of computers, mobile and pervasive platforms. The following works take place in the context of the S-CUBE European Network of Excellence.

6.3.3.1. Services adaptation in distributed and heterogeneous systems

We are still studying service adaptation in distributed and heterogeneous systems. This work covers different aspects such as structural, behavioral and environmental adaptation, distributed decision and planification of adaptation actions, adaptive allocation of resources for services. A framework called SAFDIS for "Self Adaptation For Distributed Services" has been defined and implemented. It is built as a set of services, providing functionalities useful to build an adaptation system. The analysis phase can take reactive as well as proactive decisions. This gives the ability to either react fast or to take decisions for the long term. This implies the ability to analyze the context with a variable depth of reasoning. Our implementation of the SAFDIS analysis phase also distributes and decentralizes its analysis process to spread the computational
load and make the analysis process scalable. The planning phase seeks the set of actions (the plan) needed to adapt the system according to the strategy chosen by the analysis phase. It also schedules the selected actions to ensure a coherent and efficient execution of the adaptation. The planning topic is a well known subject in AI research works and many algorithms already exist in that field to produce efficient schedules. With our SAFDIS framework, the planning phase is able to reuse these algorithms. The resulting plan of actions can have actions that can be executed in parallel.

6.3.3.2. Quality Assurance for Distributed Services

In the context of the service-centric paradigm, we have designed and developed the Qu4DS (Quality Assurance for Distributed Services) research prototype. Qu4DS is a cloud PaaS solution which fills the gap between the conception of higher-level SaaS service providers over the resource-level PaaS layer. Qu4DS provides an automatic support for service execution management by aiming at increasing the service provider’s profit. More specific, Qu4DS dynamically acquires resources according to the customer demand, deploys service instances and implements QoS assurance mechanisms in order to prevent SLA violations. Moreover, Qu4DS has been evaluated on Grid5000 and showed to be effective on reducing service provider’s costs [27].

6.3.3.3. Self-configuration for Cloud Platforms

By definition, cloud computing offers an abstraction to manage various needs and concepts such as distributed software design, the deployment of such software on dynamic resources and the management of this kind of resources. Thus it is possible to reconfigure (adapt) according to some needs the software as well as the use of the resources. However these reconfigurations that are used on different layers may also have impacts on the others. Moreover these layers are independent and so are able to adapt themselves independently of the others. In our work, we propose to use some adaptation capabilities offered for example by the infrastructure (IaaS) that manages the resources to adapt the software (SaaS). We also propose to use planning algorithms to coordinate the adaptations between them to avoid inconsistency or inefficiency due to concurrent adaptations.

6.3.3.4. Dynamic Adaptation of Chemical services

We have proposed a QoS-aware middleware for dynamic service execution. In the context of dynamic execution, a workflow is defined by composing a set of abstract activities as place holders. Each activity is bound to a suitable partner service, which is selected at run-time from a set of functional equivalent candidates with different non-functional properties such as quality of service (QoS). The service selection process is modeled as a series of chemical reactions.

6.3.3.5. Prediction of SLA violations and dynamic adaptation in workflows

During execution, run-time QoS is determined by the dynamic execution environment and thus the expected QoS is not always ensured. In addition, infrastructure failures can make a service undeliverable. The adaptive execution reflects the capability to recompose (a part of a) workflow on-the-fly in case that global SLA violation is predicted. Most techniques for predicting global SLA violation require past experiences on executions of a business process. All historical execution instances have the same structure as well as the same bindings. These solutions do not adapt to the case of dynamic execution, where for each execution, partner services are selected and bound at run time.

In order to predict global SLA violation in the context of dynamic service execution, we proposed a 2-phase prediction technique, which is fit for generic workflow composition. The prediction method works with a high accuracy for simple workflows, but when the workflow composes complicated structures (such as loops and exclusive branches), the performance degrades. The reason is that the estimation of global SLA is based on the critical path, which is not definitely executed. To solve this problem, we propose to use data mining technique to predict workflow branches and the number of loop execution. Based on predicted branches, the prediction of global SLA violation is much more accurate. The numerical evaluation will be carried out in the near future.

6.4. A Chemical Approach for Autonomous Service Computing

Participants: Héctor Fernandez, Marko Obrovac, Thierry Priol, Cédric Tedeschi.
6.4.1. Chemistry-Inspired Workflow Management System for e-Science Applications

**Participants:** Héctor Fernandez, Cédric Tedeschi, Thierry Priol.

In the research track that aims at leveraging the properties of the chemical Programming models for autonomic computing, we have built a software based on the HOCL compiler (part of the HOCL-tools) that was actually deployed and experimented over the Grid’5000 platform. The experiments have shown, that envisioning the execution workflow as an autonomic chemical process is actually viable in practice. Experimented with different well known workflow-based e-Science applications, the software showed a performance level comparable to current top-rated scientific workflow management systems [25].

6.4.2. Solving Workflow Patterns Through Molecular Composition

**Participants:** Héctor Fernandez, Cédric Tedeschi, Thierry Priol.

In the same area, but on a more conceptual point of view, we have shown how the expressive power of the chemical model can be leveraged to solve complex workflow patterns. This aspect was also integrated into the HOCL-tools and experimented over the Grid’5000 platform, following two architectures with a different level of decentralization, showing the advantages and drawbacks of decentralizing the workflow execution using a chemical workflow engine [26].

6.4.3. Scalable Atomic Capture of Molecules

**Participants:** Marko Obrovac, Cédric Tedeschi.

Capturing the reactants involved in a reaction constitutes one of the main challenges in the execution of chemical programs. Doing it at large scale is one of the essential barriers hindering the actual execution of chemical programs at large scale. We proposed a protocol solving this issue on top of a distributed hash table (DHT). The DHT secures the scalability of the communications and provides a scalable discovery of reactants. Our protocol is triggered once reactants are found. It is made of two sub-protocols being used at different stages of the computation, according to the density of possible reactions. The protocol is validated through its proof of liveness and simulations showing that it is able to switch from one sub-protocol to the other efficiently, according to the execution’s conditions [18].
6. New Results

6.1. Model Driven and Aspect Oriented Design

6.1.1. Requirements Engineering

Participants: Olivier Barais, Benoît Baudry, Benoît Combemale, Maha Driss, Jean-Marc Jézéquel, Emmanuelle Rouillé, Nicolas Sannier, Didier Voitisek.

Model-driven engineering can have a huge impact on the early design and analysis of complex systems. We have investigated modeling for requirements engineering in three areas:

- We use executable metamodeling techniques developed in the team to capture formal relationships between regulatory requirements and accepted practices in systems engineering [47], [42].
- We propose an approach for facilitating Web service selection according to user requirements. These requirements specify the needed functionality and expected QoS, as well as the composability between each pair of services. The originality of our approach is embodied in the use of Formal Concept Analysis (FCA) and its extension Relational Concept Analysis (RCA) [33] [25].
- We have analyzed a real industrial software process to illustrate the need for bridging the gap between software processes and software development tools to automate the development tools configuration, deployment, integration and adaptation [46].

6.1.2. Dynamically adaptive interactive systems

Participants: Arnaud Blouin, Jean-Marc Jézéquel, Grégory Nain.

Combining Aspect-Oriented Modeling with Property-Based Reasoning to Improve User Interface Adaptation: in this work we combined aspect-oriented modeling with an interactive system architecture to support dynamic adaptation of interactions and user interfaces [28].

6.1.3. Dynamically adaptive component-based systems

Participants: François Fouquet, Olivier Barais, Viet-Hoa Nguyen, Noël Plouzeau.

Continuous Design to Achieve Intelligent Reflection in Distributed Systems: we defined an intelligent reflection model to support fast adaptation of distributed systems by architecture modification without stopping the system. This adaptation mechanism is well suited to rapidly changing needs (continuous design of eternal systems) or fast paced modifications of the context of the running system (for instance for Internet of Things distributed systems) [34].

6.1.4. Architecture for Services-based applications

Participants: Olivier Barais, Johann Bourcier, Erwan Daubert, Jean-Marc Jézéquel.

The architecture of service-based applications can have a huge impact on their dynamic adaptability. We have investigated various framework for architecting service-based applications:

- Designing SAFDIS: a self adaptive framework for distributed applications based on services. SAFDIS includes facilities to support the coordination of distributed reconfigurations [24]. SAFDIS also takes benefit of the Infrastructure As A Service to dynamically reconfigure Software As A Service [44].
- Analyzing and improving consistency between functional and business view of telecom services architecture. This work is based on the definition of a strategic alignment of the target functional view with the target business view. Alignment is validated with a real case study implemented and deployed at Orange–France Telecom on their messaging service [22].
• Designing AutoHome: a service oriented framework to simplify the development and runtime adaptive support of autonomic pervasive applications. This includes the amalgamation of the two computing areas of Autonomics and Service Orientation, to produce a Component-based platform providing facilities. This infrastructure uniquely blends the advantages of distributed autonomic control with global conflict management in a management hierarchy [17].

6.2. Model V&V and Testing

6.2.1. Formal MDE Foundations

Participants: Benoit Baudry, Benoit Combemale.

• Formally Tracing Executions From an Analysis Tool Back to a Domain Specific Modeling Language’s Operational Semantics: in this work, we propose a formal and operational framework for tracing results back (e.g., a program crash log, or a counterexample returned by a model checker) from execution and verification tools to an original DSML’s syntax and operational semantics [31].

• A Proof Assistant Based Formalization of components in MDE: using the Coq proof assistant we propose a formalization of some operators for model fragment extraction and composition, as defined in the ReuseWare toolset [39].

• We have developed a methodology to explicitly model the context in which a temporal property must be verified. This contextual information is expressed in the requirements, and an explicit model allows to reduce the complexity of automated verification [41].

6.2.2. Pairwise testing for highly variable systems

Participants: Benoit Baudry, Aymeric Hervieu.

Variability management is a major concern for the development of software intensive systems. In particular, the explosion of variants is an issue for testing and analysis. Feature models allow to explicitly capture the variability in a formal model and get a complete view on all possible variants of the system. We have investigated pair-wise generation from feature models in order to test software product lines [36], and to evaluate QoS contracts in variable web service compositions [38].

6.2.3. Testing aspect-oriented programs

Participant: Benoit Baudry.

Aspect-oriented mechanisms introduce new risks for reliability that must be tackled by specific testing techniques in order to fully benefit from the use of this paradigm. We have investigated a monitoring mechanism of advices in an aspect-oriented program and use this information to build test cases that target faults in pointcut descriptors [18].

6.2.4. Modeling model quality metrics

Participants: Benoit Baudry, Jean-Marc Jézéquel.

We have developed a model-driven measurement approach to measure models of a domain specific modeling language. The approach uses models as unique and consistent metric specifications for the automated generation of a metric tool. The benefit from applying the approach is evaluated by four case studies [20]. In particular, we have evaluated the ability of the approach to build a tool for the measurement of requirements documents [21].
6.3. Meta-Modeling

6.3.1. Model Driven Language Engineering

Participants: Benoit Baudry, Arnaud Blouin, Juan-Jose Cadavid Gomez, Benoit Combemale, Clément Guy, Jean-Marc Jézéquel, Didier Vojtisek.

- Model-Driven Engineering and Optimizing Compilers: A bridge too far? In this work, we report and analyze an experience about the use of MDE technologies to build and evolve compiler infrastructures in the optimizing compiler domain. From this study, we highlight challenges and propose a roadmap for the cross-fertilization of the MDE and compiler domains [35], [45].
- Modeling Model Slicers: model slicing is a model operation that consists in extracting a subset of a model. Because the creation of a new DSL implies the creation from scratch of a new model slicer, we proposed the Kompren language that models and generates model slicers for any DSL [27].
- Empirical Evaluation of the Conjunct Use of MOF and OCL: we evaluate in this work the conjunct usage of MOF (Meta-Object Facility) and OCL (Object Constraint Language) in the development of Domain-Specific Modeling Languages. We observe the state of practice to understand how experts use them and find patterns on its usage, in order to provide techniques to improve the experience [29].

6.3.2. Model Transformation and Composition

Participants: Olivier Barais, Benoit Baudry, Arnaud Blouin, Mickaël Clavreul, Benoit Combemale, Xavier Dolques, Jean-Marc Jézéquel.

- Model operations such as transformation and composition declare source metamodels that are usually larger than the set of concepts and relations actually used by the operation. We have proposed and validated a static operation analyzer to retrieve the metamodel footprint of the operation [37].
- Service-Oriented Architecture Modeling: Bridging the Gap between Structure and Behavior: In this approach, we propose to detect divergences among structural and behavioral models to support a semi-automatic process of synchronization between class diagrams and workflow models [30].
- The paper propose a technique for discovering matchings between two model elements modeling the same system, but being instances of different metamodels. This is achieved by using property names and models structure thanks to the adaptation of a schema matching technique named Anchor-PROMPT [32].
- Specifying and implementing UI Data Bindings with Active Operations: based on the concept of active operations, this work proposes a framework to bind models at runtime and more precisely to bind data and their possible representations [26].
- We propose a requirement-centric approach for Web service composition which allows: (i) modeling users’ requirements with the MAP formalism and specifying required services using an Intentional Service Model (ISM); (ii) discovering and selecting relevant Web services and high QoS services; and (iv) generating automatically BPEL coordination processes by applying the model transformation technique [19].
6. New Results

6.1. Diagnosis of large scale discrete event systems

Participants: Marie-Odile Cordier, Christine Largouët, Sophie Robin, Laurence Rozé, Yulong Zhao.

The problem we deal with is monitoring complex and large discrete-event systems (DES) such as an orchestration of web services or a fleet of mobile phones. Two approaches have been studied. The first one consists in representing the system model as a discrete-event system by an automaton. In this case, the diagnostic task consists in determining the trajectories (a sequence of states and events) compatible with the sequence of observations. From these trajectories, it is then easy to determine (identify and localize) the possible faults. In the second approach, the model consists in a set of predefined characteristic patterns. We use temporal patterns, called chronicles, represented by a set of temporally constrained events. The diagnostic task consists in recognizing these patterns by analyzing the flow of observed events.

More recently, we started research on interacting with large-scale systems in a decision-oriented way. Scenario patterns were defined for exploring complex systems, based on the use of model-checking techniques.

6.1.1. Distributed monitoring with chronicles - Interleaving diagnosis and repair - Making web services more adaptive

Our work addresses the problem of maintaining the quality of service (QoS) of an orchestration of Web services (WS), which can be affected by exogenous events (i.e., faults). The main challenge in dealing with this problem is that typically the service where a failure is detected is not the one where a fault has occurred: faults have cascade effects on the whole orchestration of services. We have proposed a novel methodology to treat the problem that is not based on Web service (re)composition, but on an adaptive re-execution of the original orchestration. The re-execution process is driven by an orchestrator Manager that takes advantage of an abstract representation of the whole orchestration and may call a diagnostic module to localize the source of the detected failure. It is in charge of deciding the service activities whose results can be reused and may be skipped, and those that must be re-executed. A paper has been submitted to the CAISE conference.

6.1.2. Scenario patterns for exploring qualitative ecosystems

Our work aims at giving means of exploring complex systems, in our case ecosystems. We proposed to transform environmental questions about future evolution of ecosystems into formalized queries that can be submitted to a simulation model. The system behavior is represented as a discrete event system described by a set of interacting timed automata, the global model corresponding to their composition on shared events. To query the model, we have defined high-level generic patterns associated to the most usual types of scenarios. These patterns are then translated into temporal logic formula. The answer is computed thanks to model-checking techniques that are efficient for analysing large-scale systems. Five generic patterns have been defined using TCTL (Timed Computation Tree Logic): WhichStates, WhichDate, WhichStates, Stability, Safety. Three of them have been implemented using the model-checker UPPAAL.

The approach has been experimented on a marine ecosystem under fishing pressure. The model describes the trophodynamic interactions between fish trophic groups as well as interactions with the fishery activities and with an environmental context. A paper has been accepted for publication by the Environmental Modelling Software Journal [4].

We extended the approach to deal with “How to” queries. As before, we rely on a qualitative model in the form of timed automata and use model-checking tools to answer queries. We have recently proposed two approaches to answer questions such as “How to avoid a given situation?” (safety query). The first one exploits controller synthesis and the second one is a “generate and test” approach. We compared these two approaches in the context of an application that motivates this work, i.e the management of a marine ecosystem and the evaluation of fishery management policies. The results have been accepted for publication in [14].
6.2. Machine learning for model acquisition

Participants: Thomas Guyet, René Quiniou.

Model acquisition is an important issue for model-based diagnosis, especially as modeling dynamic systems. We investigate machine learning methods for temporal data recorded by sensors or spatial data resulting from simulation processes. We also investigate efficient methods for storing and accessing large volume of simulations data. Our main interest is extracting knowledge, especially sequential and temporal patterns or prediction rules, from static or dynamic data (data streams). We are particularly interested in mining temporal patterns with numerical information and in incremental mining from sequences recorded by sensors.

6.2.1. Mining temporal patterns with numerical information

We are interested in mining interval-based temporal patterns from event sequences where each event is associated with a type and time interval. Temporal patterns are sets of constrained interval-based events. This year, we have been working on improving the formal setting of the approach as well as its efficiency [8]. We have introduced the notion of \(\epsilon\)-covering of temporal patterns over sequences to cope with the dual nature, symbolic and numerical, of temporal patterns. The parameter \(\epsilon\) specifies the tightness of the similarity used for matching patterns and sequences. It complements the parameter \(\sigma\) representing the minimal support which is used to prune candidate patterns. The \(\epsilon\)-similar occurrences of some pattern, precisely their associated temporal intervals, are classified to characterize the different classes of numerical temporal intervals that correspond to different patterns sharing the same symbolic part. This process have been embedded in two sequential pattern mining algorithms, GSP and PrefixSpan, and we have compared their performance.

6.2.2. Incremental sequential mining

We investigate the problem of mining and maintaining frequent sequences in a window sliding on a stream of itemsets. We propose in [11] a complete and correct incremental algorithm based on a tree representation of frequent sequences inspired by PSP [52] and a method for counting the minimal occurrences of a sequence. Instead of the frequence, to a node representing a pattern is associated the set of occurrences of this pattern. The algorithm updates efficiently the tree representation of frequent sequences and their occurrences by means of two operations on the tree: deletion of the itemset at the beginning of the window (obsolete data) and addition of an itemset at the end of the window (new data). Experiments were conducted on simulated data and on real data of instantaneous power consumption.

6.2.3. Multiscale segmentation of satellite image time series

Satellite images allow the acquisition of large-scale ground vegetation. Images are available along several years with a high acquisition frequency (1 image every two weeks). Such data are called satellite image time series (SITS). In [9], we present a method to segment an image through the characterization of the evolution of a vegetation index (NDVI) on two scales: annual and multi-year. We test this method to segment Senegal SITS and compare our method to a direct classification of time series. The results show that our method using two time scales better differentiates regions in the median zone of Senegal and locates fine interesting areas (cities, forests, agricultural areas).

6.3. Decision aiding with models and simulation data

Participants: Tassadit Bouadi, Marie-Odile Cordier, Véronique Masson, Florimond Ployette, René Quiniou, Karima Sedki.

Models can be very useful for decision aiding as they can be used to play different plausible scenarios for generating the data representing future states of the modeled process. However, the volume of simulation data may be very huge. Thus, efficient tools must be investigated in order to store the simulation data, to focus on relevant parts of the data and to extract interesting knowledge from these data.
6.3.1. Exploring models thanks to scenarios: a generic framework

In the framework of the APPEAU project (see 8.2.1), that ended in December 2010, a paper, describing a generic framework for scenario exercises using models applied to water-resource management, has been written during 2011 in cooperation with all the partners and submitted to Environmental Modelling and Software. It is currently under revision.

6.3.2. A datawarehouse for simulation data

The ACASSYA project aims at providing experts or stakeholders or farmers with a tool to evaluate the impact of agricultural practices on water quality. As the simulations of the deep model TNT2 are time-consuming and generate huge data, we have proposed to store these simulation results in a datawarehouse and to extract relevant information, such as prediction rules, from the stored data. We have devised a general architecture for agro-environmental data on top of the framework Pentaho. An article presenting the principles of this architecture as well as a set of realistic scenarios and their transformation into OLAP queries has been submitted to Compag (Computers and Electronics in Agriculture).

6.3.3. Efficient computation of skyline queries in an interactive context

Skyline queries retrieve from a database the objects that optimizes multiple criteria, related to user preferences for example, or objects that are the best compromises satisfying these criteria. When data are huge such objects may shed light on interesting parts of the dataset. However, computing the skylines (i.e. retrieving the skyline points) may be time consuming because of many dominance tests. This is, especially the case in an interactive setting such as querying a data cube in the context of a datawarehouse. We have worked on how to answer efficiently to skyline queries by the materialization of precomputed skyline queries related to dynamic user preferences. An article has been submitted to the conference SIGMOD 2012.

6.3.4. Influence Diagrams for Multi-Criteria Decision-Making

For multi-criteria decision-making problems, we propose in [6] a model based on influence diagrams able to handle uncertainty, represent interdependencies among the different decision variables and facilitate communication between the decision-maker and the analyst. The model makes it possible to take into account the alternatives described by an attribute set, the decision-maker’s characteristics and preferences, and other information (e.g., internal or external factors) that influence the decision. Modeling the decision problem in terms of influence diagrams requires a lot of work to gather expert knowledge. However, once the model is built, it can be easily and efficiently used for different instances of the decision problem. In fact, using our model simply requires entering some basic information, such as the values of internal or external factors and the decision-maker’s characteristics.

6.3.5. Recommending actions from classification rules

In the framework of the SACADEAU project (see 8.2.1), a paper dedicated to building recommendation actions for a given situation, from the set of classification rules, learnt from simulation results, has been published in the KAIS journal [7].

6.4. Causal reasoning and influence diagrams

Participants: Philippe Besnard, Louis Bonneau de Beaufort, Marie-Odile Cordier, Yves Moinard, Karima Sedki.

This work stems on [23], [24], [25], [26], [27] and, for the logic programming translation, on [53], [54]. It is related to diagnosis (observed symptoms explained by faults).

The previously existing proposals were ad-hoc or, as in [29], [41], they were too close to standard logic in order to make a satisfactory diagnosis. Our proposal starts from a restricted first order logic (of the Datalog kind: no function symbols) and introduces causal formulas, built on causal atoms such as (α causes β) intended to mean: “α causes β”. The system is described thanks to these causal formulas, classical formulas, and taxonomy atoms such as (α IS_A β) (α is of kind β).
The system produces explanation atoms of the kind $(\alpha \text{ explains } \beta \text{ if possible } \{\gamma_1, \ldots, \gamma_n\})$, meaning that $\beta$ can be explained by $\alpha$ if all the $\gamma_i$’s are possible together in the context of the given data.

This year, we have improved our logic programming translation in ASP. The aim is to improve efficiency and also reduce the work of the programmer, taking advantage of the declarative aspect of this type of programming. We have applied some of these improvements to two classic riddles, in order to illustrate the power and limitations of current answer set programming systems, and we proposed a few improvements which could make the present systems yet easier to use [12], [13].

We are starting a work with some similarities to automatize the treatment of cognitive maps. The aim is to extract relevant information from these maps, which means: building a graph formalism for representing mixed causal and influence relations, and defining a framework (argumentation theory is a good candidate) to aggregate the graphs and provide inference rules in order to infer new information and relations. This work is done in the framework of the RADE2BREST project, involving Agrocampus Ouest and CNRS (GEOMER/LETG), funded by “Ministère de l’Ecologie”¹. The goal of this project is to model shellfish fishing in order to assess the impact of management pollution scenarios on the Rade de Brest. The cognitive maps result from interviews with fishermen.

¹This project is not mentioned in section 8.1 because DREAM is not a partner of this project.
6. New Results

6.1. Visual tracking

6.1.1. 3D model-based tracking

Participants: Antoine Petit, Eric Marchand.

Our 3D model-based tracking algorithm [3] was used in various contexts. First, it has been studied and tested on a mock-up of a telecommunication satellite using a 6-DOF robotic arm, with satisfactory results in terms of accuracy of the pose estimation and computational costs [41], [42]. A potential application would be the final phase of space rendezvous mission using visual navigation. Then, it has been considered for designing a visual servoing scheme able to control the walking of a humanoid robot [29].

6.1.2. Omnidirectional stereovision

Participants: Guillaume Caron, El Mustapha Mouaddib, Eric Marchand.

Omnidirectional cameras allow direct tracking and motion estimation of planar regions in images during a long period of time. However, using only one sensor leads to plane and trajectory reconstruction up to a scale factor. We proposed to develop dense plane tracking based on omnidirectional stereovision to answer this issue. The method estimates simultaneously the parameters of several 3D planes along with the camera motion using a spherical projection model formulation [20].

6.1.3. Motion estimation using mutual information

Participant: Eric Marchand.

Our work with Amaury Dame related to template tracking using mutual information as registration criterion has been extended to motion estimation applications. It has been applied to mosaicing from an image sequence [28]. The main advantage is that this approach is robust to noise, lighting variations and does not require a statistically robust estimation process.

6.1.4. Augmented reality

Participants: Pierre Martin, Hideaki Uchiyama, Eric Marchand.

We developed an approach for detecting and tracking various types of planar objects with geometrical features [45]. We combine traditional keypoint detectors with Locally Likely Arrangement Hashing (LLAH) for keypoint matching. In order to produce robustness to scale changes, we build a non-uniform image pyramid according to keypoint distribution at each scale. It demonstrates that it is possible to detect and track different types of textures including colorful pictures, binary fiducial markers and handwritings. This approach was extended to consider non-rigidly deformable markers [46].

6.2. Visual servoing

6.2.1. Micro-manipulation

Participant: Eric Marchand.

We developed an accurate nanopositioning system based on direct visual servoing [43],[17]. This technique relies only on the pure image signal to design the control law, by using the pixel intensity of each pixel as visual features. The proposed approach has been tested in terms of accuracy and robustness in several experimental conditions. The obtained results have demonstrated a good behavior of the control law and very good positioning accuracy: 89 nm, 14 nm, and 0.001 degrees in the x, y and \( \theta \) axes of a positioning platform, respectively.
6.2.2. Multi sensor-based control  
**Participants:** Olivier Kermorgant, François Chaumette.

We have designed a generic sensor-based control approach to automatically tune the weights related to the features involved as inputs of a control scheme, allowing to take constraints into account. This scheme has been applied to several configurations, such as fusing the data provided by an eye-in-hand camera and an eye-to-hand camera, ensuring the visibility constraint, and avoiding the robot joint limits [30], [31], [32], [11].

6.2.3. Visual navigation of mobile robots  
**Participants:** Eric Marchand, Andrea Cherubini, Fabien Spindler, François Chaumette.

We have developed a visual servoing scheme based on the mutual information between the images acquired by an onboard camera and a visual memory to control the orientation of a vehicle during its navigation [27]. We have also fused the data provided by a pan-tilt camera and a laser range sensor for the autonomous navigation of a mobile vehicle while avoiding obstacles [23], [22]. Real experiments with our Cycab (see Section 5.4) have been conducted on Place de Jaude in Clermont-Ferrand in the scope of the ANR Tosa CityVIP project (See Section 8.2.1).

6.2.4. Visual servoing for aircrafts  
**Participants:** Céline Teulière, Eric Marchand, Laurent Coutard, François Chaumette.

A dynamic controller has been designed for the homing of a quadri-rotor aerial vehicle [39]. A color-based tracking algorithm has also been designed and combined with an image-based visual servoing for chasing a moving target from a flying UAV [44]. Finally, a method has been developed to detect and localize an aircraft carrier in an image sequence, from which visual servoing control laws have been designed for the automatic landing [25], [26].

6.3. Medical robotics

6.3.1. Visual servoing based on ultrasound images  
**Participants:** Caroline Nadeau, Alexandre Krupa.

We developed a new approach of ultrasound image based visual servoing that directly uses the intensities of the ultrasound image pixels as visual features. This method that spares any segmentation or image processing time consuming step was initially proposed to control the 6 DOF of a conventional 2D probe for positioning and tracking tasks [38], [48]. To increase the tracking performance we also adapted this method by considering a predictive control law based on the periodicity of physiological motions [36]. Rigid motion compensation experiments were conducted in the context of the ANR USComp project (See Section 8.2.3). The method was also improved by estimating on-line the image 3D gradient required for the positioning task and extended for the use of a bi-plane ultrasound probe [37]. Finally, the use of a 3D motorized probe was also considered to compute directly the image 3D gradient and a comparison of the results obtained with the different probes (2D, bi-plan, 3D) was performed [12].

6.3.2. Autonomous control modes for ultrasound probe guidance  
**Participants:** Tao Li, Alexandre Krupa.
In the context of the ANR Prosit (See Section 8.2.2), we developed several autonomous control modes in order to assist a doctor during a robotized and teleoperated ultrasound examination (tele-echography). The robotic tasks we proposed concern: an automatic scanning of the patient by a 2D probe, a shared control mode that maintains the visibility of an anatomic element of interest while the doctor teleoperates the slave robot holding the 2D probe, an automatic positioning task that allows the doctor to retrieve a desired anatomic section that was previously captured by the doctor. The two latter modes are based on visual servoing schemes that use as input image moments extracted from the observed 2D ultrasound image. This extraction is performed thanks to an active contour (snake) based on Fourier descriptors that we developed and implemented on GPU in order to provide real-time performance [34], [47]. The proposed autonomous control modes were experimentally validated on the Lagadic medical robotics platform (see Section 5.3) and are now in the process of being integrated on the Prosit robot platform.

6.3.3. Real-time 3D ultrasound image reconstruction and 3D deformation tracking

Participants: Deukhee Lee, Alexandre Krupa.

We developed and implemented on GPU an algorithm that reconstructs in real-time a sequence of dense ultrasound volumes from a set of pre-scan 2D ultrasound images provided online by a motorized ultrasound probe [33]. Then we proposed a dense ultrasound tracking algorithm that estimates in real time both rigid and non-rigid motions of a region of interest observed in the sequence of reconstructed ultrasound volumes [33]. The algorithm consists in estimating in real-time, from intensity-value changes between successive 3D ultrasound images, motions of a set of 3D control points that describe the evolution of 3D Thin-Plate Splines (TPS) modeling the deformation. The estimated rigid motion was then used in a pose-based control scheme to automatically displace the probe held by a robot for soft tissue motion compensation. These works were conducted in the context of the ANR USComp project (See Section 8.2.3).
6. New Results

6.1. Audio and speech content processing

6.1.1. Audio motif discovery

Participants: Frédéric Bimbot, Laurence Catanese, Armando Muscariello.

*This work was performed in close collaboration with Guillaume Gravier from the Texmex project-team.*

As an alternative to supervised approaches for multimedia content analysis, where predefined concepts are searched for in the data, we investigate content discovery approaches where knowledge emerge from the data. Following this general philosophy, we pursued work on motif discovery in audio contents.

Audio motif discovery is the task of finding out, without any prior knowledge, all pieces of signals that repeat, eventually allowing variability. In 2011, we extended our recent work on seeded discovery to near duplicate detection and spoken document retrieval from examples. First, we proposed algorithmic speed ups for the discovery of near duplicate motifs (low variability) in large (several days long) audio streams, exploiting subsampling strategies [muscariello-cbmi-11]. Second, we investigated the use of previously proposed efficient pattern matching techniques to deal with motif variability in speech data [muscariello-icassp-11] in a different setting, that of spoken document retrieval from an audio example. We demonstrated the potential of model-free approaches for efficient spoken document retrieval on a variety of data sets, in particular in the framework of the Spoken Web Search task of the MediaEval 2011 international evaluation [muscariello-is-11, muscariello-mediaeval-11].

This work is carried out in the context of the Quaero Project.

6.1.2. Landmark-driven speech recognition

Participant: Stefan Ziegler.

*This work is supervised by Guillaume Gravier and Bogdan Ludusan from the Texmex project-team.*

Speech recognition is a key issue to access multimedia spoken contents. In this context, speech recognition faces several challenges among which robustness to acoustic and linguistic variability.

In 2011, we initiated research on landmark-driven speech recognition to increase robustness. The idea of this approach consists in accurately detecting in the signal landmarks corresponding to broad phonetic classes (vowels, nasals, etc.). These landmarks, which represent almost certain knowledge about the phonetic content of the signal, are then used to bias the search space in Viterbi decoding towards solutions consistent with the landmarks. We proposed a landmark detection system, which employs numerous attributes extracted from a segment based representation of speech. We use a decision tree for BPC classification, since this allows the evaluation of each BPC on its most informative attributes, selected from a large variety of attributes. Then, each segment is converted into a landmark and a probability estimate for each BPC is provided. Second, we extend a previously proposed landmark-driven decoding strategy by a more flexible implementation, which reinforces paths at the detected landmarks according to the obtained BPC probabilities. Results obtained on French broadcast news data show a relative improvement in word error rate of about 2 % with respect to the baseline.

6.2. Recent results on sparse representations

The team has had a substantial activity ranging from theoretical results to algorithmic design and software contributions in the field of sparse representations, which is at the core of the FET-Open European project (FP7) SMALL (Sparse Models, Algorithms and Learning for Large-Scale Data, see Section 7.2.1) and the ANR project ECHANGE (ECHantillonnage Acoustique Nouvelle GEnération, see, Section 6.3.1).
6.2.1. A new framework for sparse representations: analysis sparse models 
Participants: Rémi Gribonval, Sangnam Nam.

Main collaboration: Mike Davies (Univ. Edinburgh), Michael Elad (The Technion), Hadi Zayyani (Sharif University)

In the past decade there has been a great interest in a synthesis-based model for signals, based on sparse and redundant representations. Such a model assumes that the signal of interest can be composed as a linear combination of few columns from a given matrix (the dictionary). An alternative analysis-based model can be envisioned, where an analysis operator multiplies the signal, leading to a cosparse outcome. Within the SMALL project, we initiated a research programme dedicated to this analysis model, in the context of a generic missing data problem (e.g., compressed sensing, inpainting, source separation, etc.). We obtained a uniqueness result for the solution of this problem, based on properties of the analysis operator and the measurement matrix. We also considered a number of pursuit algorithms for solving the missing data problem, including an L1-based and a new greedy method called GAP (Greedy Analysis Pursuit). Our simulations demonstrated the appeal of the analysis model, and the success of the pursuit techniques presented. These results have been published in international conferences [64], [63], and a journal paper is in preparation.

Our simulations demonstrated the appeal of the analysis model, and the success of the pursuit techniques presented. These results have been published in conferences [64], [91], [92] and a journal paper submitted to Applied and Computational Harmonic Analysis is under revision [103]. Other algorithms based on iterative cosparse projections [57] as well as extensions of GAP to deal with noise and structure in the cosparse representation have been developed, with applications to toy MRI reconstruction problems and acoustic source localization and reconstruction from few measurements (submitted to ICASSP 2012).

6.2.2. Theoretical results on sparse representations and dictionary learning 
Participants: Rémi Gribonval, Sangnam Nam, Nancy Bertin.

Main collaboration: Karin Schnass (EPFL), Mike Davies (University of Edinburgh), Volkan Cevher (EPFL), Simon Foucart (Université Paris 5, Laboratoire Jacques-Louis Lions), Charles Soussen (Centre de recherche en automatique de Nancy (CRAN)) Jérôme Idier (Institut de Recherche en Communications et en Cybernétique de Nantes (IRCCyN)), Cédric Herzet (Equipe-projet FLUMINANCE (INRIA - CEMAGREF, Rennes)) Morten Nielsen (Department of Mathematical Sciences [Aalborg]), Gilles Puy, Pierre Vanderheynst, Yves Wiaux (EPFL) Mehrdad Yaghoobi, Rodolphe Jenatton, Francis Bach (Equipe-projet SIERRA (INRIA, Paris)) Boaz Ophir, Michael Elad (Technion) Mark D. Plumbley (Queen Mary, University of London)

Sparse recovery conditions for Orthogonal Least Squares: We pursued our investigation of conditions on an overcomplete dictionary which guarantee that certain ideal sparse decompositions can be recovered by some specific optimization principles / algorithms. This year, we extended Tropp’s analysis of Orthogonal Matching Pursuit (OMP) using the Exact Recovery Condition (ERC) to a first exact recovery analysis of Orthogonal Least Squares (OLS). We showed that when ERC is met, OLS is guaranteed to exactly recover the unknown support. Moreover, we provided a closer look at the analysis of both OMP and OLS when ERC is not fulfilled. We showed that there exist dictionaries for which some subsets are never recovered with OMP. This phenomenon, which also appears with ℓ1 minimization, does not occur for OLS. Finally, numerical experiments based on our theoretical analysis showed that none of the considered algorithms is uniformly better than the other. This work has been submitted for publication in a journal [108].

New links between the Restricted Isometry Property and nonlinear approximations: It is now well known that sparse or compressible vectors can be stably recovered from their low-dimensional projection, provided the projection matrix satisfies a Restricted Isometry Property (RIP). We establish new implications of the RIP with respect to nonlinear approximation in a Hilbert space with a redundant frame. The main ingredients of our approach are: a) Jackson and Bernstein inequalities, associated to the characterization of certain approximation spaces with interpolation spaces; b) a new proof that for overcomplete frames which satisfy a Bernstein inequality, these interpolation spaces are nothing but the collection of vectors admitting a representation in the dictionary with compressible coefficients; c) the proof that the RIP implies Bernstein
inequalities. As a result, we obtain that in most overcomplete random Gaussian dictionaries with fixed aspect ratio, just as in any orthonormal basis, the error of best $m$-term approximation of a vector decays at a certain rate if, and only if, the vector admits a compressible expansion in the dictionary. Yet, for mildly overcomplete dictionaries with a one-dimensional kernel, we give examples where the Bernstein inequality holds, but the same inequality fails for even the smallest perturbation of the dictionary. This work has been submitted for publication in a journal [102].

Performance guarantees for compressed sensing with spread spectrum techniques: We advocate a compressed sensing strategy that consists of multiplying the signal of interest by a wide bandwidth modulation before projection onto randomly selected vectors of an orthonormal basis. Firstly, in a digital setting with random modulation, considering a whole class of sensing bases including the Fourier basis, we prove that the technique is universal in the sense that the required number of measurements for accurate recovery is optimal and independent of the sparsity basis. This universality stems from a drastic decrease of coherence between the sparsity and the sensing bases, which for a Fourier sensing basis relates to a spread of the original signal spectrum by the modulation (hence the name "spread spectrum"). The approach is also efficient as sensing matrices with fast matrix multiplication algorithms can be used, in particular in the case of Fourier measurements. Secondly, these results are confirmed by a numerical analysis of the phase transition of the $l_1$-minimization problem. Finally, we show that the spread spectrum technique remains effective in an analog setting with chirp modulation for application to realistic Fourier imaging. We illustrate these findings in the context of radio interferometry and magnetic resonance imaging. This work has been presented at a conference [93] and accepted for publication in a journal [105].

Dictionary learning: An important practical problem in sparse modeling is to choose the adequate dictionary to model a class of signals or images of interest. While diverse heuristic techniques have been proposed in the literature to learn a dictionary from a collection of training samples, there are little existing results which provide an adequate mathematical understanding of the behaviour of these techniques and their ability to recover an ideal dictionary from which the training samples may have been generated.

In 2008, we initiated a pioneering work on this topic, concentrating in particular on the fundamental theoretical question of the identifiability of the learned dictionary. Within the framework of the Ph.D. of Karin Schnass, we developed an analytic approach which was published at the conference ISCCSP 2008 [13] and allowed us to describe "geometric" conditions which guarantee that a (non overcomplete) dictionary is "locally identifiable" by $l_1$ minimization.

In a second step, we focused on estimating the number of sparse training samples which is typically sufficient to guarantee the identifiability (by $l_1$ minimization), and obtained the following result, which is somewhat surprising considering that previous studies seemed to require a combinatorial number of training samples to guarantee the identifiability: the local identifiability condition is typically satisfied as soon as the number of training samples is roughly proportional to the ambient signal dimension. The outline of the second result was published in conferences [12], [25]. These results have been published in the journal paper [15].

This year we have worked on extending the results to noisy training samples with outliers. A journal paper is in preparation, and the results will be presented at a workshop at NIPS 2011.

Analysis Operator Learning for Overcomplete Cosparse Representations: Besides standard dictionary learning, we also considered learning in the context of the cosparse model. We consider the problem of learning a low-dimensional signal model from a collection of training samples. The mainstream approach would be to learn an overcomplete dictionary to provide good approximations of the training samples using sparse synthesis coefficients. This famous sparse model has a less well known counterpart, in analysis form, called the cosparse analysis model. In this new model, signals are characterized by their parsimony in a transformed domain using an overcomplete analysis operator. We proposed two approaches to learn an analysis operator from a training corpus, both published in the conference EUSIPCO 2011 [79], [67].

The first one uses a constrained optimization program based on $L_1$ optimization. We derive a practical learning algorithm, based on projected subgradients, and demonstrate its ability to robustly recover a ground truth analysis operator, provided the training set is of sufficient size. A local optimality condition is derived,
providing preliminary theoretical support for the well-posedness of the learning problem under appropriate conditions. Extensions to deal with noisy training samples are currently investigated, and a journal paper is in preparation.

In the second approach, analysis "atoms" are learned sequentially by identifying directions that are orthogonal to a subset of the training data. We demonstrate the effectiveness of the algorithm in three experiments, treating synthetic data and real images, showing a successful and meaningful recovery of the analysis operator.

**Connections between sparse approximation and Bayesian estimation:** Penalized least squares regression is often used for signal denoising and inverse problems, and is commonly interpreted in a Bayesian framework as a Maximum A Posteriori (MAP) estimator, the penalty function being the negative logarithm of the prior. For example, the widely used quadratic program (with an \( \ell^1 \) penalty) associated to the LASSO / Basis Pursuit Denoising is very often considered as MAP estimation under a Laplacian prior in the context of additive white Gaussian noise (AWGN) reduction.

A first result, which has been published in IEEE Transactions on Signal Processing [35], highlights the fact that, while this is one possible Bayesian interpretation, there can be other equally acceptable Bayesian interpretations. Therefore, solving a penalized least squares regression problem with penalty \( \phi(x) \) need not be interpreted as assuming a prior \( C \cdot \exp(-\phi(x)) \) and using the MAP estimator. In particular, we showed that for any prior \( P_X \), the minimum mean square error (MMSE) estimator is the solution of a penalized least square problem with some penalty \( \phi(x) \), which can be interpreted as the MAP estimator with the prior \( C \cdot \exp(-\phi(x)) \). Vice-versa, for certain penalties \( \phi(x) \), the solution of the penalized least squares problem is indeed the MMSE estimator, with a certain prior \( P_X \). In general \( dP_X(x) \neq C \cdot \exp(-\phi(x))dx \).

A second result, obtained in collaboration with Prof. Mike Davies and Prof. Volkan Cevher (a paper is under revision) characterizes the "compressibility" of various probability distributions with applications to underdetermined linear regression (ULR) problems and sparse modeling. We identified simple characteristics of probability distributions whose independent and identically distributed (iid) realizations are (resp. are not) compressible, i.e., that can be approximated as sparse. We prove that many priors which MAP Bayesian interpretation is sparsity inducing (such as the Laplacian distribution or Generalized Gaussian distributions with exponent \( p \leq 1 \)), are in a way inconsistent and do not generate compressible realizations. To show this, we identify non-trivial undersampling regions in ULR settings where the simple least squares solution outperform oracle sparse estimation in data error with high probability when the data is generated from a sparsity inducing prior, such as the Laplacian distribution.

6.2.3. Wavelets on graphs

**Participant:** Rémi Gribonval.

**Main collaboration:** Pierre Vandergeynst, David Hammond (EPFL)

Within the framework of the SMALL project 7.2.1, we investigated the possibility of developing sparse representations of functions defined on graphs, by defining an extension to the traditional wavelet transform which is valid for data defined on a graph.

There are many problems where data is collected through a graph structure: scattered or non-uniform sampling, sensor networks, data on sampled manifolds or even social networks or databases. Motivated by the wealth of new potential applications of sparse representations to these problems, the partners set out a program to generalize wavelets on graphs. More precisely, we have introduced a new notion of wavelet transform for data defined on the vertices of an undirected graph. Our construction uses the spectral theory of the graph laplacian as a generalization of the classical Fourier transform. The basic ingredient of wavelets, multi-resolution, is defined in the spectral domain via operator-valued functions that can be naturally dilated. These in turn define wavelets by acting on impulses localized at any vertex. We have analyzed the localization of these wavelets in the vertex domain and showed that our multi-resolution produces functions that are indeed concentrated at will around a specified vertex. Our theory allowed us to construct an equivalent of the continuous wavelet transform but also discrete wavelet frames.
Computing the spectral decomposition can however be numerically expensive for large graphs. We have shown that, by approximating the spectrum of the wavelet generating operator with polynomial expansions, applying the forward wavelet transform and its transpose can be approximated through iterated applications of the graph Laplacian. Since in many cases the graph Laplacian is sparse, this results in a very fast algorithm. Our implementation also uses recurrence relations for computing polynomial expansions, which results in even faster algorithms. Finally, we have proved how numerical errors are precisely controlled by the properties of the desired spectral graph wavelets. Our algorithms have been implemented in a Matlab toolbox that has been released in parallel to the main theoretical article [16]. We also plan to include this toolbox in the SMALL project numerical platform.

We now foresee many applications. On one hand we will use non-local graph wavelets constructed from the set of patches in an image (or even an audio signal) to perform de-noising or in general restoration. An interesting aspect in this case, would be to understand how wavelets estimated from corrupted signals deviate from clean wavelets. In a totally different direction, we will also explore the applications of spectral graph wavelets constructed from brain connectivity graphs obtained from whole brain tractography. Our preliminary results show that graph wavelets yield a representation that is very well adapted to how the information flows in the brain along neuronal structures.

6.2.4. Algorithmic breakthrough in sparse approximation : LoCOMP

Participants: Rémi Gribonval, Frédéric Bimbot, Ronan Le Boulch.

Main collaborations: Pierre Vandergheynst (EPFL), Boris Mailhé (former team member, now with Queen Mary University, London)

Our team had already made a substantial breakthrough in 2005 when first releasing the Matching Pursuit ToolKit (MPTK, see Section 5.3) which allowed for the first time the application of the Matching Pursuit algorithm to large scale data such as hours of CD-quality audio signals. In 2008, we designed a variant of Matching Pursuit called LoCOMP (ubiquitously for LOw Complexity Orthogonal Matching Pursuit or Local Orthogonal Matching Pursuit) specifically designed for shift-invariant dictionaries. LoCOMP has been shown to achieve an approximation quality very close to that of a full Orthonormal Matching Pursuit while retaining a much lower computational complexity of the order of that of Matching Pursuit. The complexity reduction is substantial, from one day of computation to 15 minutes for a typical audio signal [20], [19]. The main effort this year has been to integrate this algorithm into MPTK to ensure its dissemination and exploitation, and a journal paper has been published [22].

6.3. Emerging activities on compressive sensing and inverse problems

6.3.1. Nearfield acoustic holography (ECHANGE ANR project)

Participants: Rémi Gribonval, Nancy Bertin.

Main collaborations: Albert Cohen (Laboratoire Jacques-Louis Lions, Université Paris 6), Laurent Daudet, Gilles Chardon, François Ollivier, Antoine Peillot (Institut Jean Le Rond d’Alembert, Université Paris 6)

Compressed sensing is a rapidly emerging field which proposes a new approach to sample data far below the Nyquist rate when the sampled data admits a sparse approximation in some appropriate dictionary. The approach is supported by many theoretical results on the identification of sparse representations in overcomplete dictionaries, but many challenges remain open to determine its range of effective applicability. METISS has chosen to focus more specifically on the exploration of Compressed Sensing of Acoustic Wavefields, and we have set up the ANR collaborative project ECHANGE (ECHantillonage Acoustique Nouvelle GEnération) which began in January 2009. Rémi Gribonval is the coordinator of the project.

In 2010, the activity on ECHANGE has concentrated on Nearfield acoustic holography (NAH), a technique aiming at reconstructing the operational deflection shapes of a vibrating structure, from the near sound field it generates. In this application scenario, the objective is either to improve the quality of the reconstruction (for a given number of sensors), or reduce the number of sensors, or both, by exploiting a sparsity hypothesis which helps regularizing the inverse problem involved.
Contributions of the team in this task spans: notations and model definitions, experimental setting design and implementation, choice of an adapted dictionary in which the sparsity hypothesis holds, improved acquisition strategies through pseudo-random sensor arrays and/or spatial multiplexing of the inputs, experimental study of robustness issues, and theoretical study of potential success guarantees based on the restricted isometry property (which revealed being not verified in our case, despite improved experimental performance).

A paper about robustness issues and spatial multiplexing (an alternative to building antennas with random sensor position) was published in GRETSI [88]. A journal paper is under revision.

6.3.2. Sparse reconstruction for underwater acoustics (ECHANGE ANR project)

Participants: Rémi Gribonval, Valentin Emiya, Nikos Stefanakis, Nancy Bertin.

Main collaborations: Jacques Marchal, Pierre Cervenka (UPMC Univ Paris 06)

Underwater acoustic imaging is traditionally performed with beamforming: beams are formed at emission to insonify limited angular regions; beams are (synthetically) formed at reception to form the image. We proposed to exploit a natural sparsity prior to perform 3D underwater imaging using a newly built flexible-configuration sonar device. The computational challenges raised by the high-dimensionality of the problem were highlighted, and we described a strategy to overcome them. As a proof of concept, the proposed approach was used on real data acquired with the new sonar to obtain an image of an underwater target. We discussed the merits of the obtained image in comparison with standard beamforming, as well as the main challenges lying ahead, and the bottlenecks that will need to be solved before sparse methods can be fully exploited in the context of underwater compressed 3D sonar imaging. This work has been submitted to ICASSP 2012 and a journal paper is in preparation.

6.3.3. Audio inpainting (SMALL FET-Open project)

Participants: Rémi Gribonval, Valentin Emiya.

Main collaborations: Amir Adler, Michael Elad (Computer Science Department, The Technion, Israel); Maria G. Jafari, Mark D. Plumbley (Centre for Digital Music, Department of Electronic Engineering, Queen Mary University of London, U.K.).

Inpainting is a particular kind of inverse problems that has been extensively addressed in the recent years in the field of image processing. It consists in reconstructing a set of missing pixels in an image based on the observation of the remaining pixels. Sparse representations have proved to be particularly appropriate to address this problem. However, inpainting audio data has never been defined as such so far.

METISS has initiated a series of works about audio inpainting, from its definition to methods to address it. This research has begun in the framework of the EU Framework 7 FET-Open project FP7-ICT-225913-SMALL (Sparse Models, Algorithms and Learning for Large-Scale data) which began in January 2009. Rémi Gribonval is the coordinator of the project. The research on audio inpainting has been conducted by Valentin Emiya in 2010 and 2011.

The contributions consist of:

- defining audio inpainting as a general scheme where missing audio data must be estimated: it covers a number of existing audio processing tasks that have been addressed separately so far – click removal, declipping, packet loss concealment, unmasking in time-frequency;
- proposing algorithms based on sparse representations for audio inpainting (based on Matching Pursuit and on $l_1$ minimization);
- addressing the case of audio declipping (i.e. desaturation): thanks to the flexibility of our inpainting algorithms, they can be constrained so as to include the structure of signals due to clipping in the objective to optimize. The resulting performance are significantly improved. This work has been reported in [47] and it will appear as a journal paper [96].
Current and future works deal with developing advanced sparse decomposition for audio inpainting, including several forms of structured sparsity (e.g., temporal and multichannel joint-sparsity) and several applicative scenarios (declipping, time-frequency inpainting).

6.4. Music Content Processing and Music Information Retrieval

6.4.1. Acoustic music modeling

Participants: Nancy Bertin, Emmanuel Vincent.

Main collaborations: R. Badeau (Télécom ParisTech), J. Wu (University of Tokyo)

Music involves several levels of information, from the acoustic signal up to cognitive quantities such as composer style or key, through mid-level quantities such as a musical score or a sequence of chords. The dependencies between mid-level and lower- or higher-level information can be represented through acoustic models and language models, respectively.

Our acoustic models are based on nonnegative matrix factorization (NMF) and variants thereof. NMF models an input short-term magnitude spectrum as a linear combination of magnitude spectra, which are adapted to the input under suitable constraints such as harmonicity and temporal smoothness. While our previous work considered harmonic spectra only, we proposed the use of wideband spectra to represent attack transients and showed that this resulted in improved pitch transcription accuracy [77]. Our past work on the convergence properties of NMF was also disseminated [50].

We used the resulting model parameters to identify the musical instrument associated with each note, by means of a Support Vector Machine (SVM) classifier trained on solo data, and obtained improved instrument classification accuracy compared to state-of-the-art Mel-Frequency Cepstral Coefficient (MFCC) features [42], [78].

6.4.2. Music language modeling

Participants: Frédéric Bimbot, Emmanuel Vincent.

Main collaboration: S.A. Raczynski (University of Tokyo, JP)

We pursued our pioneering work on music language modeling, with a particular focus on the joint modeling of "horizontal" (sequential) and "vertical" (simultaneous) dependencies between notes by log-linear interpolation of the corresponding conditional distributions. We identified the normalization of the resulting distribution as a crucial problem for the performance of the model and proposed an exact solution to this problem.

We also applied the log-linear interpolation paradigm to the joint modeling of melody, key, chords and meter, which evolve according to different timelines. In order to synchronize these feature sequences, we explored the use of beat-long templates consisting of several notes as opposed to short time frames containing a fragment of a single note.

Both of these studies are ongoing.

6.4.3. Music structuring

Participants: Frédéric Bimbot, Gabriel Sargent, Emmanuel Vincent.

External collaboration: Emmanuel Deruty (as an independent consultant)

The structure of a music piece is a concept which is often referred to in various areas of music sciences and technologies, but for which there is no commonly agreed definition. This raises a methodological issue in MIR, when designing and evaluating automatic structure inference algorithms. It also strongly limits the possibility to produce consistent large-scale annotation datasets in a cooperative manner.

We have pursued our investigations on autonomous and comparable blocks, based on principles inspired from structuralism and generativism for producing music structure annotation. This work has allowed consolidating the methodology and producing additional annotations (over 400 pieces) [53].
We have also developed an algorithm aiming at the automatic inference of autonomous and comparable blocks using the timbral and harmonic content of music pieces, in combination with a regularity constraint [72]. Tested within the QUAERO project and during the MIREX 2011 campaign [94], the algorithm ranked among state-of-the-art methods.

6.5. Source separation

6.5.1. A general framework for audio source separation

**Participants:** Alexis Benichoux, Frédéric Bimbot, Charles Blandin, Ngoc Duong, Rémi Gribonval, Nobutaka Ito, Alexey Ozerov, Emmanuel Vincent.

*Main collaborations: H. Tachibana (University of Tokyo, JP), N. Ono (National Institute of Informatics, JP)*

Source separation is the task of retrieving the source signals underlying a multichannel mixture signal. The state-of-the-art approach, which we presented in a survey chapter [95], consists of representing the signals in the time-frequency domain and estimating the source coefficients by sparse decomposition in that basis. This approach relies on spatial cues, which are often not sufficient to discriminate the sources unambiguously. Last year, we proposed a general probabilistic framework for the joint exploitation of spatial and spectral cues [39] that was disseminated in several invited talks [43], [44]. This framework relies in particular on the thesis of Ngoc Duong, which was defended this year [30]. It makes it possible to quickly design a new model adapted to the data at hand and estimate its parameters via the EM algorithm. As such, it is expected to become the basis for a number of works in the field, including our own.

Since the EM algorithm is sensitive to initialization, we devoted a major part of our work to reducing this sensitivity. One approach is to set probabilistic priors over the model parameters, including spatial position priors [56] or temporal continuity priors [55]. A complementary approach is to initialize the parameters in a suitable way using source localization techniques specifically designed for environments involving multiple sources and possibly background noise [33], [54], [83]. In a longer-term perspective, we also investigated the design and exploitation of sparsity priors over time-domain acoustic transfer functions [52], [82].

6.5.2. Exploiting filter sparsity for source localization and/or separation

**Participants:** Alexis Benichoux, Prasad Sudhakar, Emmanuel Vincent, Rémi Gribonval, Frédéric Bimbot.

*Main collaboration: Simon Arberet (EPFL)*

Estimating the filters associated to room impulse responses between a source and a microphone is a recurrent problem with applications such as source separation, localization and remixing.

We considered the estimation of multiple room impulse responses from the simultaneous recording of several known sources. Existing techniques were restricted to the case where the number of sources is at most equal to the number of sensors. We relaxed this assumption in the case where the sources are known. To this aim, we proposed statistical models of the filters associated with convex log-likelihoods, and we proposed a convex optimization algorithm to solve the inverse problem with the resulting penalties. We provided a comparison between penalties via a set of experiments which shows that our method allows to speed up the recording process with a controlled quality tradeoff. This work was presented at two conferences [52], [82] and a journal paper including extensive experiments with real data is in preparation.

We also investigated the filter estimation problem in a blind setting, where the source signals are unknown. We proposed an approach for the estimation of sparse filters from a convolutive mixture of sources, exploiting the time-domain sparsity of the mixing filters and the sparsity of the sources in the time-frequency (TF) domain. The proposed approach is based on a wideband formulation of the cross-relation (CR) in the TF domain and on a framework including two steps: (a) a clustering step, to determine the TF points where the CR is valid; (b) a filter estimation step, to recover the set of filters associated with each source. We proposed for the first time a method to blindly perform the clustering step (a) and we showed that the proposed approach based on the wideband CR outperforms the narrowband approach and the GCC-PHAT approach by between 5 dB and 20 dB. This work has been published at ICASSP 2011 [49] and submitted for publication as a journal paper.
On a more theoretical side, we studied the frequency permutation ambiguity traditionally incurred by blind convolutive source separation methods. We focussed on the filter permutation problem in the absence of scaling, investigating the possible use of the temporal sparsity of the filters as a property enabling permutation correction. The obtained theoretical and experimental results highlight the potential as well as the limits of sparsity as an hypothesis to obtain a well-posed permutation problem. This work has been submitted for publication as a journal paper [99].

6.5.3. Towards real-world separation and remixing applications

Participants: Valentin Emiya, Alexey Ozerov, Laurent Simon, Emmanuel Vincent.

Shoko Araki (NTT Communication Science Laboratories, JP), Cédric Févotte (Télécom ParisTech, FR), Antoine Liutkus (Télécom ParisTech, FR), Volker Hohmann (University of Oldenburg, DE)

Following our founding role in the organization of a regular source separation evaluation campaign (SiSEC), we wrote an invited paper summarizing the outcomes of the three latest campaigns [41]. While some challenges remain, this paper highlighted that progress has been made and that audio source separation is closer than ever to successful industrial applications. This is also exemplified by the i3DMusic project and the contract recently signed with MAIA Studio.

In order to exploit our know-how for these real-world applications, we investigated issues such as how to implement our algorithms in real time and how best to exploit human input or metadata [68], [70]. In addition, while the state-of-the-art quality metrics previously developed by METISS remain widely used in the community, we proposed a new set of perceptually motivated metrics which greatly increase correlation with subjective assessments [34].

6.5.4. Source separation for multisource content indexing

Participants: Kamil Adiloglu, Alexey Ozerov, Emmanuel Vincent.

Main collaborations: J. Barker (University of Sheffield, UK), M. Lagrange (IRCAM, FR)

Another promising real-world application of source separation concerns information retrieval from multi-source data. Source separation may then be used as a pre-processing stage, such that the characteristics of each source can be separately estimated. The main difficulty is not to amplify errors from the source separation stage through subsequent feature extraction and classification stages. To this aim, we proposed a principled Bayesian approach to the estimation of the uncertainty about the separated source signals [45] and propagated this uncertainty to the features. We then exploited it in the training of the classifier itself, thereby greatly increasing classification accuracy [69].

While our work in this direction was initially motivated by music applications (e.g. artist recognition by separating the vocals from the accompaniment), we eventually applied it to noise-robust speech recognition, which is a better defined task [71]. In order to motivate further work by the community, we created a new international evaluation campaign on that topic (CHiME) [86].
6. New Results

6.1. Motion Sensing and analysis

6.1.1. Sensing human activity for detecting falling motions

Participants: Franck Multon [contact], Richard Kulpa, Anthony Sorel, Edouard Auvinet.

Sensing human activity is a very active field of research, with a wide range of applications ranging from entertainment and serious games to personal ambient living assistance. MimeTIC aims at proposing original methods to process raw motion capture data in order to compute relevant information according to the application.

In personal ambient living monitoring, we have collaborated with University of Montreal, Department of Computer Science and Operations Research (DIRO) which main activity is biomedical engineering. A co-supervised student is addressing two complementary problems: detecting people falling in everyday environment and providing easy-to-use clinical gait analysis systems for early detection of potential risks of falling. These two problems have been addressed by reconstructing the visual hull of a subject according to synchronized classical of depth cameras. As visual hull is based on videos it’s subject to occlusions which generally occur in natural environment, such as a room with furniture. We have adapted the classical visual hull algorithm in order to be less sensible to occlusions. We also have proposed an index based on 3D silhouette vertical distribution which enhance this property to tackle occlusion problems [1]. This index is based on a ratio: the volume above a given threshold divided by the total body volume. It has been successfully applied to dozens of falling scenarios involving natural occlusions with furniture. The second problem consists in extracting relevant indexes in gait that could enable clinicians to identify elderly people who have a risk of falling. Classical indexes are based on gait regularity and asymmetry in dual tasks protocols (such as walking while counting downward). 3D silhouettes intrinsically contain all the required information in a unique representation contrary to multi-point motion capture systems. However extracting the relevant information from 3D volumes is complex. We have proposed an original approach based on statistical analysis of the volumes in order to compute indexes for gait asymmetry while simply using 3 depth-cameras (Microsoft Kinects) [1] (see figure 4).

Figure 4. 3D silhouettes reconstructed with three depth-cameras in order to analyze gait asymmetry.
In entertainment and serious games, the problem is different as we need to accurately know the action performed by the user in order to react in a convenient manner. Collaboration with Artefacto Company enabled us to develop such motion recognition methods in serious games scenarios. Given motion capture data provided by an optical motion capture system lead to large state vectors in which the relevant information is hidden. Mixture of Gaussians is generally used as an input of Hidden Markov Models to recognize a motion according to this raw data. To simplify, features are generally introduced in order to capture the relevant geometrical property of the motion with either general information (such as joint angles or Cartesian positions) or application-specific information. The former type of information has the advantage to be generic but leads to recognizers that are very sensitive to style and morphology variations. We have proposed a new generic feature based on morphology-independent representation that enables to tackle this problem (submitted to Eurographics2012). The recognition rate is above 75% for very similar upper-limb motions (see figure 5) while classical methods fail to recognize the same motions (recognition rate below 50%).

![Figure 5. The motion of a user equipped with reflective markers is automatically recognized with a HMM system based on morphology-independent features.](image)

6.1.2. The Joyman: a novel immersive locomotion device for virtual environments

**Participant:** Julien Pettré [contact].

We proposed a novel interface called Joyman, designed for immersive locomotion in virtual environments [18]. Whereas many previous interfaces preserve or stimulate the users proprioception, the Joyman aims at preserving equilibrioception in order to improve the feeling of immersion during virtual locomotion tasks. The proposed interface is based on the metaphor of a human-scale joystick. The device has a simple mechanical design that allows a user to indicate his virtual navigation intentions by leaning accordingly. We also propose a control law inspired by the biomechanics of the human locomotion to transform the measured leaning angle into a walking direction and speed - i.e., a virtual velocity vector. We aim at using this interface to enable natural interaction with virtual humans with low-cost devices. The Joyman is patented and was presented at the Emerging Technologies, Siggraph Asia, Hong-Kong [20].

These are joint results with the VR4i team (Anatole Lécuyer and Maud Marchal).

6.2. VR and Sports

**Participants:** Richard Kulpa, Benoit Bideau [contact], Brault Sébastien, Burns Anne-Marie.
In the past, we have worked on the interaction between two opponents in handball. We have designed a framework to animate virtual throwers in a reality center and to analyze the gestures of real goalkeepers whose objective was to intercept the corresponding virtual balls. This VR framework was then validated by showing that behaviors in real and virtual environment were similar. These works have been extended by using perception-action coupling and perception-only studies to evaluate the anticipation of opponents. In order to evaluate the importance of perceived parameters, the ball and/or the character animation was successively hidden to determine their importance and the same kind of study was done on the graphical level of details.

These works have been extended to the study of deceptive movements and gaps evaluation in rugby. Combining perceptual analysis based on the use of cutoffs with biomechanical analysis, we have extracted important kinematic information that could explain differences between experts and novices. Indeed, thanks to the cutoffs, it is possible to determine how early each of these two levels of practice can perceive the correct final direction of the opponent. Then this information is correlated to kinematical parameters of this player. Finally, we have embedding these knowledge on the evaluation of novices and expertis to create models of rugby defenders. We are currently working on integrating these models in a VR experiment in which the real user is this time the attacker and our model the virtual defender.

Concurrently, studies are experimented to determine if VR can be used for training in sports [9]. The first step was to compare if trainees learned the same way in real situation, facing a video of the lesson or facing a virtual teacher that is animated from the motions of the real course. Based on evaluation of an expert, the results showed that the three groups evolved the same way and reached the same level of practice. The second step is then to have more experts to complete the evaluation but also to combine these results with objective analyses based on kinematics data.

This work is partially funded by the Biofeedback project (DGCIS "Serious Gaming" project) of the M2S laboratory, University Rennes 2. Its goal was to create a training tool that can be used and configured by coaches in order to train athletes to repetitive motions such as katas in karate. The evaluation is made by an automatic module that compares the kinematics of the trainee to a database of expert movements.

6.3. Crowds

Participants: Julien Pettré [contact], Richard Kulpa, Anne-Hélène Olivier, Samuel Lemercier, Yijiang Zhang, Jonathan Perrinet.

6.3.1. Perception of collision in crowds

We designed a level-of-detail (LOD) selection function to determine where collision avoidance constraints in crowd simulation can be relaxed without being perceived by spectators [4]. Collision avoidance is probably the most time consuming parts of crowd simulator. However, when only believable results are required, we argue that visually similar results can be obtained a low computational costs based on macroscopic crowd simulation. Based on a perception study, we determined the conditions for collision to be or not to be detected. We discovered that the camera tilt angle was playing a great effect on perception, whereas distance to camera (usually used in previous works) was only the third most important factor to be considered.

6.3.2. Mixed Reality Crowds

In the task of making virtual crowds and real people interact together, we explore a mixed reality solution [22]. The seamless integration of virtual characters into dynamic scenes captured by video is a challenging problem. In order to achieve consistent composite results, both the virtual and real characters must share the same geometrical space and their interactions must adhere to the physical coherence criteria. One essential question is how to detect the motion of real objects - such as real characters moving in the video - and how to steer virtual characters accordingly to avoid unrealistic collisions. We propose an online solution. First, by analysis of the input video, the motion states of the real objects are recovered into a common world 3D coordinate system. Meanwhile, a simplified accuracy measurement is defined to represent the confidence of the motion estimate. Then, under the constraints imposed by the real dynamic objects, the motion of virtual characters are accommodated by a uniform steering model. The final step is to merge virtual objects back the
real video scene by taking into account visibility and occlusion constraints between real foreground objects and virtual ones.

6.3.3. Experiments on crowds

Evaluating crowd simulation models is a difficult task. In the frame of the ANR PEDIGREE project, we put in a lot of effort to perform experiments on groups of walking people in order to dispose of a reference database on groups motion. In order to obtain high-quality data, we measure people locomotion by using optoelectronic motion capture systems. In 2011, we starting obtaining detailed analysis on such motion after large efforts on motion analysis and processing. We had to develop dedicated reconstruction techniques because of the challenging conditions in which we performed our motion capture [12]. We submitted two papers on following modeling and simulation stages (submitted to Eurographics 2012 and Physical Review E).

6.4. Interactive Virtual Cinematography

Participants: Marc Christie [contact], Christophe Lino.

The domain of Virtual Cinematography explores the operationalization of rules and conventions pertained to camera placement, light placement and staging in virtual environments. In 2011, two major challenges were tackled (i) the proposition of intelligent interactive assistants integrating users in the process of selecting viewpoints and editing a virtual movie, with the capability of adapting to the user choices, and (ii) the design and implementation of evaluation functions for precisely ranking the quality of viewpoints of a virtual 3D environment.

Our intelligent assistant is designed around (i) an intelligent cinematography engine that can compute, at the request of the filmmaker, a set of suitable camera placements (called suggestions) for starting a shot, representing semantically and cinematically distinct choices for visualizing the current narrative, considering established cinema conventions of continuity and composition along with the filmmaker’s previous selected suggestions, and also his or her manually crafted camera compositions, by a machine learning component that adapts shot editing preferences from user-created camera edits, (ii) a user interface, where the suggestions are presented as small movie frames, arranged in a grid whose rows and columns correspond to visual composition properties of the suggested cameras, and (iii) a motion-tracked camera system that makes the user able of modifying the low-level parameters of the camera in shots in the same way a real operator would. The result of this work [16] is a novel workflow based on interactive collaboration of human creativity with automated intelligence that enables efficient exploration of a wide range of cinematographic possibilities, and rapid production of computer-generated animated movies. A full prototype has been built and demonstrated at ACM Multimedia conference [15] as well as ParisFX conference. A patent protecting this technology is currently under evaluation [25].

The second challenge is related to the design of efficient and precise metrics for measuring the quality of viewpoints. For efficiency, we have proposed parallel GPU-based evaluation techniques for the estimation of multiple viewpoints [8] coupled within a Genetic Algorithm (Swarm Particle Optimization) to rapidly explore the space of possible viewpoints. For preciseness, we have designed a large range of quality functions relative to screen composition and transition between shots, and employed these functions to either automatically generate movies from actions occurring in the virtual environment [13] or interactively generating movies by letter the users select best shots and best transitions between shots [14].

Finally we have been exploring the use of tactile devices to the interactive construction of narratives following Prop’s computational model of stories [10].

6.5. Biomechanics and Motion Analysis

6.5.1. Balance in highly dynamic situations

Participants: Franck Multon [contact], Ludovic Hoyet.
Balance is a key problem in humans as people stand on two feet which leads to a small base of support area compared to the overall body volume. This unstable state has been widely analyzed in static situation but is still difficult to understand when velocity and acceleration reach ineligible values. We thus have proposed an experimental protocol in order to evaluate if criteria published in the literature for specific motions could be generalized to any dynamic motions (see figure 7). To this end, each studied criterion was tested on various dynamic motions and the number of false falling alarms was reported in each case: the number of frames where the criterion detects loss of balance while the subject is actually balanced. The tested criteria where: the projection of the center of mass on the ground which should remain in the base of support, the Zero Moment Point widely used in robotics, the Zero Rate of Angular Momentum, the Foot Rotation Index and the extrapolated center of mass which was introduced in biomechanics recently. The results demonstrate that none of the criteria succeeded in correctly predicting loss of balance in highly dynamic motions [7]. It thus demonstrate the need to continue some fundamental work on this topic which is a key problem in many applications, including robotics, detection and prevention of falls in the Elderly, understanding performance in sports, improving realism in virtual human simulation...

**Figure 6.** Director’s Lens: a new workflow for interactive virtual cinematography. For a spatial partition analysis of a 3D environment, our tools automatically generates ranges of viewpoint suggestions following classical cinematographic conventions. Cameramen then select the suggestions and refine them using a 6DOF tracked device.

**Figure 7.** Dynamic motions that have been used to evaluate balance criteria published in previous works.

**6.5.2. Interaction strategies between two walkers to avoid collision**
Participants: Armel Crétual, Julien Pettré [contact], Anne-Hélène Olivier, Jan Ondrej, Antoine Marin.

In the everyday life situation where two humans walk in the same nearness, each can be considered as a moving obstacle for the other one. They adapt their locomotion with respect to this external disturbance to avoid any collision. Collision avoidance between two humans has been largely neglected in the literature despite it lets us expect specific interactions. The main question we raised was to identify the conditions that induce avoidance manoeuvres in locomotor trajectories: what are the relations between the respective positions and velocities which induce motion adaptations? To answer this question, we proposed an original experiment: thirty participants were asked to walk two-by-two in a motion captured area. We assigned them locomotion tasks in order to provoke varied situations of potential future collisions (see figure 8 a). Following the hypothesis of a reciprocal interaction, we suggested a variable which is common to both of the walkers, the Minimum Predicted Distance (MPD), to predict the actual presence of physical interactions as well as to describe their properties. At each instant t, MPD was computed as the distance the walkers would meet if they did not perform motion adaptation after this instant t. Results showed that walkers adapted their motions only when required, i.e., when initial MPD was too low (<1m). We concluded that human walkers are able to accurately estimate future crossing distance and to mutually adapt it. The evolution of MPD enabled decomposing collision avoidance into 3 successive phases: observation, reaction, and regulation (see figure 8 b). Respectively, these phases corresponded to periods of time when, first, MPD was constant, second, increased to acceptable values by motion adaptation and, third, reached a plateau and even slightly decreased. This final phase demonstrates that collision avoidance is actually performed with anticipation. Future work is needed to inspect individual motion adaptations and to relate them with variation of MPD.

Figure 8. a) Experimental set-up to study collision avoidance between 2 walkers. b) Mean MPD evolution for trials for which initial MPD is below 1m. Interaction follows 3 successive phases: observation, reaction and regulation phases. Normalized time was computed between the time when participants were able to see each other to the time they crossed.

6.5.3. Quantification of pathological gait in adults

Participants: Armel Crétau [contact], Kristell Bervet.

Quantifying gait deviation is still a challenge in adults patient follow-up within a rehab process. This quantification can be done on several levels. Among them, the most useful for practitioners are surely kinematics and muscular activation. On the first one, Gillette Gait Index (GGI) has now become a common tool in rehabilitation centers to assess gait abnormalities. However, one limitation of this index is that it is based on some peak values and is thus sensitive to measurement noise. A new index, the Gait Deviation Index (GDI) which is based on joints angles patterns has been developed by the same team to avoid this problem. Nevertheless, both of them have only been validated in children with cerebral palsy. On the second level, no
satisfying global index has yet been developed. The first part of our study was to validate the GDI in adults. From a database of 74 healthy subjects and 48 patients we did demonstrate that GDI is a relevant index to quantify kinematic gait pattern in adults. Then, we developed a new index called KeR-EGI (for Kerpa-Rennes EMG-based Gait Index) which accounts for the muscular activation patterns of patients. There also, results were conclusive relying on a good correlation between GDI and KeR-EGI. Finally, our recommendation to practitioners was to use both of these two index as they account for complementary aspects of pathology. This allow to better understand if gait disorder is more due to neurological injury or on the opposite to mechanical constraints such as joint stiffness.

6.5.4. Modeling gesture in sports: tennis serve

Participants: Nicolas Bideau [contact], Guillaume Nicolas, Benoit Bideau, Caroline Martin, Richard Kulpa.

Most experimental studies on tennis focus on the segmental coordination in connection with the ball speed, but do not consider the resulting traumatisms. To this end, we currently develop an inverse dynamics modeling approach, based on musculoskeletal parameters. As a first step to this work, we calculated the joint constraints on the upper limb in the tennis serve, for professional and regional players. Eleven high level, professional players were compared to seven regional players during this specific motion. Each player was equipped with 42 reflective markers and tennis serve was analyzed using an optoelectronical system composed of 12 infrared cameras cadenced at 300Hz. For each player, values of force, power and internal work (in absolute value) were calculated for the three joints (shoulder, elbow, wrist) using a musculoskeletal model. The results showed that professional players produced higher power and internal work for each joint in comparison with the regional players. Results also showed a decrease in the values of internal work from the distal (wrist joint) to the proximal joint (shoulder joint). These results may explain shoulder pain in tennis, which is commonly depicted for high-level players in epidemiological studies. A first perspective to this work is to better take individual parameters (inertia, muscle parameters, pre-constraint, etc.) into account. Another perspective is to test various objective functions in order to predict which parameter is optimized during tennis serve.

6.5.5. Modeling gesture in sports: fin swimming

Participants: Nicolas Bideau [contact], Guillaume Nicolas, Benoit Bideau, Richard Kulpa.

In swimming, experimental approaches are commonly used to analyze performance. However, due to obvious limitations in experimental approaches (impossibility to standardize any situations etc.), it is difficult to characterize surrounding fluid. To overcome this limitation, we currently develop analysis, modeling and simulation of aquatic locomotion, using CFD computer simulation and new methods based on animation of virtual characters.

A first application of this topic enables to evaluate the influence of swim fin flexibility on efficiency during swimming based on a CFD structure interaction model. Finite elements simulations are carried out for various material properties and various prescribed kinematics. Besides the significant effect of flexibility on propulsive forces, the results indicate that the propulsive efficiency is greatly influenced by the stroke frequency and the initial angle of attack. For the selected material properties, the results show that efficiency increases from 3.6 percents to 11.9 percents when the stroke frequency is increased from 0 to 1.7 Hz. Moreover efficiency is clearly increased from 5.0 percents to 24.2 percents when increasing the angle of attack from 0 to 45 degrees. Therefore, an interesting prospect of the present work could be an enhancement of the design of better performing swim fins.

A second application of this topic related to aquatic propulsion deals with a new method to evaluate cross-sectional area based on computer animation of swimming. Indeed, reducing cross sectional area (CSA) during starts and turns is a key part of performance optimisation. Different methods have been used to obtain this parameter without any standard: total human body volume to the power 2/3, wetted area or frontal area based on planimetry technique (PT). These different methods can lead to discrepancies in drag values. Recently, we used two synchronized camcorders to evaluate drag parameters during the different phases of an undulatory stroke cycle.
However, such a technique needs accurate synchronization and calibration of the different camcorders views. The aim of this study is to provide a new method based on animation of virtual characters to obtain instantaneous cross-sectional area in an undulatory stroke cycle. Its main advantage is to obtain cross-sectional area as well as biomechanical analysis with a single camcorder in a sagittal plan and without space calibration. From this, we intend to better understand swimming hydrodynamics and the way CSA influences active drag. More generally, this approach has been designed to provide new practical insights into swimming analysis protocols.

6.6. Path planning and environment analysis

6.6.1. Space-Time planning in dynamic environments

Participants: Fabrice Lamarche [contact], Thomas Lopez.

When automatically populating 3D geometric databases with virtual humanoids, modeling the navigation behavior is essential since navigation is used in most exhibited behaviors. In many application fields, the need to manage navigation in dynamic environments arises (virtual worlds taking physics laws into account, numerical plants in which step stools can be moved,...). This study focuses on the following issue: how to manage the navigation of virtual entities in such dynamic environments where topology may change at any time i.e. where unpredictable accessibility changes can arise at runtime. In opposition to current algorithms, movable items are not only considered as obstacles in the environment but can also help virtual entities in their navigation.

The proposed algorithm [17] splits that problem into two complementary processes: a topology tracking algorithm and a path planning algorithm. The aim of the topology tracking algorithm is to continuously detect and update topological relations between moving objects i.e. accessibility or obstruction, while storing temporal information when recurring relations are observed. The path planning algorithm uses this information to plan a path inside the dynamic environment. The coupling of those algorithms endows a virtual character with the ability to immediately use inserted / moved object to reach previously unreachable locations. Moreover, this algorithm is able to find a path through moving platforms to reach a target located on a surface that is never directly accessible.

6.6.2. Automated environment analysis

Participants: Fabrice Lamarche [contact], Carl-Johan Jorgensen.

To populate a virtual environment, modeling the navigation behavior is crucial. This behavior relies on the ability of planning a path inside a complex environment, which itself relies on an adequate representation of the environment structure. Most often, virtual environments are represented has 3D databases that are analyzed to produce data structures that are suitable for path planning and navigation. However, without any user intervention, those data structures lack of information about the nature of identified navigable zones that are crucial for navigation credibility.

We proposed an environment analysis algorithm [11] that automatically extracts a meaningful spatial representation of 3D virtual environments, suitable for spatial reasoning. This algorithm automatically differentiates indoor, outdoor and covered parts of the environment. It separates buildings into floors linked by stairs and represent floors as rooms linked by doorsteps. On this basis, we propose a natural hierarchical representation of the environment. This representation is used for spatial reasoning including zone selection and multi-criterion path planning that enhances path credibility.
6. New Results

6.1. Analysis and modeling for compact representation and navigation

6.1.1. Joint projection/filling method for virtual view synthesis

Participants: Christine Guillemot, Vincent Jantet.

This study is carried out in collaboration with INSA/IETR (Luce Morin). Associated with a view synthesis method, a multi-view plus depth video allows the generation of virtual views of the scene from any viewpoint. Many algorithms have thus been developed to synthesize virtual views from one or several input views video plus depth data. These rendering algorithms are either based on Image-Based Rendering (IBR) techniques or Geometry-Based Rendering (GBR) techniques, according to the amount of 3D information they use. IBR techniques require limited geometric information to synthesize intermediate views and allow the generation of photo-realistic virtual views at the expense of virtual camera freedom. GBR techniques require detailed 3D models of the scene to synthesize arbitrary viewpoints (points of view). GBR techniques are sensitive to the accuracy of the 3D model, which is difficult to estimate from real multi-view videos. Depth-Image-Based Rendering (DIBR) techniques include hybrid rendering methods between IBR and GBR techniques. DIBR methods are based on warping equations, which project a reference view onto a virtual viewpoint. Each input view is defined by a "color" (or "texture") map and a "depth" map, which associates a depth value to each image pixel.

In classical DIBR schemes, the rendering proceeds in several distinct steps, each one designed to solve a specific problem. First, the input depth map is warped onto the virtual viewpoint. The obtained warped depth map contains disocclusions, cracks and ghosting artifacts. Second, this virtual depth map is filtered with a median filter, in order to remove the cracks, and then to dilate disocclusion areas on the background side, in order to avoid ghosting artifacts during view synthesis. Third, the filtered depth map is used in a backward warping to compute the color of each pixel of the virtual view. Fourth, this resulting depth map is inpainted, to fill in disocclusion areas. Finally, this complete depth map is used by a depth-aided inpainting algorithm to fill in disocclusions in the color map. However, all these steps are inter-dependent, and errors introduced by each one are amplified by the following one. Connectivity information is lost during the first projection step, as shown in Fig. 2. Without this connectivity information, every inpainting method fails to fill in background disocclusions if the disoccluded area is surrounded by foreground objects. This case may happen each time a foreground object is not convex, and contains holes, as shown in Fig. 2 -(a). As a result, depth-aided inpainting uses wrong foreground patches to fill in background disocclusions, producing annoying artifacts, as shown in Fig. 2 -(b).

We have developed two DIBR techniques, both based on a novel forward projection technique, called the Joint Projection Filling (JPF) method [16]. The JPF method performs forward projection, using connectivity information to fill in disocclusions in a single step. The first proposed DIBR method is designed to extrapolate virtual views from a single input view plus depth video sequence. The synthesis of virtual depth maps by the JPF method avoids the use of dedicated filtering and inpainting processes and leads to synthesized depth maps of higher quality. The second proposed DIBR method is designed to interpolate intermediate views from multiple input view plus depth sequences. This interpolation method uses the Floating Texture approach to register multiple inputs view plus depth sequences before blending. The JPF method fills in disocclusion areas during the projection, to ensure that geometrical structures are well preserved. The method uses the occlusion-compatible ordering presented by McMillan, which uses epipolar geometry to select a pixel scanning order. The occlusion-compatible ordering is used to handle disocclusions gracefully. Cracks are filled in by interpolation of neighboring pixels, whereas disocclusions are only filled in by background pixels. This technique can also be used with non-rectified views, avoiding prior creation of parallax maps.
Figure 2. Virtual depth map synthesized by three forward projection methods. The point-based projection method generates cracks and disocclusions (top-left). Median filtering and directional inpainting fills holes with foreground depth (top-middle). The proposed JPF method fills cracks and disocclusions with realistic background (top-right). Synthesized view with disocclusion (bottom-left), Synthesized depth maps, obtained with a Navierstokes inpainting algorithm (column 2), with the developed JPF method (column 4) and the corresponding synthesized views with the two depth maps.

6.1.2. 2D/3D image inpainting for virtual view synthesis

Participants: Josselin Gauthier, Christine Guillemot, Mouid Keskes, Olivier Le Meur.

Inpainting methods play an important role in a wide range of applications. Removing text and advertisements (such as logos), removing undesired objects, noise reduction and image reconstruction from incomplete data are the key applications of inpainting methods. Algorithms can be classified into two categories: PDE (Partial Derivative Equation)-based schemes and examplar-based schemes. The former uses diffusion schemes in order to propagate structures in a given direction. Their drawback is the introduction of blur due to diffusion. The latter relies on the sampling and the copying of texture from the known parts of the picture.

We have proposed a novel inpainting algorithm combining the advantages of both aforementioned methods. As in Criminisi et al’s approach, the proposed method involves two steps: first, a filling order is defined to favor the propagation of structure in the isophote direction. Second, a template matching is performed in order to find the best candidates to fill in the hole. Compared to previous approaches, the main contributions concern the use of structure tensors to define the filling order instead of field gradients. The structure tensor is defined as follow:

$$J = \sum_{i=\{R, G, B\}} \nabla I_i \nabla I_i^T$$

(1)

$J$ is the sum of the scalar structure tensors $\nabla I_i \nabla I_i^T$ of each image channel $I_i$ ($i \in \{R, G, B\}$). Information about local geometry can be deduced by computing the eigenvalues and eigenvectors of $J$. The local vector geometry is computed from the structure tensor $J$. Its eigenvectors $v_{1,2}$ ($v_i \in \mathbb{R}^n$) define an oriented orthogonal basis and its eigenvalues $\lambda_{1,2}$ define the amount of structure variation. $v_1$ is the orientation with the highest fluctuations (orthogonal to the image contours), and $v_2$ gives the preferred local orientation.

eigenvector (having the smallest eigenvalue) indicates the isophote orientation. The use of structure tensor allows to retrieve a more coherent local geometry. The computation of the filling order as proposed by Criminisi et al. is then replaced by a term coming from PDE-based schemes, called Coherence Enhancing Diffusion. The use of structure tensor in an exemplar-based approach leads to a more robust algorithm that visually improves the quality of the inpainted areas.

Additionally, the simple template matching originally used in previous methods has been improved by using a K-nearest neighbor approach. The weights of the linear combination of the first K best candidates are adjusted by taking into account that all candidate patches are not equally reliable. Note that the number K is also locally adjusted in function of the local spatial complexity.

The 2D inpainting algorithm described above has been extended to deal with 3D content. In this work, the goal is to synthesize novel views directly from the original images. Image-based rendering (IBR) is commonly used to render a virtual view. It generates a nearby viewpoint image by projecting a point from the reference view to the virtual view using the depth data. However, when the viewpoint is shifted, occluded regions in the original viewpoint are dissoccluded. Handling these disocclusions (holes) is a difficult problem. We propose to use an extension of the 2D inpainting method to fill in these holes. For this goal, we have modified the computation of the structure tensor by adding the depth information. Equation (2) is simply modified as follow:

\[ J = \sum_{i=(R,G,B,Z)} \nabla I_i \nabla I_i^T \] (2)

where \( Z \) represents the depth map. As previously, the tensor is used to compute the filling order. A directional term is also included in order to favor a filling direction. Specifically, when the viewpoint is shifted from left to right in the horizontal direction, occluded regions in the left image appear in the right image around the right side of the object. Therefore, it is recommended to start the filling from the right to the left. This filling is performed by a modified template matching using texture information as well as depth data. Figure 3 illustrates the inpainting quality for different approaches.

![Figure 3. Virtual synthesized view. From left to right: original view projected into the new viewpoint; disocclusions filled by Criminisi’s approach, Daribo’s approach and the proposed method.](image)

6.1.3. Computational modelling of visual attention

Participants: Josselin Gauthier, Olivier Le Meur.

6.1.3.1. Eye-movement study:

In 2011, we have investigated whether two populations of visual fixation exist in 2D context. The question is simple: do all visual fixations have the same role in the free viewing of natural scenes? Recent studies suggest that there are at least two types of visual fixations: focal and ambient fixations. The former is believed to be used to inspect local areas accurately, whereas the latter is used to obtain the context of the scene.
From a collaboration with Technicolor (P. Guillotel and C. Chamaret) and LUTIN (T. Baccino), we found new evidence to support a focal-ambient dichotomy. Our results published in the journal i-Perception [14] indicate that the determining factor to classify the visual fixations is the saccade amplitude. We proposed an automatic system to cluster visual fixations in two groups using four types of natural scene images. From this automatic classification, the terms focal saliency map and ambient saliency map have been introduced. The dependence on the low-level visual features and the time course of these two kinds of visual fixations were examined. Our results demonstrate that there is an interplay between both fixation populations and that focal fixations are more dependent on low-level visual features than ambient fixations. These results might have a strong impact on both the computational modelling of visual attention and their performance assessment.

A second study related to eye-movement dealt with the role of the binocular disparity depth cue in the deployment of visual attention. To address this point, we compared eye tracking data recorded while observers viewed natural images in 2D and 3D conditions. The influence of disparity on the inter-observers congruency, saliency, center and depth bias was first examined. Results show that visual exploration in depth layer detection task is affected by the binocular disparity. In particular, participants tend to look first at closer areas just after the stimuli onset with the introduction of disparity, and then direct their gaze to more widespread locations. Our results has been submitted in the journal Cognitive Computation.

6.1.3.2. Model of visual attention:
Since 1998 with the publication of the influential work of Itti, Koch and Niebur [11], the computational modelling of the visual attention has known an increasing interest. The former models only used the low-level visual features for getting a saliency map. They perform well in a number of cases in predicting where an observer would look at. However, to improve the quality of the prediction, it seems unavoidable to use high-level information in order to account for visual deployment.

This work aims at designing a computational model mixing low-level and high-level features. Among the different factors influencing our gaze, we have focused our works on two cues: the dominant depth and the horizon line position. The dominant depth and the spatial position of the horizon line were inferred from the low-level visual features. A training database has been set up to perform a learning. Results indicate that the proposed model outperforms state-of-the-art models [37].

From behavioural studies on eye-movement in a 3D context, we have proposed a model of visual attention able to predict saliency of 3D pictures. The method developed aims at using the depth cue, the central bias and the low-level visual features. The predicted saliency is obtained by linearly combining these cues. The weights of the linear combination are learnt from a training database and are time-dependent. This study is under revision in the journal Cognitive Computation.

6.1.3.3. Predicting the inter-observer visual congruency:
This work aims at predicting the inter-observer visual congruency (IOVC), indicating the congruence or the variability among different subjects looking at the same image [35]. Predicting this congruence is of interest for image processing applications where the visual perception of a picture matters such as website design, advertisement, etc. We proposed a computational model of the IOVC. This new model is a mixture of low-level visual features extracted from the input picture. Model’s parameters are learned by using a large eye-tracking database. In this study, we also proposed a new scheme to compute the depth of field of a picture. Finally, once the training and the feature extraction have been carried out, a score ranging from 0 (minimal congruency) to 1 (maximal congruency) is computed. A value of 1 indicates that observers would focus on the same locations and suggests that the picture presents strong locations of interest. To illustrate the interest of the proposed model, we have used it to automatically rank personalized photograph. Figure 4 illustrates the proposed approach.

6.1.4. Visual cues analysis and modelling
Participants: Safa Cherigui, Christine Guillemot.
Figure 4. Top: pictures having high IOVC (first five) and pictures having low IOVC (last five). Bottom: saliency maps of pictures. Bright areas correspond to the most salient parts.

This work is carried out in collaboration with Technicolor (D. Thoreau, Ph. Guillotel, P. Perez) and aims at designing a compression algorithm based on the concept of epitomes. An epitome is a condensed representation of an image (or a video) signal containing the essence of the textural properties of this image. Different forms of epitomes have been proposed, such as a patch-based probability model learned either from still image patches or from space-time texture cubes taken from the input video. These probability models together with appropriate inference algorithms, are useful for content analysis inpainting or super-resolution. Another family of approaches makes use of computer vision techniques, like the KLT tracking algorithm, in order to recover self similarities within and across images. In parallel, another type of approach consists in extracting epitome-like signatures from images using sparse coding and dictionary learning.

The method developed aims at tracking self-similarities within an image using a block matching (BM) algorithm \[ 25 \]. The epitome is constructed from disjoint pieces of texture (“epitome charts”) taken from the original image and a transform map which contains translational parameters (see Fig. 5 - middle row). Those parameters keep track of the correspondences between each block of the input image and a block of the epitome. An Intra image compression scheme based on the epitome has been developed showing a rate saving of up to 12.

6.2. Representation and compression of large volumes of visual data

6.2.1. 3d representations for multi-view video sequences

Participants: Christine Guillemot, Vincent Jantet.

Multi-view plus depth video content represent very large volumes of input data which need to be compressed for storage and transmission to the rendering device. The huge amount of data contained in multi-view sequences indeed motivates the design of efficient representation and compression algorithms. In collaboration with INSA/ETRI (Luce Morin), we have studied layered depth image (LDI) and layered depth video (LDV) representations as a possible compact representation format of multi-view video plus depth data. LDI give compact representations of 3D objects, which can be efficiently used for photo-realistic image-based rendering (IBR) of different scene viewpoints, even with complex scene geometry. The LDI extends the 2D+Z representation, but instead of representing the scene with an array of depth pixels (pixel color with associated depth values), each position in the array may store several depth pixels, organised into layers.

Various approaches exist to construct LDI, which all organize layers by visibility. The first layer contains all pixels visible from a chosen reference viewpoint. The other layers contain pixels hidden by objects in previous layers. With classical construction solutions, each layer may contain pixels from the background and pixels from objects in a same neighbourhood, creating texture and depth discontinuities within the same layer. These
Figure 5. Epitome and reconstructed images: (top row) Original images (columns a and b); (middle) Epitomes; (bottom) Reconstructed images from epitome texture and transform map.
discontinuities are blurred during the compression process which in turn significantly reduces the rendering quality.

We have thus developed a novel object-based LDI representation which is more tolerant to compression artifacts, as well as being compatible with fast mesh-based rendering techniques [34]. This representation organises LDI pixels into two separate layers (foreground and background) to enhance depth continuity (see Fig. 6). The construction of this object-based LDI makes use of a foreground-background region-growth segmentation algorithm followed by inpainting of both colour and texture images to have a complete background layer (without the holes corresponding to disocclusion areas). The costly inpainting algorithm is thus processed once, during the LDI classification, and not during each view synthesis, which helps to speed up the rendering step.

![Figure 6. Object-based LDI: (top) Foreground and background layers; (bottom) Rendering results classical and object LDI.](image)

6.2.2. From sparse to spread representations

**Participant:** Jean Jacques Fuchs.

Sparse representations, where one seeks to represent a vector on a redundant basis using the smallest number of basis vectors, appear to have numerous applications. The other extreme, where one seeks a representation that uses all the basis vectors, might be of interest if one manages to spread the information nearly equally over all of them. Minimizing the $\ell_\infty$-norm of the vector of weights is one way to find such a representation. Properties of the solution and a dedicated fast algorithm have been developed. While the application of such models in robust data coding and in improving achievable data rates over amplitude constrained channels seems to be wishful thinking, its use in indexing techniques appears to be promising. In this context, one further replaces the optimal vector by its sign vector (potentially associated with a re-evaluated scalar weight) to get a binary vector that is not only cheap to store and (somehow) easy to search for but also allows for an explicit reconstruction unlike all other Hamming embedding functions used to map real vectors into binary vectors.
6.2.3. On-line dictionary learning methods for prediction

Participants: Christine Guillemot, Mehmet Turkan.

One crucial question to the problem of sparse approximation, and hence also of prediction based on sparse approximation, is the choice of the dictionary. Various advanced dictionary learning schemes have been proposed in the literature for the sparse signal approximation problem, so that the dictionary used is well suited to the data at hand. The popular dictionary learning algorithms include the K-SVD, the Method of Optimal Directions (MOD), Sparse Orthonormal Transforms (SOT), and (Generalized) Principle Component Analysis (PCA). However, the above learning methods are often used off-line since their computational complexity, which results from the number and the dimension of training samples, makes them inappropriate for online learning. In addition, these methods are adapted to the learning of basis to be used for approximating input data vectors, but not to the problem of predicting unknown samples from noisy observed samples in a causal neighborhood.

In 2011, we have developed a method for on-line training dictionaries adapted to the prediction problem [41]. Let A be the input dictionary, which is divided into two sub-dictionaries: A_c and A_t. The goal is to have a simple on-line dictionary learning method which is adapted to the prediction problem, i.e., which will learn both sub-dictionaries so that sparse vectors found by approximating the known samples (the template) using the first sub-dictionary (A_c) will also lead to a good approximation of the block to be predicted when used together with the second sub-dictionary (A_t). When dealing with the prediction problem, the sparse signal approximation is indeed first run with a set of masked basis functions (dictionary A_c), the masked samples corresponding to the location of the pixels to be predicted. The principle of the approach is to first search for the linear combination of basis functions which best approximates known sample values in a causal neighborhood, and keep the same linear combination of basis functions, but this time with the unmasked samples (dictionary A_t) to approximate the block to be predicted. The decoder similarly runs the algorithm with the masked basis functions and taking the previously decoded neighborhood as the known support. The use of masked basis functions converts the complete approximation problem into an overcomplete approximation problem. Because of its simplicity, of the limited number of training samples it requires, can be used for online learning of dictionaries, i.e. while doing the block-wise encoding of the image. The training samples are all possible previously coded/decoded texture patches (blocks of pixels) within a search window located in a causal neighborhood of the block to be predicted.

6.2.4. Neighbor embedding methods for image prediction and inpainting

Participants: Christine Guillemot, Mehmet Turkan.

The problem of texture prediction as well as image inpainting can be regarded as a problem of texture synthesis. Given observations, or known samples in a spatial neighborhood, the goal is to estimate unknown samples of the block to be predicted or of the patch to be filled in inpainting. We have developed texture prediction methods as well as a new inpainting algorithm based on neighbor embedding techniques which come from the area of data dimensionality reduction. The methods which we have more particularly considered are Locally Linear Embedding (LLE) and Non-negative Matrix Factorization (NMF). The first step in the developed methods consists in searching, within the known part of the image, for the K nearest (KNN) patches to the set of known samples in the neighborhood of the block to be predicted or of samples to be estimated in the context of inpainting. This first step can be seen as constructing a dictionary matrix by stacking in the matrix columns the vectorized K-NN texture patches. The non-negative dictionary matrix $A \in \mathbb{R}^{N \times M}$, is formed by K nearest neighbors to the vector formed by the known samples in the neighborhood of the samples to be predicted. These K nearest neighbors are texture patches of the same shape taken from the known part of the image. This dictionary can then be used for approximating the known samples by masking the rows of the matrix which correspond to the position of the unknown samples, solving a least squares problem under the constraint of sum-to-one of the weights in the case of LLE, or under the constraint of non-negativity of the weights for NMF. It is actually a variant of NMF since one of the components matrices is fixed (the one corresponding to the dictionary matrix) and only the matrix containing the weights of the linear approximation must then be found. The approaches are thus intended to explore the properties of the manifolds on which the
input texture patches are assumed to reside. The underlying assumption is that the corresponding uncomplete and complete patches have similar neighborhoods on some nonlinear manifolds. The new prediction methods give RD performances which are significantly better than the ones given by the H.264 Intra prediction modes, in particular for highly textured images [21], the highest gain being achieved with NMF.

A new examplar-based inpainting algorithm using neighbor embedding techniques has been developed. A new priority order has been proposed in order to inpaint first patches containing structures or contour information. The methods have also been shown to enhance the quality of inpainted images when compared to classical examplar-based solutions using simple template matching techniques to estimate the missing pixels, (see Fig. 7).

![Figure 7. Inpainting results: (left) mask of the image to be inpainted; (middle) Inpainting results with classical examplar-based inpainting; (right) Inpainting results with LLE (right).](image)

6.2.5. Lossless coding for medical images

**Participants:** Claude Labit, Jonathan Taquet.

Last year, we developed a hierarchical oriented prediction (HOP) algorithm, for resolution scalable lossless and near lossless compression of biomedical images. In 2011, the algorithm has been slightly improved with an iterative optimization of the predictors in order to get better results on less noisy/smooth images [39].

Recently, there have been a growing interest for the compression of an emerging imaging modality: the virtual microscopy (VM). It is used in anatopathology and may produce huge images of more than 1 Gigabytes. We have studied the efficiency for lossless and lossy compression of our previously developed algorithms HOP and OWD (optimized wavelet decomposition) and of two extensions of OWD: near-lossless and/or region of interest (ROI) coding. The lossless results, which are slightly better than JPEG-LS and JPEG-2000 standards with about 3:1 compression ratio, show that lossless compression is not suited to VM. By compressing only the information area (ROI) which represents about 20 percents of the size of test images, 9:1 ratio could be obtained, and combined with near-lossless approach, depending on the required quality, ratio can reach 17:1 with no visual losses to more than 30:1 with some visual losses (or approximately about 6:1 for ROI only data). We have concluded that it would probably be better to use lossy or efficient quality scalable compression. Because those images have specific contents (cellular tissue for example) we have also introduced and investigated new learning based methods. We have developed an optimization process for designing multiple KLT (Karhunen-Loeve Transform) in order to get orthonormal bases that are optimal for decorrelation and quality scalability. This learning approach has been applied as an a-posteriori transform of a wavelet decomposition in order to propose transforms with no blocking artefacts. A fully quality-scalable coding algorithm allows to obtain interesting PSNR improvements compared to the optimized coding process.
of JPEG-2000. Gain is around 0.5 dB for 16:1 compression of ROI only data, and more than 1 dB for 8:1 compression ratio.

6.3. Distributed processing and robust communication

6.3.1. Loss concealment based on video inpainting

Participants: Mounira Ebdelli, Christine Guillemot, Olivier Le Meur.

In 2011, we have started developing a loss concealment scheme based on a new video examplar-based inpainting algorithm. The developed video inpainting approach relies on new patch priority functions as well as on a motion confidence-aided neighbor embedding techniques. Neighbor embedding approaches aim at approximating input vectors (or data points) as a linear combination of their neighbors. The search of the weights of the linear combination (i.e. of the embedding) are formulated as constrained least squares problems. When using the locally linear embedding, the constraint is that the sum of the weights is equal to 1. We have also considered non-negative matrix factorization to solve the problem, in which case the constraint is that the weights and the other vector are non-negative. The motion confidence introduced in the neighbor embedding improves the robustness of the algorithm in the sense that it limits the error propagation effects which otherwise result from uncertainties on the motion information of the unknown pixels to be estimated. A new patch similarity measure which accounts for the correlation between motion information has been defined for the $K$-NN search inherent to neighbor embedding techniques. Evaluations of the algorithm in a context of video editing (object removal) are on-going. The next step will be to assess the performance of the approach in a context of loss concealment, that is to estimate unknown pixels after decoding when the corresponding transport packets have been lost on the transmission network.

6.3.2. Unequal Erasure Protection and Object Bundle Protection

Participant: Aline Roumy.

In 2011, we started a new collaboration in the framework of the Joint INRIA/Alcatel Lucent lab. In this work, carried out with V. Roca (Planete, INRIA), B. Sayadi and R. Imad (Alcatel Lucent), we proposed and analyzed a novel technique capable of providing both an unequal erasure protection service and an object bundle protection service.

**Unequal Erasure Protection:** When a data flow contains information of different priority levels, it is natural to try to offer an unequal protection where the high priority data benefits from a higher protection than the rest of data. In this work we focused on the “erasure channel”, for instance the Internet where the UDP/IP datagram integrity is guaranteed by the physical layer FCS (or CRC) and the UDP checksum. In this context UEP refers to an Unequal Erasure Protection (rather than Error) and the FEC code being used is one of the various Application-Layer Forward Erasure Correction (AL-FEC) codes that have been designed and standardized in the past years, like Reed-Solomon, one of the LDPC variants, or Raptor(Q) codes. Offering an unequal protection in this context can be achieved by one of the following three general approaches: by using dedicated UEP-aware FEC codes, by using a dedicated UEP-aware packetization scheme, or by using an UEP-aware signaling scheme. In this work we ignored the first approach as we wanted to reuse existing AL-FEC codes. Instead we focused on and compared the last two approaches and more precisely the well known Priority Encoding Transmission (PET) scheme that belongs to the UEP-aware packetization category and a Generalized Object Encoding (GOE) scheme, we proposed [53], that belongs to the UEP-aware signaling category. Through a careful modeling of both proposals [55], whose accuracy has been confirmed by simulations, we have demonstrated that the protection performance (i.e. erasure resiliency and average decoding delay) of both approaches are equivalent, not only asymptotically but also in finite length conditions. In fact the key differences between these approaches become apparent when applying them in practical systems. Such metrics as the simplicity of the solution, the number of packets processed, the maximum memory requirements, the number of FEC encoding and decodings, as well as the system of linear equations complexity (number of variables) are in favor of the GOE approach.
Object Bundle Protection: we considered the use of PET, more precisely an extension called Universal Object Delivery (UOD), and GOE in situations where one needs to send a bundle of small object (e.g. files). If both solutions can address this need, we showed that once again the GOE scheme is highly recommendable for practical realizations. This is mostly due to the lack of flexibility of the PET/UOD approach. For instance the limited size of a packet creates an upper bound to the number of objects that can be considered together (e.g. UOD limits this number to 255), the symbol size has a coarse granularity (e.g. UOD requires symbols to be multiple of 4 bytes when used with RaptorQ codes) which can create rounding problems with certain sets of objects (i.e. the actual packet size may be significantly shorter than the target, and/or the actual code rate significantly different than its target). GOE has no such constraints. In particular GOE offers the possibility to adjust the packet interleaving to the use-case and channel erasure features. One can easily trade robustness in front of long erasure bursts for very short decoding delays of high priority objects and low memory requirements, which can be a key asset in case of small, lightweight terminals or timely delivery services. This feature may be sufficiently important to justify by itself the use of a GOE FEC Scheme [55].

6.3.3. Distributed compressed sensing

Participants: Aline Roumy, Velotiaray Toto-Zarasoa.

This work has been performed in collaboration with E. Magli and G. Coluccia (Politecnico di Torino) in the framework of the FP7 IST NOE NEWCOM++ (Jan. 2008 - Apr. 2011). A new lossy compression scheme for distributed and sparse sources under a low complexity encoding constraint has been proposed in [26]. This problem naturally arises in wireless sensor networks. Indeed, nodes of a sensor network may acquire temperature readings over time. The temperature may vary slowly over time, and hence consecutive readings have similar values. However, they also have inter-sensor correlation, as the sensors may be in the same room, in which the temperature is rather uniform. The question hence arises of how to exploit intra- and inter-sensor correlations without communication between the sensors and with a low complexity acquisition process in order to save energy consumption at the sensor. Therefore, we consider continuous, correlated, distributed and sparse (in some domain) sources and perform lossyuniversal compression under a low encoding complexity constraint.

In order to meet the low complexity encoding constraint, the encoding stage is performed by a lossy distributed compressed sensing (CS). More precisely, the proposed architecture is based on the joint use of CS to capture memory of a signal, and DSC to take advantage of inter-sensor correlations. First, we showed that the resilience of CS to quantization error also holds in the distributed setup. Moreover, the optimal number of measurements can be chosen as the one guaranteeing (close-to-)perfect reconstruction. In addition, using joint decoding, dequantization and reconstruction techniques allows to boost performance even further. The joint use of CS and DSC allows to save 1.18 bit per source sample for the same PSNR quality w.r.t. the non-distributed CS scheme. Compared to the DSC scheme (without CS), we observe a gain increasing with the rate for the same PSNR quality. All these results makes the proposed scheme an attractive choice for environments such as sensor networks, in which the devices performing acquisition and processing are severely constrained in terms of energy and computations.

6.3.4. Super-resolution as a communication tool

Participants: Marco Bevilacqua, Christine Guillemot, Raul Martinez-Noriega, Aline Roumy.

In 2011, we started a new collaboration in the framework of the Joint INRIA/Alcatel Lucent lab. In this work, carried out with M.L. Alberi (Alcatel Lucent), we proposed a novel technique capable of producing a high-resolution (HR) image from a single low-resolution (LR) image. This method that belongs to the class of single-image super-resolution (SR), offers the promise of overcoming the inherent limitations of the video acquisition and transmission systems. More precisely, one can think of sending a low resolution video to adapt to the complexity constraint of the encoder and/or the bandwidth limitation of the network, while the decoder reconstructs a high-resolution video.
As a first step toward the more ambitious goal of compressing video through SR, we proposed a novel method for single-image super-resolution based on a neighbor embedding technique. Each low-resolution input patch is approximated by a linear combination of nearest neighbors taken from a dictionary. This dictionary stores low-resolution and corresponding high-resolution patches taken from natural images and is thus used to infer the HR details of the super-resolved image. The entire neighbor embedding procedure is carried out in a feature space. Features which are either the gradient values of the pixels or the mean-subtracted luminance values are extracted from the LR input patches, and from the LR and HR patches stored in the dictionary. The algorithm thus searches for the $K$ nearest neighbors of the feature vector of the LR input patch and then computes the weights for approximating the input feature vector. The so-obtained weights are finally used to compute a linear combination of the corresponding HR patches, which yields the super-resolved image. The use of a positive constraint for computing the weights of the linear approximation is shown to have a more stable behavior than the use of sum-to-one constraint and lead to significantly higher PSNR values for the super-resolved images.
TEXMEX Project-Team

6. New Results

6.1. Advanced algorithms of data analysis, description

6.1.1. Advanced description techniques

6.1.1.1. Image joint description and compression

**Participant:** Ewa Kijak.

*This is a joint work with the TEMICS project-team (J. Zepeda and C. Guillemot).*

In the context of ANR project ICOS-HD ended at December 2010, in collaboration with Christine Guillemot from TEMICS, we investigated sparse representations methods for local image description. We have developed methods for learning dictionaries to be used for sparse signal representations. These methods lead to dictionaries which have been called Iteration-Tuned Dictionaries (ITDs), Basic ITD (BITD), Tree-Structured ITD (TSITD) and Iteration-Tuned and Aligned Dictionaries (ITAD). All three proposed ITD schemes (BITD, TSITD and ITAD) have been shown to outperform the state-of-the-art learned dictionaries in terms of PSNR versus sparsity. The performance of these dictionaries has also been assessed for both compression and de-noising applications. ITAD in particular has been used to produce a new image codec that outperforms JPEG2000 for a fixed image class and leads in 2011 to two new publications [49], [20].

6.1.1.2. Bag-of-colors

**Participant:** Hervé Jégou.

*This is joint work with Christian Wengert (Kooaba) and Matthijs Douze (INRIA LEAR and SED project-teams).*

This work investigates [48] 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.

6.1.1.3. Aggregating local image descriptors into compact codes

**Participant:** Hervé Jégou.

*This is joint work with Matthijs Douze (INRIA LEAR and SED project-teams), Patrick Pérez (Technicolor), Florent Perronnin (Xérox Research Center Europe) and Cordelia Schmid (INRIA LEAR).*

This work [19] addresses the problem of large-scale image search and consolidates and extends results from a previous work [78]. Different ways of aggregating local image descriptors into a vector are compared, and the Fisher vector is shown to achieve better performance than the reference bag-of-visual words approach for any given vector dimension. We then 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. Searching a 100 million image dataset takes about 250 ms on one processor core.

6.1.2. Browsing multimedia databases

**Participant:** Laurent Amsaleg.

*This is a joint work with Björn Pór Jónsson and Grímur Tómasson from the School of Computer Science, Reykjavik University, Iceland.*
Since the introduction of personal computers, personal collections of digital media have been growing ever larger. It is therefore increasingly important to provide effective browsing tools for such collections. We have proposed a multi-dimensional model for media browsing, called ObjectCube, based on the multi-dimensional model commonly used in OLAP applications. We implemented a prototype of a media browser based on the ObjectCube model. We then ran evaluations of its performance using three different underlying data stores and photo collections of up to one million photos.

6.1.3. Advanced data analysis techniques

6.1.3.1. Factorial analysis and output display for text and textual streams mining

Participant: Annie Morin.

Textual data can be easily transformed in frequency tables and any method working on contingency tables can be used to process them. Besides, with the important amount of available textual data, we need to find convenient ways to process the data and to get invaluable information. It appears that the use of factorial correspondence analysis allows us to get most of the information included in the data. We start exploring temporal changes in textual data and mainly focus on the visualization of results: we try to detect the topics if they have not already been identified and to study the evolution of the previous vocabulary inside a topic through time. In fact, as with economical datasets, we try to find seasonal components and cycling components in the documents and to characterize these components.

6.1.3.2. Intensive use of SVM for text and image mining

Participants: François Poulet, Thanh Nghi Doan.

Support Vector Machines (SVM) and kernel methods are known to provide accurate models but the learning task usually needs a quadratic program, so this task for very large datasets requires a large memory capacity and a long time. We have developed new algorithms. The first versions of the algorithms were based on a CPU distributed software program, then we have used GP-GPU (General Purpose GPU) versions to significantly improve the algorithm speed (130 times faster than the CPU one, 2500 times faster than libSVM, SVMPerf or CB-SVM). We have extended the least squares SVM algorithm (LS-SVM) to adapt the algorithm to datasets having a very large number of dimensions and have applied boosting to LS-SVM for datasets having simultaneously a very large number of vectors and dimensions on standard computers. In image classification, the usual frameworks involve three steps: feature extraction, building codebook by feature quantization and training the classifier with a standard classification algorithm (eg. SVM). However, task complexity becomes very large when applying this approach on large scale datasets like the ImageNet dataset containing more than 14 million images and 21,000 classes. The complexity is both about the time needed to perform each task and the memory and disk usage (eg. 11TB are needed to store the SIFT descriptors computed on the full datasets). Efficient algorithms must be used into these three steps: - obviously, the descriptors computed for one image are independant of the other image ones, so they can be computed in a parallel way, - the quantization step usually uses a k-means algorithm, we have developed different versions of parallel k-means algorithms to use on GPU or a cluster of CPUs, - for the learning task, we have developed a parallel version of LibSVM. The first results on the ten largest classes of ImageNet dataset are promising [55], we have developed a fast and efficient framework for large scale image classification.

6.1.4. Security of media

6.1.4.1. Security of content based image retrieval

Participants: Thanh Toan Do, Ewa Kijak, Laurent Amsaleg, Teddy Furon.

Over the years, the level of maturity reached by content-based retrieval systems (CBRSs) has significantly increased. CBRSs have so far been used in very friendly settings where cultural enrichments are paramount. CBRSs are also used in quite different settings where the control, the surveillance and the filtering of multimedia information are central, such as for copyright enforcement systems. While an abundant literature assesses that today’s CBRSs are robust against general-purpose attacks, we address in this work the security of content-based retrieval systems. Because of our expertise, we focus on security of content-based image retrieval, where images are described by SIFT descriptors and indexed by NV-Tree. We proved in one
preliminary study that a real system fails to match a specifically attacked image and its quasi-copy, breaking its otherwise excellent copyright protection performances. After proposing specific attacks that aim to disturb the descriptor detection stage by both prevent some key-points of being detected and create new ones \[ 75 \], \[ 74 \], we pursue the work by considering attacks dedicated to the description computation stage.

6.1.4.2. Estimation of the false alarm probability in watermarking and fingerprinting

**Participant:** Teddy Furon.

A key issue in watermarking and fingerprinting applications is to satisfy the requirement on the probability of false detection or false accusation. Assume commercial contents are encrypted and watermarked and that future consumer electronics storage devices have a watermark detector. These devices refuse to record a watermarked content since it is copyrighted material. The probability of false alarm is the probability that the detector considers an original piece of content (which has not been watermarked) as protected. The movie that a user shot during his holidays could be rejected by his storage device. This absolutely non user-friendly behavior really scares consumer electronics manufacturers.

In fingerprinting, users’ identifiers are embedded in purchased contents. When this content is found in an illegal place (e.g. a P2P network), the copyright holders decode the hidden message, find an identifier, and thus they can trace the traitor, i.e. the customer who has illegally broadcast his copy. However, the task is not that simple because dishonest users might collude. For security reason, anti-collusion codes have to be employed. Yet, these solutions have a non-zero probability of error (defined as the probability of accusing an innocent). This probability should be, of course, extremely low, but it is also a very sensitive parameter: anti-collusion codes get longer (in terms of the number of bits to be hidden in content) as the probability of error decreases. Fingerprint designers have to strike a trade-off, which is hard to conceive when only rough estimation of the probability of error is known. The major issue for fingerprinting algorithms is the fact that embedding large sequences implies also assessing reliability on a huge amount of data, which may be practically unachievable without using rare event analysis.

In collaboration with the team-projects ASPI and ALEA, we developed a novel strategy for simulating rare events and an associated Monte Carlo estimation of tail probabilities. Our method uses a system of interacting particles and exploits a Feynman-Kac representation of that system to analyze their fluctuations. Our precise analysis of the variance of a standard multilevel splitting algorithm reveals an opportunity for improvement. This leads to a novel method that relies on adaptive levels and produces, in the limit of an idealized version of the algorithm, estimates with optimal variance. Some numerical results show performance close to the idealized version of our technique for these practical applications. This work has been published in the journal Statistics and computing [ 13 ]. Algorithms for estimating extreme probabilities and quantiles are implemented as a Matlab package.

6.1.4.3. New decoders for fingerprinting

**Participant:** Teddy Furon.

So far, the accusation process of a Tardos fingerprinting code is based on single decoders which compute a score per user. Users with the highest score or whose scores is above a threshold are then deemed guilty. In the past years, we have contributed to this approach bringing two improvements: the ‘learn and match’ strategy aims at estimating the collusion process and using the matched score function; a rare event analysis translates this score into a more meaningful probability of being guilty. A fast implementation computes the scores of one million of users within 0.2 second on a regular laptop. Therefore, contrary to common belief, although a single decoder is exhaustive with a linear complexity in \( O(n) \), it is not slow.

This fast implementation allows us to propose iterative decoders. A first idea is that conditioning by the identities of some colluders bring more discrimination power to the score function. The first iteration is thus a single decoder, users we are extremely confident to accuse are enrolled as side information. The next iteration computes new scores for the remaining users etc. A second idea is that information theory proves that a joint decoder computing scores for pairs, triplets, or in general \( t \)-tuples is more powerful than single decoders working with scores for single users. However, nobody did try them for large scale setups since the number of \( t \)-tuples is in \( O(n^t) \). We propose in a first iteration to use a single decoder, to prune out users who are
definitively innocents (because their scores are low) and keeping $O(\sqrt{n})$ individual suspects. The second iteration is a joint decoding working on pairs of users etc. Iteratively, we prune out enough users such that it is manageable to run a joint decoder on bigger $t$-tuples. This work has been done under a collaboration of TEMICS, and published in a series of conference papers [37], [38], [36]. A journal version has been submitted to IEEE Trans. on Information Forensics and Security. A Tardos code software suite (generation of code, collusion attacks, accusation algorithms) is available as a C package.

6.1.4.4. **Protocols for fingerprinting**

**Participant:** Teddy Furon.

A key assumption of the fingerprinting schemes developed so far is that the colluders may know their own codewords but they ignore the codeword of any other innocent user. Otherwise, the collusion can very easily forge a pirated content framing an innocent user because it contains a sequence close enough to his/her codeword. This puts a lot of pressure on the versioning mechanism which creates the personal copy of the content in accordance to a codeword. For instance, suppose that the versioning is done in the user’s setup box, the unique codeword being loaded into this device at the manufacture. If the code matrix ends up in the hands of an untrustworthy employee, then the whole fingerprinting system is pulled down. This is one argument of the motivation for designing cryptographic protocols for the construction, the versioning and the accusation. We have proposed a new asymmetric fingerprinting protocol dedicated to the state-of-the-art Tardos codes. We believe that this is the first such protocol, and that it is practically efficient. The construction of the fingerprints and their embedding within pieces of content is based on oblivious transfer and do not need a trusted third party. Note, however, that during the accusation stage, a trusted third party, like a Judge, is necessary like in any asymmetric fingerprinting scheme we are aware of. This work has been done in collaboration with the team-project TEMICS, Lab-STICC Telecom Bretagne and University College London, and presented at Information Hiding [22]. Ana Charpentier defended her PhD. thesis in October 2011 [72].

6.1.4.5. **Reconstructing an image from its local descriptors**

**Participant:** Hervé Jégou.

We show [47] that an image can be approximately reconstructed based on the output of a black-box local description software such as those classically used for image indexing. Our approach consists first in using an off-the-shelf image database to find patches which are visually similar to each region of interest of the unknown input image, according to associated local descriptors. These patches are then warped into input image domain according to interest region geometry and seamlessly stitched together. Final completion of still missing texture-free regions is obtained by smooth interpolation. As demonstrated in our experiments, visually meaningful reconstructions are obtained just based on image local descriptors like SIFT, provided the geometry of regions of interest is known. The reconstruction allows most often the clear interpretation of the semantic image content. As a result, this work raises critical issues of privacy and rights when local descriptors of photos or videos are given away for indexing and search purpose.

6.2. **Multi-dimensional indexing and clustering**

6.2.1. **Improved NV-tree**

**Participant:** Laurent Amsaleg.

This is a joint work with Björn Pór Jónsson from the School of Computer Science, Reykjavik University, Iceland and with Herwig Lejsek, Videntifier Technologies, Iceland.

We have further improved the NV-Tree (Nearest Vector Tree) indexing techniques. It addresses the specific, yet important, problem of efficiently and effectively finding the approximate $k$-nearest neighbors within a collection of a few billion high-dimensional data points. The NV-Tree is a very compact index, as only six bytes are kept in the index for each high-dimensional descriptor. It thus scales extremely well when indexing large collections of high-dimensional descriptors. The NV-Tree efficiently produces results of good quality, even at such a large scale that the indices cannot be kept entirely in main memory any more. We have demonstrated this with extensive experiments using a collection of 2.5 billion SIFT (Scale Invariant Feature Transform) descriptors. Additional experiments involving more than 30 billion SIFT descriptors show results are still of a good quality and that disks are handled as efficiently as they can be.
6.2.2. **Indexation of time series**  
**Participants:** Laurent Amsaleg, Romain Tavenard.

Dynamic Time Warping (DTW) is the most popular approach for evaluating the similarity of time series, but its computation is costly. Therefore, simple functions lower bounding DTW distances have been designed, accelerating searches by quickly pruning sequences that could not possibly be best matches. The tighter the bounds, the more they prune and the better the performance. Designing new functions that are even tighter is difficult because their computation is likely to become complex, canceling the benefits of their pruning. It is possible, however, to design simple functions with a higher pruning power by relaxing the no false dismissal assumption, resulting in approximate lower bound functions. We have discover how very popular approaches accelerating DTW such as LB_Keogh and LB_PAA can be made more efficient via approximations. The accuracy of approximations can be tuned, ranging from no false dismissal to potential losses when aggressively set for great response time savings. At very large scale, indexing time series is mandatory. These approximate lower bound functions can be used with iSAX. Furthermore, we have also observed that a $k$-means-based quantization step for iSAX gives significant performance gains.

6.2.3. **Improved image indexing with asymmetric Hamming embedding**  
**Participants:** Patrick Gros, Mihir Jain, Hervé Jégou.

We have proposed [28] an improved asymmetric Hamming Embedding scheme for large scale image search based on local descriptors. The comparison of two descriptors relies on an vector-to-binary code comparison, which limits the quantization error associated with the query compared with the original Hamming Embedding method. The approach is used in combination with an inverted file structure that offers high efficiency, comparable to that of a regular bag-of-features retrieval systems, and consistently improves the search quality over the symmetric version on the two datasets used for the evaluation.

6.2.4. **Compression techniques for nearest neighbor search**  
**Participants:** Laurent Amsaleg, Teddy Furon, Hervé Jégou, Romain Tavenard.

*Part of this work on this topic was done in cooperation with Matthijs Douze and Cordelia Schmid (INRIA/LÉAR).*

6.2.4.1. **Re-ranking with source coding**

An extension of our previous work on source coding techniques for high-dimensional indexing has been proposed [29]. The goal is to index a large set of vectors, as large as 1 billion vectors, with limited CPU and memory usage. Based on the product quantization-based indexing technique [18], we show that it is interesting to add an additional level of processing 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. Compared with the original work [18], the proposed re-ranking technique is shown to obtain better trade-off with respect to memory, efficiency and search quality.

6.2.4.2. **Anti-sparse coding for approximate nearest neighbor search**

Following recent works on Hamming Embedding techniques, we propose [67] a binarization method that aim at addressing the problem of nearest neighbor search for the Euclidean metric by mapping the original vectors into binary vectors ones, which are compact in memory, and for which the distance computation is more efficient.

Our method is based on the recent concept of anti-sparse coding, which exhibits here excellent performance for approximate nearest neighbor search. Unlike other binarization schemes, this framework allows, up to a scaling factor, the explicit reconstruction from the binary representation of the original vector. We also show that random projections which are used in Locality Sensitive Hashing algorithms, are significantly outperformed by regular frames for both synthetic and real data if the number of bits exceeds the vector dimensionality, i.e., when high precision is required.
6.2.5. Architecture-aware indexing techniques for solid state disks

Participants: Laurent Amsaleg, Gylfi Gudmundsson.

This is a joint work with Björn Pórl Jónsson from the School of Computer Science, Reykjavik University, Iceland.

The scale of multimedia data collections is expanding at a very fast rate. In order to cope with this growth, the high-dimensional indexing methods used for content-based multimedia retrieval must adapt gracefully to secondary storage. Recent progress in storage technology, however, means that algorithm designers must now cope with a spectrum of secondary storage solutions, ranging from traditional magnetic hard drives to state-of-the-art solid state disks. We have analyzed the impact of storage technology on a simple, prototypical high-dimensional indexing method for large scale query processing. We found that while the algorithm implementation deeply impacts the performance of the indexing method, the setup of the underlying storage technology is equally important.

6.3. New techniques for linguistic information acquisition and use

6.3.1. NLP for document description

6.3.1.1. Semantic annotation of multimedia documents based on textual data

Participants: Ali Reza Ebadat, Vincent Claveau, Pascale Sébillot, Ewa Kijak.

This work is done in the framework of the Quaero project (see below).

On this subject, TExMEX is implied in three tasks of the Quaero project.

The first task concerns the extraction of terminology from documents. The objective of this work is to study the development and the adaptation of methods to automate the acquisition and the structuring of terminologies. In this context, in 2011, we have undergone a new evaluation of terminology extraction systems. Here again, our system, relying on TermoStat (see previous reports) ranked first for the tracks in which we participated. We have also continued our work the use of morphology for biomedical terminologies. This approach relies on the decomposition of terms into morphemes and the translation of these morphemes into Japanese (kanji) subwords. The kanji characters thus offer a semantic way to access the semantics of the morpheme and allow us to detect semantic relations between them. We have tested this approach on more languages and have proved its relevance for information retrieval problems.

The second task aims at extracting semantic and ontological relations from documents. Indeed, detecting semantic and ontological relations in texts is a key to describe a domain and thus manipulate cleverly documents. In 2011, we developed a new relation extraction system based on k-nearest-neighbors and language modeling. It has been tested in the framework of the Quaero evaluation campaign and ranked first or second for all tracks. We have also developed a clustering technique for named entities. It relies on new representation schemes called bag-of-vectors (or bag-of-bags-of-features), which perform better than the classical bag-of-word approach.

The last task directly deals with the semantic annotation of multimedia documents based on textual data, for, very often, many textual or language-related data can be found in multimedia documents or come along such documents. For example, a TV-broadcast, contains speech that can transcribed, Electronic Program Guide and standard program guide information, closed captions, associated websites, etc. All these sources offers a way to exploit complementary information that can be used to semantically annotate multimedia document. During this year, we finished the development of a football multimedia corpus. It contains the video of several matches, the speech transcripts, associated textual data from specialized websites. All these media have been manually annotated in terms of events, named entities, specialized relations (fouls, replacements, etc) and other relevant information. This corpus will be distributed under LGPL-LR license.

6.3.1.2. Text recognition in videos

Participants: Khaoula Elagouni, Pascale Sébillot.
This work is done in the context of a joint Texmex/Orange Ph.D. thesis supported by a CIFRE grant with Orange Labs.

We aim at helping multimedia content understanding by obtaining benefit from textual clues embedded in digital video data. In 2011, we proposed an Optical Character Recognition-based method to recognize natural scene texts in images, avoiding the conventional character segmentation step. The text image is scanned with multi-scale windows and a robust recognition model is applied on each window, relying on a neural classification approach, to identify non valid characters and recognize valid ones. A graph model is used to represent spatial constraint between recognition results, and to determine the best sequence of characters. Some linguistic knowledge is also incorporated in the graph to remove errors due to recognition confusions. The method was evaluated on the ICDAR 2003 database of scene text images and outperforms state-of-the-art approaches. This work will be presented at DAS2012.

6.3.1.3. DEFT evaluation campaign participation

Participants: Vincent Claveau, Christian Raymond.

Christian Raymond and Vincent Claveau participated to DEFT (http://deft2011.limsi.fr/). Two tasks were proposed: the first one was called "the diachronic variation task" whose objective was to identify the writing year of some OCR newspapers from 1801 to 1944. The second one was a abstract/article pairing task. Their approaches based on boosting and k-nearest neighbors was ranked first on the difficult diachronic task.

6.3.2. Oral and textual information retrieval

6.3.2.1. Graded-inclusion-based Information retrieval systems

Participants: Vincent Claveau, Laurent Ughetto.

Our work on this topic is done in close collaboration with Olivier Pivert from the Pilgrim project-team of IRISA Lannion.

Databases (DB) querying mechanisms, and more particularly the division of relations was at the origin of the Boolean model for Information Retrieval Systems (IRSS). This model has rapidly shown its limitations and is no more used in Information retrieval (IR). Among the reasons, the Boolean approach do not allow to represent and use the relative importance of terms indexing the documents or representing the queries. However, this notion of importance can be captured by the division of fuzzy relations. This division, modeled by fuzzy implications, corresponds to graded inclusions. Theoretical work conducted by the Pilgrim project-team have shown the interest of this operator in IR.

Our first work was to investigate the use of graded inclusions to model the information retrieval process. In this framework, documents and queries are represented by fuzzy sets, which are paired with operations like fuzzy implications and T-norms. Through different experiments, we have shown that only some among the wide range of fuzzy operations are relevant for information retrieval. When appropriate settings are chosen, it is possible to mimic classical systems, thus yielding results rivaling those of state-of-the-art systems. These positive results have validated the proposed approach, while negative ones have given some insights on the properties needed by such a model.

More recently, the links between our fuzzy model and other classical IR models have been studied. It has been shown that our fuzzy implication-based model can be interpreted as several classical models: an Extended Boolean Model, a Logical Model, a Vector Space Model or a Language Model in IR.

6.3.2.2. Information retrieval in TV streams using automatic speech recognition

Participants: Guillaume Gravier, Patrick Gros, Julien Fayolle, Fabienne Moreau, Christian Raymond.
Automatic speech recognition outputs are by nature incomplete and uncertain, so much that lexical indexes of speech are not sufficient to overcome the errors due to out-of-vocabulary words and to most of the named entities, consisting in important semantic information from the discourse. Using if necessary a phonetic index is a solution to retrieve partially the mis-recognized words but at the price of a lower precision because the phonetic representation is also noisy. We proposed this year (still to be submitted) an indexation method which jointly model lexical and phonetic levels with finite-state transducers, offering the possibility to take a lexical path or a phonetic path between two synchronization nodes. The edges are weighted by a vector of features (edition scores, confidence measures, durations) that will be used in a supervised manner to estimate the reliability of the returned result at the search step. The experiments have shown the complementary of lexical-phonetic representations and their contribution for a task of spoken utterance retrieval using named entity queries.

6.4. New processing tools for audiovisual documents

6.4.1. TV stream structuring

6.4.1.1. Repetition detection-based TV structuring
**Participants:** Vincent Claveau, Guillaume Gravier, Patrick Gros, Emmanuelle Martienne, Abir Ncibi.

We work on the issue of structuring large TV streams. More precisely, we focus on the problem of labeling the segments of a stream according to their types (e.g., programs, commercial breaks, sponsoring, etc). Contrary to existing techniques, we wanted to take into account the sequential aspect of the data, and thus we used Conditional Random Fields (CRF), a classifier which has proved useful to handle sequential data in other domains like computational linguistics or computational biology. During this year, we proved the relevance of CRF in the framework of TV segments labeling. We conducted different experiments, either on manually or automatically segmented streams, with different label granularities, and demonstrated that this approach rivals existing ones. The use of this model for semi-supervised and unsupervised learning are under study.

6.4.2. Program structuring

6.4.2.1. Audiovisual models for event detection in videos
**Participants:** Guillaume Gravier, Patrick Gros, Cédric Penet.

*This work was performed in close collaboration with Technicolor as external partner.*

Following our work on the detection of audio concepts related to violence in movie soundtracks [58], we developed a system for the detection of violent scenes in movies, combining multimodal features. We investigated multimodal fusion strategies and temporal integration exploiting Bayesian networks as a joint distribution model. Several strategies for learning the structure of the Bayesian networks were compared, resulting in a complete system for violence detection. The system was evaluated on the Violent Scenes Detection task of the MediaEval 2011 international evaluation [42] that we co-organized with Technicolor and the University of Geneva [62]. A fair amount of time was dedicated this year to the organization of the evaluation campaign which includes defining the task and metrics, supervising the annotation, recruiting participants, analyzing the results and organizing the corresponding workshop session.

6.4.2.2. Unsupervised multimedia content mining
**Participants:** Guillaume Gravier, Anh Phuong Ta.

*This work on audio content discovery was partially carried out in collaboration with Armando Muscariello and Frédéric Bimbot from the Metiss project-team.*

As an alternative to supervised approaches for multimedia content analysis, where predefined concepts are searched for in the data, we investigate content discovery approaches where knowledge emerge from the data. Following this general philosophy, we pursued work on motif discovery in audio and video content.
Audio motif discovery is the task of finding out, without any prior knowledge, all pieces of signals that repeat, eventually allowing variability. In 2011, we extended our recent work on seeded discovery to near duplicate detection and spoken document retrieval from examples. First, we proposed algorithmic speed ups for the discovery of near duplicate motifs (low variability) in large (several days long) audio streams, exploiting subsampling strategies [39]. Second, we investigated the use of previously proposed efficient pattern matching techniques to deal with motif variability in speech data [40] in a different setting, that of spoken document retrieval from an audio example. We demonstrated the potential of model-free approaches for efficient spoken document retrieval on a variety of data sets, in particular in the framework of the Spoken Web Search task of the MediaEval 2011 international evaluation [41].

Video structure is often enforced through editing rules which result in a set of shots defining an event that repeats throughout the video with a high visual and audio similarity. Typical such shots are anchor persons and close-up on guests in talk-shows. We recently proposed an unsupervised multimodal approach to discover such events exploiting audio and visual consistency between two sets of independent nested clusters, one for each modality [21]. In 2011, we extended the approach in two directions. First, we improved the selection of consistent audio and visual clusters and the unsupervised selection of positive and negative examples exploiting redundancy between nested clusters. Second, we extended the method to discover several audio-visualy consistent events rather than a single one in our previous work, thus enabling the use of unsupervised mining as a pre-processing step for video structure analysis.

6.4.2.3. Topic segmentation with vectorization and morpho-mathematics

Participant: Vincent Claveau.

Our work on this topic is done in close collaboration with Sébastien Lefèvre from the SEASIDE project-team of IRISA Vannes.

Segmenting a program into topics is an important step for fine-grained structuring of TV streams. Based on our work on vectorization (see previous reports), we have developed a new segmentation technique using speech transcripts. Making an analogy with image segmentation, we have adapted the watershed transform to handle these textual data and more precisely the distances computed by vectorization between possible segments. This method has been tested on different TV collections (news, reports) as well as more usual text collection used for segmentation evaluation. In every cases, our technique has outperformed any state-of-the-art approaches.

6.4.3. Using speech to describe and structure video

Participants: Camille Guinaudeau, Guillaume Gravier, Ludivine Kuznik, Bogdan Ludusan, Pascale Sébillot.

Speech can be used to structure and organize large collections of spoken documents (videos, audio streams, etc) based on semantics. This is typically achieved by first transforming speech into text using automatic speech recognition (ASR), before applying natural language processing (NLP) techniques on the transcripts. Our research focuses firstly on the adaptation of NLP methods designed for regular texts to account for the specific aspects of automatic transcripts. In particular, we investigate a deeper integration between ASR and NLP, i.e., between the transcription phase and the semantic analysis phase.

In 2011, we mostly focused on robust transcription, hierarchical topic segmentation and collection structuring. On the one hand, we investigated the use of broad phonetic landmarks and syllable prominence to improve large vocabulary speech recognition by guiding the Viterbi search process. Several mechanisms to incorporate landmarks into the search space were studied. Significant improvements were observed on radio broadcast news data in the French language. On the other hand, we pursued our work on unsupervised topic adaptation, focusing on the automatic selection of out-of-vocabulary words combining phonetic and morpho-syntactic criteria.
Linear topic segmentation has been widely studied for textual data and recently adapted to spoken contents. However, most documents exhibit a hierarchy of topics which cannot be recovered using linear segmentation. We investigated hierarchical topic segmentation of TV programs exploiting the spoken material. Recursively applying linear segmentation methods is one solution but fails at the lowest levels of the hierarchy when small segments are targeted, in particular when transcription errors jeopardize lexical cohesion. We proposed new probabilistic measures of the lexical cohesion to emphasize the contribution of words that appears only locally, thus attenuating the impact of words which contributed to the segments at an upper level of the hierarchy [11].

Finally, we initiated work in collaboration with INA on structuring a large collection of news reports. The idea is to automatically create links and threads between reports in several months of broadcast news shows, based either on the documentary records of the shows and/or on the automatic transcripts. As preliminary step towards this goal, we investigated distances between documentary records in an information retrieval setting so as to construct a nearest neighbor graph. The next step consists in exploiting graph clustering methods.

Our research in speech for TV content structuring was illustrated through the Texmix demonstration (see Section 5.2) which exploits most of our achievements in the field, including transcription, topic segmentation and collection structuring.
6. New Results

6.1. Physical modelling and simulation

6.1.1. Modal analysis for haptic manipulation of deformable models

Participants: Zhaoguang Wang, Georges Dumont [contact].

Real-time interaction between designer and deformable mock-up in VR (Virtual Reality) environment is a natural and promising manner to evaluate designing feasibility. Using finite element method (FEM) for solving this issue leads to high fidelity simulation but to simulation rates that do not meet the requirements (1000Hz) of real time haptic applications. We have proposed a two-stage method based on linear modal analysis. In this method, different modal subspaces, related to use scenarios, are pre-computed offline. These data are then combined online with respect to a simulation division scheme to obtain real time deformations of the parts with respect to the modal response. Two main features are developed in the method. First, we apply an adapted meshing method during the pre-computation process. This method allows to automatically switch between different modal subspaces depending on the interaction region. Second, we divide the real time deformation computation into two separate modules by extracting sub-matrixes from the pre-computed modal matrixes. This separates the haptic simulation loop from the whole deformation computation and thus preserves the haptic response. This work was presented in WINVR 2011 conference [31] is accepted for publication [8] and was the subject of the PhD Thesis of Zhaoguang Wang, that was defended in june 2011 [3].

6.1.2. Real-time mechanical simulation of brittle fracture

Participants: Loëiz Glondu, Georges Dumont [contact], Maud Marchal [contact].

Simulating brittle fracture of stiff bodies is now commonplace in computer graphics. However, simulating the deformations undergone by the bodies in a realistic way remains computationally expensive. Thus, physically-based simulation of brittle fracture in real-time is still challenging for interactive applications. We are currently working on a new physically-based approach for simulating realistic brittle fracture in real-time. Our method is composed of two main original parts: (1) a fracture initiation model based on modal analysis and a new contact force model and (2) a fracture propagation model based on a novel physically-based algorithm (Figure 2). First results of this method have been published in [32].

Adding physical properties to objects within a virtual world can not generally be handled in real-time during a simulation. For that reason, it is still difficult nowadays to physically simulate fragments of fractured objects or parts of teared/cut objects. We have proposed a method for handling the real-time physical simulations of arbitrary objects that are represented by their surface mesh. Our method is based on a pre-computed shape database in which physical data are stored for a wide variety of objects. When a query object needs to be physically simulated in the virtual world, a similarity search is performed inside the database and the associated physical data are extracted. Our approach proposes to compare three different similarity search methods that fit with our real-time needs. Our results show that our approach has a great potential for the physical simulation of arbitrary objects in interactive applications. These results have been published in the Eurographics International Workshop on Virtual Reality Interaction and Physical Simulation (Vriphys) [21].

6.1.3. Collision detection in large scale environments with High Performance Computing

Participants: Quentin Avril, Valérie Gouranton [contact], Bruno Arnaldi.
Virtual reality environments are becoming increasingly large and complex and real-time interaction level is becoming difficult to stably insure. Indeed, because of their complexity, detailed geometry and specific physical properties, these large scale environments create a critical computational bottleneck on physical algorithms. Our work focused on the first step of the physical process: the collision detection. These algorithms can sometimes have a quadratic complexity. Solving and simplifying the collision detection problem is integral to alleviating this bottleneck. Hardware architectures have undergone extensive changes in the last few years that have opened new ways to relieve this computational bottleneck. Multiple processor cores offer the ability to execute algorithms in parallel on one single processor. At the same time, graphics cards have gone from being a simple graphical display device to a supercomputer. These supercomputers now enjoy attention from a specialized community dealing solely with physical simulation. To perform large scale simulations and remain generic on the runtime architecture, we proposed unified and adaptive mapping models between collision detection algorithms and the runtime architecture using multi-core and multi-GPU architectures. We have developed innovative and effective solutions to significantly reduce the computation time in large scale environments while ensuring the stability and reproducibility of results (cf. Figure 3).

We proposed a new pipeline of collision detection with a granularity of parallelism on multicore processors or multi-GPU platforms[11]. It enables simultaneous execution of different stages of the pipeline and a parallel internal to each of these steps. This was the subject of the PhD thesis of Quentin Avril[1].

6.1.4. Assessment of inverse dynamics method for muscle activity analysis

Participants: Georges Dumont [contact], Charles Pontonnier.

The use of virtual reality tools for ergonomics applications is a very important challenge. In order to improve the design of workstations, an estimation of the muscle forces involved in the work tasks has to be done. Several methods can lead to these muscle forces. In this study, we try to assess the level of confidence for results obtained with an inverse dynamics method from real captured work tasks. The chosen tasks are meat cutting tasks, well known to be highly correlated to musculoskeletal troubles appearance in the slaughter industry.

The experimental protocol consists in recording three main data during meat cutting tasks, and analyse their variation when some of the workstation design parameters are changing[26].

1. External (cutting)force data: for this purpose, a 3D instrumented knife has been designed in order
to record the force applied by the subject during the task;

2. Motion Capture data: for this purpose, we use a motion capture system with active markers (VisualEyez II, Phoenix Technologies, Canada);

3. EMG data: several muscle activities are recorded using electromyographic electrodes, in order to compare these activities to the ones obtained from the inverse dynamics method.

Then the motion is replayed in the AnyBody modeling system (AnyBody, Aalborg, Denmark) in order to obtain muscle forces generated during the motion. A trend comparison is then done [27], comparing recorded and computed muscle activations. Results show that most of the computed activations are qualitatively close from the recorded ones (similar shapes and peaks), but quantitative comparison leads to major differences between recorded and computed activations (the trend followed by the recorded activations in regard of a workstation design parameter, such as the table height, is not obtained with the computed activations). We currently explore those results to see if the fact that co-contraction of single joints muscles is badly estimated by classical inverse dynamics method can be a reason of this issue. We also work on the co-contraction simulation in order to improve the results [28].

This work has been done in collaboration with the Center for Sensory-motor Interaction (SMI, Aalborg University, Aalborg, Denmark), particularly Mark de Zee (Associate Professor) and Pascal Madeleine (Professor). Charles Pontonnier spent a 9 months post-doctoral fellowship at SMI from December 2010 to August 2011.

6.2. Multimodal immersive interaction

6.2.1. Brain-Computer Interaction based mental state

Participants: Anatole Lécuyer [contact], Bruno Arnaldi, Laurent George, Yann Renard.

In [20], presented at IEEE EMBS conference, we have explored the use of electrical biosignals measured on scalp and corresponding to mental relaxation and concentration tasks in order to control an object in a video game as illustrated in Figure 4. To evaluate the requirements of such a system in terms of sensors and signal processing we compared two designs. The first one used only one scalp electroencephalographic (EEG) electrode and the power in the alpha frequency band. The second one used sixteen scalp EEG electrodes and machine learning methods. The role of muscular activity was also evaluated using five electrodes positioned on the face and the neck.
Results show that the first design enabled 70% of the participants to successfully control the game, whereas 100% of the participants managed to do it with the second design based on machine learning. Subjective questionnaires confirm these results: users globally felt to have control in both designs, with an increased feeling of control in the second one. Offline analysis of face and neck muscle activity shows that this activity could also be used to distinguish between relaxation and concentration tasks. Results suggest that the combination of muscular and brain activity could improve performance of this kind of system. They also suggest that muscular activity has probably been recorded by EEG electrodes.

In [19], presented in the 5th International Brain-Computer Interface Conference, we introduce the concept of Brain-Computer Interface (BCI) inhibitor, which is meant to standby the BCI until the user is ready, in order to improve the overall performance and usability of the system. BCI inhibitor can be defined as a system that monitors user’s state and inhibits BCI interaction until specific requirements (e.g. brain activity pattern, user attention level) are met.

We conducted a pilot study to evaluate a hybrid BCI composed of a classic synchronous BCI system based on motor imagery and a BCI inhibitor (Figure 5 ). The BCI inhibitor initiates the control period of the BCI when requirements in terms of brain activity are reached (i.e. stability in the beta band).

Preliminary results with four participants suggest that BCI inhibitor system can improve BCI performance.

### 6.2.2. Navigating in virtual worlds using a Brain-Computer Interface

**Participants:** Anatole Lécuyer [contact], Jozef Legény.

When a person looks at a light flickering at a constant frequency, we can observe a corresponding electrical signal in their EEG. This phenomenon, located in the occipital area of the brain is called Steady-State Visual-Evoked Potential (SSVEP).
In [7] we introduce a novel paradigm for a controller using SSVEP. Compared to the state-of-the-art implementations which use static flickering targets, we have used animated and moving objects. In our example applications we have used animated butterflies flying in front of the user as shown in Figure 6. A study has revealed that, at the cost of decreased performance, this controller increases the personal feeling of presence.

These results show that integrating visual SSVEP stimulation into the environment is possible and that further study is necessary in order to improve the performance of the system.

6.2.3. Walking-in-place in virtual environments

Participants: Anatole Lécuyer [contact], Maud Marchal [contact], Léo Terziman, Bruno Arnaldi, Franck Multon.

The Walking-In-Place interaction technique was introduced to navigate infinitely in 3D virtual worlds by walking in place in the real world. The technique has been initially developed for users standing in immersive setups and was built upon sophisticated visual displays and tracking equipments. We have proposed to revisit the whole pipeline of the Walking-In-Place technique to match a larger set of configurations and apply it notably to the context of desktop Virtual Reality. With our novel “Shake-Your-Head” technique, the user has the possibility to sit down, and to use small screens and standard input devices for tracking. The locomotion simulation can compute various motions such as turning, jumping and crawling, using as sole input the head movements of the user (Figure 7).

In a second study [29] we analyzed and compared the trajectories made in a Virtual Environment with two different navigation techniques. The first is a standard joystick technique and the second is the Walking-In-Place (WIP) technique. We proposed a spatial and temporal analysis of the trajectories produced with both techniques during a virtual slalom task. We found that trajectories and users’ behaviors are very different across the two conditions. Our results notably showed that with the WIP technique the users turned more often and navigated more sequentially, i.e. waited to cross obstacles before changing their direction. However, the users were also able to modulate their speed more precisely with the WIP. These results could be used to optimize the design and future implementations of WIP techniques. Our analysis could also become the basis of a future framework to compare other navigation techniques.

6.2.4. Improved interactive stereoscopic rendering: SCVC

Participants: Jérôme Ardouin, Anatole Lécuyer [contact], Maud Marchal [contact], Eric Marchand.
Frame cancellation comes from the conflict between two depth cues: stereo disparity and occlusion with the screen border. When this conflict occurs, the user suffers from poor depth perception of the scene. It also leads to uncomfortable viewing and eyestrain due to problems in fusing left and right images.

In [10], presented at the IEEE 3DUI conference, we propose a novel method to avoid frame cancellation in real-time stereoscopic rendering. To solve the disparity/frame occlusion conflict, we propose rendering only the part of the viewing volume that is free of conflict by using clipping methods available in standard real-time 3D APIs. This volume is called the Stereo Compatible Volume (SCV) and the method is named Stereo Compatible Volume Clipping (SCVC).

Black Bands, a proven method initially designed for stereoscopic movies is also implemented to conduct an evaluation. Twenty two people were asked to answer open questions and to score criteria for SCVC, Black Bands and a Control method with no specific treatment. Results show that subjective preference and user’s depth perception near screen edge seem improved by SCVC, and that Black Bands did not achieve the performance we expected.

At a time when stereoscopic capable hardware is available from the mass consumer market, the disparity/frame occlusion conflict in stereoscopic rendering will become more noticeable. SCVC could be a solution to recommend. SCVC’s simplicity of implementation makes the method able to target a wide range of rendering software from VR application to game engine.

6.2.5. Six degrees-of-freedom haptic interaction

Participants: Anatole Lécuyer [contact], Maud Marchal [contact], Gabriel Cirio.

Haptic interaction with virtual objects is a major concern in the virtual reality field. There are many physically-based efficient models that enable the simulation of a specific type of media, e.g. fluid volumes, deformable and rigid bodies. However, combining these often heterogeneous algorithms in the same virtual world in order to simulate and interact with different types of media can be a complex task.

In [5], published at IEEE Transactions on visualization and Computer Graphics, we propose a novel approach that allows real-time 6 Degrees of Freedom (DoF) haptic interaction with fluids of variable viscosity. Our haptic rendering technique, based on a Smoothed-Particle Hydrodynamics (SPH) physical model, provides a realistic haptic feedback through physically-based forces. 6DoF haptic interaction with fluids is made possible thanks to a new coupling scheme and a unified particle model, allowing the use of arbitrary-shaped rigid bodies. Particularly, fluid containers can be created to hold fluid and hence transmit to the user force feedback coming from fluid stirring, pouring, shaking or scooping. We evaluate and illustrate the main features of our approach through different scenarios, highlighting the 6DoF haptic feedback and the use of containers.
The Virtual Crepe Factory [14] illustrates this approach for 6DoF haptic interaction with fluids. It showcases a 2-handed interactive haptic scenario: a recipe consisting in using different types of fluid in order to make a special pancake also known as "crepe". The scenario (Figure 8) guides the user through all the steps required to prepare a crepe: from the stirring and pouring of the dough to the spreading of different toppings, without forgetting the challenging flipping of the crepe. With the Virtual Crepe Factory, users can experience for the first time 6DoF haptic interactions with fluids of varying viscosity.

Figure 8. A complete use-case of our approach: a virtual crepe preparation simulator. The user manipulates a bowl (left hand, left haptic device) and a pan (right hand, right haptic device).

In [15], presented at the IEEE Virtual Reality Conference, we propose the first haptic rendering technique for the simulation and the interaction with multistate (Figure 9) media, namely fluids, deformable bodies and rigid bodies, in real-time and with 6DoF haptic feedback. The shared physical model (SPH) for all three types of media avoids the complexity of dealing with different algorithms and their coupling. We achieve high update rates while simulating a physically-based virtual world governed by fluid and elasticity theories, and show how to render interaction forces and torques through a 6DoF haptic device.

Figure 9. 6DoF haptic interaction in a medical scenario. Fluid blood pours from the deformable intestine when the user penetrates it with the rigid probe.

6.2.6. Joyman: a human-scale joystick for navigating in virtual worlds
Participants: Maud Marchal [contact], Anatole Lécuyer, Julien Pettré.
We have proposed a novel interface called Joyman (Figure 10), designed for immersive locomotion in virtual environments. Whereas many previous interfaces preserve or stimulate the users proprioception, the Joyman aims at preserving equilibrioception in order to improve the feeling of immersion during virtual locomotion tasks. The proposed interface is based on the metaphor of a human-scale joystick. The device has a simple mechanical design that allows a user to indicate his virtual navigation intentions by leaning accordingly. We have also proposed a control law inspired by the biomechanics of the human locomotion to transform the measured leaning angle into a walking direction and speed - i.e., a virtual velocity vector. A preliminary evaluation was conducted in order to evaluate the advantages and drawbacks of the proposed interface and to better draw the future expectations of such a device.

This principle of this new interface was published at international conference IEEE 3DUI [25] and a patent has been filed for the interface. A demonstration of this interface was proposed at ACM Siggraph Asia Emerging Technologies [33].

![Image](image.jpg)

*Figure 10. Prototype of the "Joyman"*

### 6.2.7. Interactions within 3D virtual universes

**Participants:** Thierry Duval [contact], Valérie Gouranton [contact], Bruno Arnaldi, Laurent Aguerreche, Cédric Fleury, Thi Thuong Huyen Nguyen.

Our work focuses upon new formalisms for 3D interactions in virtual environments, to define what an interactive object is, what an interaction tool is, and how these two kinds of objects can communicate together. We also propose virtual reality patterns to combine navigation with interaction in immersive virtual environments.

We have worked upon generic interaction tools for collaboration, based on multi-point interaction. In that context we have studied the efficiency of one instance of our Reconfigurable Tangible Device, the RTD-3, for collaborative manipulation of 3D objects compared to state of the art metaphors [9]. We have setup an experiment for collaborative distant co-manipulation (figure 1) of a clipping plane inside for remotely analyzing 3D scientific data issued from an earthquake simulation.

### 6.3. Collaborative work in CVE’s

#### 6.3.1. The immersive interactive virtual cabin (IIVC)

**Participants:** Thierry Duval [contact], Valérie Gouranton [contact], Alain Chauffaut, Bruno Arnaldi, Cédric Fleury.

We are still improving the architecture of our Immersive Interactive Virtual Cabin to improve the user’s immersion with all his real tools and so to make the design and the use of 3D interaction techniques easier, and to make possible to use them in various contexts, either for different kinds of applications, or with different kinds of physical input devices.
The IIVC is now fully implemented in our two VR platforms: OpenMASK 5.1 and Collaviz 7.1.3.

6.3.2. **Generic architecture for 3D interoperability**

**Participants:** Thierry Duval [contact], Valérie Gouranton, Cédric Fleury, Rozenn Bouville Berthelot, Bruno Arnaldi.

Our goal is to allow software developers to build 3D interactive and collaborative environments without bothering with the 3D graphics API they are using. This work is the achievement of the IIVC software architecture. We have proposed PAC-C3D (Figure 11), a new software architectural model for collaborative 3D applications, in order to provide a higher abstraction for designing 3D virtual objects, and in order to provide interoperability, making it possible to share a virtual universe between heterogeneous 3D viewers [17], [16].

![Figure 11. The PAC-C3D software architectural model makes interoperability possible between heterogeneous 3D viewers](image)

We also study how to offer interoperability between virtual objects that are loaded in the same virtual environment but that are described using different formats. This is why we have proposed a generic architecture for enabling interoperability between 3D formats (Figure 12), the Scene Graph Adapter [12]. Our SGA is now able to allow events coming from a 3D format to act upon data provided in another format, such as X3D events operating on Collada data [4].

6.4. **Immersia Virtual Reality room**

**Participants:** Georges Dumont [contact], Alain Chauffaut [contact], Ronan Gaugne [contact], Rémi Félix, Marwan Badawi, Bruno Arnaldi, Thierry Duval, Valérie Gouranton.

The team was the first in France to host a large-scale immersive virtual reality equipment known as Immersia. This platform, with full visual and sound immersion, is dedicated to real-time, multimodal (vision, sound, haptic, BCI) and immersive interaction. The Immersia platform is a key node of the european transnational VISIONAIR infrastructure and will be open in 2012 to the access of foreign research projects. It will accommodate experiments using interactive and collaborative virtual-reality applications that have multiple local or remote users. Our new wall has four faces: a front, two sides and a ground. Dimensions are 9.6 m wide, 2.9 m deep and 3.1 m height. The visual reproduction system combines eight Barco Galaxy NW12 projectors and three Barco Galaxy 7+ projectors. Visual images from Barco projectors are rendered on glass screens. They are adjusted for the user’s position, and this, together with their high resolution and homogeneous coloring, make them very realistic. The ART localization system, constituted of 16 ART-track2 cameras, enables real
Figure 12. Our architecture allows the loading of any 3D graphics format simultaneously in any available rendering engine. The scene graph adapter is an interface that adapts a scene graph (SG) of a given format into a renderer scene graph and which also allows the rendering part to request this scene graph.

objects to be located within the U-shape. Sound rendering is provided by a Yamaha processor, linked either to Genelec speakers with 10.2 format sound or Beyer Dynamic headsets with 5.1 virtual format sound, controlled by the user’s position.