# SCIENTIFIC REPORT 2010

## Microsoft Research-INRIA Joint Centre

CONTENTS

# TIMELINE

## 2005

**26 April 2005: MOU Signed.** François d'Aubert, the French Minister for Research, Gilles Kahn, Chairman and CEO of INRIA, and Steve Ballmer, CEO of Microsoft Corporation, sign a memorandum of understanding (MOU) and announce the creation of a joint research laboratory in France. "*The international stature of French research in computer science is confirmed by the decision by Microsoft to enter a partnership with INRIA and develop a joint laboratory in France. We value the opportunity for INRIA to work closely and openly with Microsoft*," says Gilles Kahn." *By working with internationally leading organisations such as INRIA, Microsoft is able to drive momentum in new kinds of science and computing, building on strong foundations*," says Andrew Herbert.

**26 October 2005: Framework Agreement Signed.** François Goulard, the French Minister for Research, Gilles Kahn, and Bill Gates, Chairman of Microsoft, sign a final framework agreement. A new laboratory will be created on the Plateau de Saclay, close to the INRIA Saclay-Ile de France campus, and near the Paris Sud University and Ecole Polytechnique.

## 2006

**Summer 2006: Track A Starts.** The Joint Centre is fully operational and moves in its new INRIA building at "Parc Orsay Université" on the Plateau de Saclay. A Management Committee is appointed with members at parity from Microsoft Research and INRIA. We start the first three projects in the "Software Security and Trustworthy Computing" track: Mathematical Components, led by Georges Gonthier (Microsoft Research Cambridge), Secure Distributed Computations and their Proofs, led by Cédric Fournet (Microsoft Research Cambridge), Tools and Methodologies for Formal Specifications and their Proofs led by Damien Doligez (INRIA) along with Leslie Lamport (Microsoft Research Silicon Valley).

**Fall 2007: Track B Starts.** We start three projects in the "Computational Sciences and Scientific Information Interaction" track: ReActivity, led by Wendy Mackay (INRIA) and Jean-Daniel Fekete (INRIA), Dynamic Dictionary of Mathematical Functions, led by Bruno Salvy (INRIA), and Adaptative Combinatorial Search for E-Sciences, led by Youssef Hamadi (Microsoft Research Cambridge) and March Schoenauer (INRIA).

**January 2009: Forum 2009.** We celebrate the second anniversary of the Joint Centre with a full day conference at École polytechnique: "Microsoft Research-INRIA Forum 2009". Three years of research at the Joint Centre is presented and our recent software is demonstrated by our researchers and students. "The success of this laboratory shows the path to follow in this strategic area" says Luc Rousseau, Head of the "Direction Générale des Entreprises" at the Ministry of Industry, in his closing address. We have 200 attendees.

**October 2009: Renewal of the Framework Agreement.** In the presence of Valérie Pecresse, Minister of Research and Higher education, Steve Ballmer, CEO of Microsoft Corp, Michel Cosnard, CEO of INRIA and Andrew Herbert, Managing Director, Microsoft Research Cambridge sign the Renewal of the Framework Agreement guaranteeing the existence of the Joint Centre until 2013.

2007 2008 2009 2010 2011

**Fall 2008: Seventh Project Starts.** Our seventh project starts, in the "Computational Sciences and Scientific Information Interaction" track: Image and Video Mining for Sciences and Humanities, led by Jean Ponce (Ecole Normale Supérieure). Since 2006, 3 PHD thesis have been completed and defended.

**December 2010: Project A-Brain.** Researchers at Neurospin (Saclay) and INRIA (Rennes) start a project investigating the benefits of using Cloud computing for studying correlations between brain images and genetic data.

**April 2011: Forum 2011.** We celebrate the fourth anniversary of the Joint Centre with a full day conference open to all at Campus Microsoft in Issy-les-Moulineaux in the presence of the following invited experts: Thierry Coquand, Bastian Leibe, Peter Paule, Catherine Plaisant, David Pointcheval, and Thomas Stuetzle.

# OVERVIEW

**How can one check that a program or a mathematical proof is correct, or that program does not leak confidential data over the network?**

**How can research in Computer Science provide new software tools to other sciences?**

**These are problems that we try to solve at the Microsoft Research-INRIA Joint Centre.**

Mathematical logic, invented to formalise meta-mathematics, has been used during half a century to specify properties of programs and to prove their correctness. Our ambition to mechanise mathematical proofs in pure mathematics is following the same trend and many theorems require lengthy proofs.
For instance, the Four-Colour theorem, which states that any planar graph can be coloured with four colours without assigning two adjacent nodes the same colour, is known to have a proof requiring several thousand distinct cases. Similarly, the Odd Order theorem, stating that any odd order finite group is solvable, is a fundamental theorem in the classification of finite groups and involves a proof of several hundred pages. These theorems need proofs expressed in a logic that a computer can check. We have done so for the Four-Colour theorem with a proof carried out in the Coq proof-assistant. We have already made substantial progress in doing the same for the Odd Order Theorem also in Coq.

Interestingly, the technology used in these long proofs comes from software technology where software components can be assembled to create large programs. In mathematical proofs, the choice of good mathematical components is crucial to facilitate their development. This is the objective of one project at the Joint Centre. In the long term, these mathematical components may become standard and popular. One can envision a revolution in the writing of textbooks for mathematics where formal proofs are both human readable and machine checkable. Furthermore, we can expect proof technology to impact the verification of properties of computer programs.

Security and concurrent programming are other areas where formal proofs are strongly needed. Protocols for security are now extremely complex, and concurrent programs are highly non-deterministic with fine grained interleaving during execution. The design of Secure Distributed Computations can be performed with the help of high-level primitives provided by programming languages, thus making it easier to program applications such as electric voting, electronic commerce, or auditing. But one needs to prove the safety of their translation to low-level protocols. Security properties are difficult to state in precise formal settings, often based on process calculi such as the ones studied in concurrency theory over the last thirty years. These properties require complex proofs since attacks by compromised participants are usually subtle. These proofs may need computer-assistance or use model checking methods. A project at the Joint Centre studies the design of Secure Distributed Computations and their Proofs. Another project is interested in correctness proofs of concurrent programs based on the Temporal Logic of Actions (TLA+), where we consider real-world examples, whose formal specifications bring substantial improvements over other methodologies, in order to improve the proof language and tools and to develop methods and design patterns for TLA+.

In the project for building a Dynamic Dictionary of Mathematical Functions (DDMF) at the Joint Centre, our aim is to automate the computation of numerous mathematical formulae needed in Calculus and Analysis. Engineers and researchers traditionally use encyclopedias or books which list tables of properties of useful functions, for instance Abramowitz and Stegun's Handbook of Mathematical Functions. With the help of Computer Algebra systems and recent results on special functions, considered as implicit solutions of linear differential equations, it is now possible to automatically build a dynamic encyclopedia providing up to 60DDMF encyclopedia, nicely interfaced to standard web navigators, will provide a new and attractive tool for engineers and researchers.

In the experimental sciences, there is a strong need for technical environments to help in capturing researcher activity and in visualising the history of results. Traditionally, this is achieved by a mix of laboratory notebooks and computers to log the various steps of experiments. In the Reactivity project at the Joint Centre, we explore how to capture and visualise user activity, and how to enable scientists and groups of scientists to reflect upon, interact with, and improve their research processes. Typical applications are for biologists or historians.

In the project named Adaptive Combinatorial Search for e-Sciences, the objective is to merge techniques of heuristic-based solvers (from constraint-based programming) and fine tuning of solver parameters (from evolutionary algorithms) in order to apply them to search and optimisation for complex scientific problems. The goal is also to increase the practicality of these techniques and provide a software tool that is easy to use by scientists who are not necessarily expert in computer science.

The project named Scientific Image and Video Data Mining is devoted to providing tools in Computer Vision which can be used by scientists in Environmental Sciences, Archaeology and Sociology. In this ambitious project, we plan to work on the detection of salient changes in multi-temporal satellite images with application to the assessment of natural damage, on the mining of historical collections of photographs and paintings with application to archaeology, and on the mining of TV broadcasts with application to sociology.

Finally, the new project A-Brain investigates the benefits of using the large computing power provided by Cloud computing (the Azure platform) for studying correlations between high-definition brain images and genetic data.

Thus, at the Joint Centre, INRIA and Microsoft Research have demonstrated a fruitful collaboration in creating an attractive centre for Computer Science research.

**Jean-Jacques LÉVY**
*Director of the Microsoft Research-INRIA Joint Centre*

# SCIENTIFIC REPORT 2010

## Microsoft Research-INRIA Joint Centre

## INTRODUCTION

The research programme at the Joint Centre is divided into two main tracks.

- **Track A**, "*Software Security and Trustworthy Computing*", comprises three projects: Mathematical Components, Tools and Methodologies for Formal Specifications and for Proofs, and Secure Distributed Computations and their Proofs.
- **Track B**, "*Computational Sciences and Scientific Information Interaction*", comprises four projects: Dynamic Dictionary of Mathematical Functions, ReActivity, Adaptive Combinatorial Search for E-Science, and Image and Video Mining for Science and Humanities.

Track A is focused on the application of mathematics to increasing the security and reliability of software and computing systems through formal specifications, tools for verification, and theorem proving or computer assisted proofs. These areas of research rely on a long history of cooperation between researchers at INRIA and Microsoft Research through conferences and exchanges of researchers and Phd/postdoc students. The tools used are mainly the ones of Mathematical Logic and the Theory of Programming Languages. Track A addresses new specification and proof techniques for distributed systems and Web services; it also treats formal proofs of mathematical theorems, such as the Four-Colour theorem or the Odd Order theorem, whose proof complexity is at the limits of current prover technology to handle.

Track B focuses on new software tools and applications for effective management, analysis of, and interaction with, increasingly highly complex scientific data. This relates to the areas of computer algebra, data visualisation, vision, computer human interfaces, constraint programming and evolutionary algorithms. This research track extends

**Jean-Jacques Lévy** graduated from Ecole Polytechnique in 1968 and received his PhD at the University of Paris 7 in 1978. He is senior researcher at INRIA, professor at Ecole Polytechnique since 1992 where he has been teaching Programming, and Director of the Microsoft Research-INRIA Joint Centre since 2006. Jean-Jacques Lévy worked in Operating Systems, the Semantics of Programming Languages, the syntax of the Lambda calculus, Interactive Graphics and CAD for VLSI and Concurrency theory.

the Microsoft European Science Initiative (ESI), focused on enabling and accelerating "new kinds" of science and computing: new fields emerging at the intersection of computer science and the natural sciences.

A new Track C based on applications of Cloud Computing started at end of 2010. It comprises a new project on Neuroimaging on top of the Azure platform. This project is part of the Microsoft Extreme Computing Group Cloud Computing Initiative in Europe.

In the area of S**oftware Security and Trustworthy Computing**, progress has been done in the three projects of Track A.

- In **Mathematical Components**, thanks to the tools developed in the first two years of the project, we designed in 2009 new techniques to combine the various mathematical structures that we had previously developed in a single coherent hierarchy. We then extended this hierarchy with a comprehensive development on linear algebra, which in turn allowed us to formalise linear group representation theory, the last outstanding prerequisite of the first part of the Feit-Thompson proof. These developments allowed us to start tackling the actual proof of the Odd Order Theorem at the start of 2010, and by the end of 2010 we had entirely completed the Local Analysis part of the proof, which is a little over half of it, as well as most of the character and finite field theory prerequisites for the remainder of the proof. This is major progress, and it now appears likely that we will complete the work in the next 12-18 months. During this work, there has been several releases of ssreflect, the Coq plug-in we designed to formalize mathematics. This work on finite groups is still done within an active cooperation with research projects Marelle at INRIA–Sophia-Antipolis, Typical at INRIA–Saclay.

- In **Secure Distributed Computations and their Proofs**, programming abstractions for security were explored, implemented, and verified (multiparty sessions versions 1 and 2, (refinement) type systems, secure information flows, secure logs). This lead to a better understanding of the relation between cryptography and formal proofs, with the formalizations of classic results of computational cryptography (in Coq), and the development of new verification tools to validate implementations of crypto-primitives or protocols. This work is done with an active collaboration of INRIA-Rocquencourt and INRIA-Sophia-Antipolis. There is also a strong interaction between work at the Joint Centre and the F7 project at MSRC. Publication activity has been high in majors ACM and IEEE security conferences. Several prototype packages were released, especially a compiler for secure sessions on top of both F#/.NET and Ocaml/OpenSSL, a verified reference implementation for a subset of the TLS standard, and the Coq library CertiCrypt formalizing game-based proof techniques for properties of computational cryptography.

- In **Tools and Methodologies for Formal Specifications and for Proofs**, the TLA+2 language has been designed, together with a proof system, interfaced with Isabelle-HOL, Coq and Zenon. The resulting system TLAPS was released on the web, with a graphical ToolBox giving access to an incremental proof manager, a model check and a parser. A safety proof of the Byzantine paxos consensus algorithm can now be performed within TLAPS. This work has been done by a close collaboration between MSR Silicon Valley (Lamport), INRIA Rocquencourt, INRIA Lorraine, and external assistance of Gonthier at the Joint Centre.

The general publication record of Track A is high. Two articles were accepted at the ACM conferences POPL'09 and POPL'10, plus 2 others also presented at POPL'10 and POPL'11 by a researcher of Track A (Zappa Nardelli) on related research (multicore memory models, type systems), which is not inside the projects of the Joint Centre. Leifer and Zappa Nardelli were also PC members at POPL'09 and POPL'11 and other researchers (Barthe, Blanchet, Fournet) were PC members or organizers of international conferences and workshops (IEEE CSF'09, CSF'10, CSF'11).

In the area of **New Software Tools and Applications for effective management, analysis of, and interaction** with, increasingly highly complex scientific data, progress was also done in the four projects of Track B, although most of them are very young projects.

- The **Dynamic Dictionary of Mathematical Functions** (DDMF) is now available on the Web (release 1.5). Its engine is based on a new representation of so-called "special functions" of mathematics using linear differential equations as a data-structure, instead of the more traditional tree based approach. The GUI works with standard internet navigators. The resulting dictionary is dynamic as opposed to static dictionaries based on manual and static precalculations of properties of mathematical functions. DDMF also relies on strong research in Computer Algebra, as shown by the high score of publications at ISSAC, the best conference of the domain, where Bostan is a PC member in 2010.

- In the **ReActivity** project, several prototypes have been completed. The publication record is high with seven papers presented at the ACM conference CHI'09. Fekete was a PC member at CHI'08; Mackay at CHI'09;

a workhop on the theme of Reactivity is organised by Reactivity researchers at CHI'09.

- The project **Adaptive Combinatorial Search for e-Sciences** benefits of the twofold expertise on Constraint Programming (Hamadi at MSRC) and Evolutionary Algorithms (Schoenauer at INRIA Saclay) coupled with know-how in Machine Learning (Sebag, CNRS, at INRIA Saclay). This combination lead to original results in both Constraint Programming (learning how to choose the best heuristic at any node of the search tree) and Evolutionary Computation (using Multi-Armed Bandit algorithms for operator selection). Furthermore, the ideas of CMA-ES, the state-of-the-art method in Evolutionary Continuous Optimization, originally developed by Hansen, have been extended to any other search algorithm, and to multi-objective optimization, with application to parameter identification for some biological model of the morphogenesis of drosophila.

- The project named **Image and Video Mining for Science and Humanities, Scientific Image and Video Data Mining**, addresses: (i) Mining historical collections of photographs and paintings with applications to archaeology and cultural heritage preservation; (ii) Mining and analysis of TV broadcasts with applications to sociology; (iii) the problem of detection, identification and tracking of dynamical geophysical events, with application to risk assessment and weather forecast. This project comprises experts in computer vision from ENS, INA, INRIA (Paris, Grenoble and Rennes) and MSR (Cambridge, Redmond and Bangalore).

- In the area of Applications of Cloud Computing, the project **A-Brain** was created in 2010 with researchers from Neurospin (Saclay) and INRIA (Rennes). The project investigates the benefits of using tools of Cloud computing (the Azure platform) for studying correlations between brain images and genetic data. The computing model will be based on BlobSeer developed at INRIA Rennes.

The three projects of Track A were prolonged for three new years (2010-2012). At end of 2010, the ReActivity project ended and a new project led by Nicholas Ayache and Antonio Criminisi on Automatic indexation of time/series 4D cardiac MR images was created, although not yet started. The three other projects of Track B were also renewed for three years (2011-2013).

Thus, the work achieved in the last two years is still impressive. As in the past years, the research has been published in renowned conferences; innovative softwares were designed and implemented in Track A and B. All projects but one have an INRIA and a MSR component. CNRS, ENS and Univ. of Paris-Sud also participate to several projects. Two post-docs at the Joint Centre and one MSR researcher (Bhargavan) won permanent positions at INRIA (Saclay); another post-doc obtained a professorship position at Purdue University. Nine PhD theses and one Habilitation have been defended. Two researchers (Ponce, Bhargavan) won ERC grants.

Several projects have very ambitious goals, like proving with computer assistance, the main theorem in the classification of finite groups, or providing a complete dynamic dictionary for useful mathematical functions. New research themes have been designed in Security, such as designing primitives in programming languages for secure sessions or mixing cryptology and information flow. The project on computer vision also bring new tools to social sciences and the preservation of the historical heritage. And successful usage of Cloud computing might boost analyses of neuro-images.

Finally, the Joint Centre has now 20 resident researchers on site, 16 PhD students and 11 post-docs. During the last 2 years, the Centre funded 24 person-years of PhD students, 24 person-years of Post-docs. The Centre hosted 22 PhD students, 28 post-docs, 4 long-term invited professors. It benefits from the collaboration of 35 researchers from INRIA, of 5 from other French academic institutions, of 14 researchers from Microsoft Research. (See tables in appendix.)

The following sections describe the current research in the seven projects. ■

**Track A** **This project started on summer 2006.**

# MATHEMATICAL COMPONENTS

## OVERVIEW

> **Our goal in this project is to demonstrate that formalized mathematical theories can, like modern software, be built out of components.**

Formalized mathematical theories can, like modern software, be built out of components. By components we mean modules that comprise both the static (objects and facts) and dynamic (proof and computation methods) contents of theories. We develop a general platform for mathematical components, based on the Coq "ssreflect" extension that was used to carry out the formalization of the Four Colour Theorem. Although we are using the formalization of a seminal result in Group Theory, the Odd Order Theorem, to drive our development, the components that we are developing — for Logic, Combinatorics, Set Theory, Algebra, Linear Algebra, Group Theory, Graph Theory, and Finite Field Theory — are widely usable.

**Georges Gonthier** graduated from Ecole Normale Supérieure in Paris, he received his PhD from the University of Paris 11, worked at AT&T Bell laboratories for 2 years and at INRIA for 13 years. He is presently a senior researcher at Microsoft Research Cambridge, which he joined in 2003. His work includes semantics of reactive languages, optimal reduction of functional programs, verification of concurrent memory management for Ocaml and IBM Java run-times, the join calculus model of concurrency (used in the design of Cw and Visual Basic), concurrent analysis of the Ariane 5 flight software, formal properties of security and fully computer-checked proof of the Four Colour Theorem.

## HIGHLIGHTS

In the 2009-10 period all the ground work completed in the first two years of the project came to fruition. In early 2009 we developed new techniques to combine the various mathematical structures we had previously developed in a single coherent hierarchy. We then proceeded to extend this hierarchy with a comprehensive development on linear algebra, which in turn allowed us to formalise linear group representation theory, the last outstanding prerequisite of the first part of the Feit-Thompson proof.

These developments allowed us to start tackling the actual proof of the Odd Order Theorem at the start of 2010, and by the end of 2010 we had entirely completed the Local Analysis part of the proof, which is a little over half of it, as well as most of the character and finite field theory prerequisites for the remainder of the proof. This is major progress, and it now appears likely that we will complete the work in the next 12-18 months.

## RESEARCH

The long-term scientific objective of the Mathematical Components project is to demonstrate that state-of-the art formal method tools can be successfully applied to research-level mathematics.

More precisely, our thesis is that the explanation for the limitations of the existing libraries of formal mathematics can be traced to those of the modularity tools used to compose elements of those libraries. Our plan was to combine modern software engineering concepts such as components with the advanced type system and type inference procedures of the Coq proof system to overcome those limitations.

We decided to drive and test our work with a real large-scale example: the proof of the Feit-Thompson, or Odd Order Theorem. This landmark result asserts that all finite groups of odd order are solvable, and sparked the effort towards the complete Classification of Finite Simple Groups. It is also the first example of a large — 255 pages — proof in group theory that is hard to survey: it reportedly takes a year for a trained specialist to understand the proof.

During the first two years we started to develop the basic and intermediate component libraries that would be required for the Feit Thompson proof. This work, which was described in the previous edition of this report, was very successful, and identified a specific and unique feature of the Coq type inference system — Canonical Structures — as the basis of the solution to the mathematical components problem (see par. *Librairy Organisation*).

In the last two years we have developed this approach, exploiting Canonical Structures to create a whole new range of library organisation techniques, starting to create a comprehensive set of libraries for general algebra, making major progress towards the completion of the Feit-Thompson formalisation itself in the last year (it is now 60% complete), and finally starting to investigate new applications for our libraries, partly as part of a new EU-funded project, FORMATH.

## TEAM

| **Team leader** | **GONTHIER** | **Georges** | ***Microsoft Research Cambridge*** |
|---|---|---|---|
| Researcher | BARRAS | Bruno | *INRIA Saclay-Île-de-France* |
| Researcher | BERTOT | Yves | *INRIA Sophia Antipolis-Méditerranée* |
| Researcher | MAHBOUBI | Assia | *INRIA Saclay-Île-de-France* |
| Researcher | RIDEAU | Laurence | *INRIA Sophia Antipolis-Méditerranée* |
| Researcher | THERY | Laurent | *INRIA Sophia Antipolis-Méditerranée* |
| Researcher | WERNER | Benjamin | *INRIA Saclay-Île-de-France* |
| Post doc | COHEN | Cyril | *Ecole Polytechnique* |
| Post doc | LEROUX | Stéphane | *MSR-INRIA Joint Centre* |
| Post doc | MAHBOUBI | Assia | *MSR-INRIA Joint Centre* |
| Post doc | MELQUIOND | Guillaume | *MSR-INRIA Joint Centre* |
| Post doc | TASSI | Enrico | *University of Bologna* |
| PHD student | CANO | Guillaume | *Université Nice Sophia Antipolis* |
| PHD student | COHEN | Cyril | *Ecole Polytechnique* |
| PHD student | GARILLOT | Francois | *Ecole Normale Supérieure de Paris* |
| PHD student | O'CONNOR | Russell | *MSR-INRIA Joint Centre* |
| PHD student | OULD BIHA | Sidi | *Université Nice Sophia Antipolis* |
| PHD student | PASCA | Iona | *Université Nice Sophia Antipolis* |
| PHD student | SPIVAK | Arnaud | *Ecole Polytechnique* |
| PHD student | TASSI | Enrico | *University of Bologna* |
| PHD student | ZUMKELLER | Roland | *Ecole Polytechnique* |

*The permanent researchers of the Mathematical Components project are Georges Gonthier (Principal Researcher, Microsoft Research), Benjamin Werner and Yves Bertot (Senior Researcher, Inria), and Assia Mahboubi, Laurence Rideau, and Laurent Théry (Researcher, Inria). We have had Jeremy Avigad as a visiting professor (on sabbatical from CMU) for the 2009-10 academic year, and Enrico Tassi has been a postdoc with us since November 2009. In addition, Russel O'Connor started a joint postdoc with us and McMaster University (Canada); he was with us from April to July in 2010, and will be returning for the same period in 2011 and 2012. We had three PhD students with the project at the start of 2009, Sidi Ould Biha, Ioana Pasca, and François Garillot; the first two defended their thesis in 2010. Three new PhD student joined us in 2009-10: Cyril Cohen in October 2009, and Maxime Denès and Guillaume Cano in 2010. Finally, Maxime Denès, Benjamin Lesage and Salil Josh were interns with us in 2009, as was Stephania Dumbraya in 2010.*

**Careers:** *Our two PhD students that defended their thesis in 2009, Sidi Ould Biha and Ioana Pasca, are respectively post docs at Tsingua University in Beijing and at LIP (Laboratoire de l'Informatique du Parallélisme) in Lyon.*

**Visitors:** *Andrea Asperti (University of Bologna, Italy) in 2008. Jérémy Avigad (CMU University, USA) in 2009.*

## LIBRARY ORGANISATION

Canonical Structures are a very simple mechanism: they are simply hints that let the type inference algorithm of Coq extrapolate the value of certain records from a single one of their fields. In combination with other features of the Coq system — dependent types, higher-kinded records, convertibility, coercions and user notations with inherited attributes — this apparently crude mechanism turns out to be extremely expressive, basically because it allows a library designer to reprogram the behaviour of value matching, unification and type inference to suit his specific needs.

- **Structure hierarchies:** In mathematics, structures are a means of associating certain operations and properties to sets; mathematical structures are organised by a complex web of inheritance relations; for example fields are a refinement of rings, while algebras are both vector spaces and rings.

  Type theoretic structures (records with type fields) serve a similar purpose in formalised mathematics; however, inheritance between such structures had been limited to structural inclusion, which is limited to single inheritance and interacts poorly with the type inference heuristics of Coq. We have discovered that it is possible to use canonical structures to program explicitly inheritance relations amongst structures. This has allowed us to adopt a new organisation for structures, called *packed classes*, that is sufficiently expressive and efficient to represent the large collection of mathematical structures required by the Odd Order theorem proof.
- **Quotation inference:** Unlike their programming language counterpart, type classes, canonical structures can be keyed on values as well as types. We have exploited this to reuse the packed class organisation to represent the hierarchy of morphisms between structures, and to capture key concepts of group theory (subgroups and group action and representation). We have also shown that the same feature can be used to perform term quotation, and exploited this to represent the concept of direct subspace sum in linear algebra, and of characteristic subgroup in group theory.
- **Quantifier elimination:** Another potential application of quotation is that it makes it possible to incorporate decision procedures into type inference. Assia Mahboubi supervised Salil Joshi's internship on the development of a reflexive tactic for first order linear integer arithmetic for the Coq system. This work has lead to a full implementation and formal proof of correctness of Cooper's algorithm.

  Also, Assia Mahboubi and Cyril Cohen formalised the general Tarski quantifier elimination procedure for closed fields, using the continuation-passing style programming technique to streamline the description and proof of the procedure.

### TOOLS AND SOFTWARE

Version 1.2 of the ssreflect was released in August 2009 and featured a major expansion of the core ssreflect 1.2 theory to cover primes, summations (and more generally "big operators"), finite sets, functions, and groups, and general algebra with matrices and polynomials, including a complete algebraic hierarchy. It also features comprehensive internal documentation for each module in the library.

Version 1.3 of ssreflect, released in March 2011, includes numerous improvements to the ssreflect proof language and its implementation, and extends further the theory library to cover general linear algebra, and advanced group theory including a comprehensive treatment of non-linear and linear group representation theory, all with extended documentation. We have also published a comprehensive ssrflect tutorial.

## CONSTRUCTIVE ALGEBRA

Because the Odd Order theorem is a result in finite group theory, we expect its proof to be constructive; this should be expressible in Coq, since the core logic of Coq is also constructive, and we have set ourselves this as an additional goal. To do so we had to devise replacements for some of the non-constructive general algebra results used in the classic proof. This has led to some new and intriguing math.

- **Linear algebra:** The obvious decision procedure for linear algebra is Gaussian elimination, but upon formalising this we realised that this procedure covers all basic linear algebra — for instance, it implements a proof of the incomplete basis theorem. We developed on this basis an "untyped" model of linear algebra, where matrices represent everything from vectors and homomorphisms to bases and subspaces. Matrix multiplication and its algebraic laws are consistent with these multiple interpretations, and this leads to a considerable simplification of the formalisation. The usual zoology of linear algebra objects can alternatively be recreated by adding simple wrapper types.
- **Linear group representations:** We exploited the matrix model to develop a fairly complete formalisation of linear group representations, including results such as Maschke's theorem, the Schur lemmas, the Jacobson density theorem, Clifford's theorem, and the Wedderburn structure theorem for semisimple rings. We used a variant of the Gödel encoding to formalise the inherently non-constructive results that were required, such as the

existence of a socle or of closure fields.

- **Field theory:** Russel O'Connor developed and formalised a new, constructive proof of the primitive element theorem, which makes it possible to combine the subrings generated by roots of different polynomials. This provides the basis for the formalisation of both algebraic numbers and Galois theory.

## THE FEIT THOMPSON PROOF

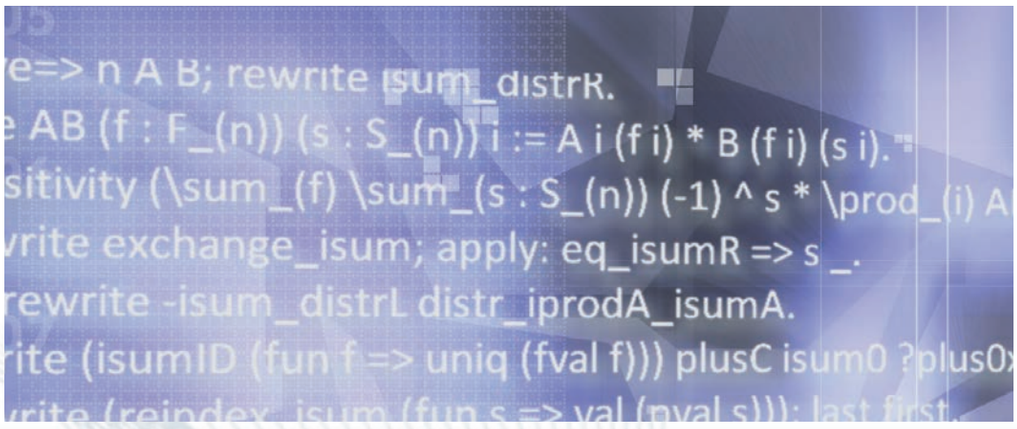
The proof splits in two parts: Local Analysis, which characterises the structure of subgroups of a hypothetical minimal counter-example to the theorem, and Character Theory, which exploits Local Analysis to derive global norm inequalities on the characters of the group, that lead to the final contradiction. In 2010 we completed the formalisation of the first part; this is about 60% of the proof.

- **The Uniqueness theorem:** This is the "deep" result that underpins Local Analysis; it asserts that any proper subgroup of a minimal counter-example that has "dimension" (more properly, *rank*) greater than 2 is included in a *unique* maximal subgroup. While this appears to be a purely combinatorial result, its proof relies heavily on linear representations via the p-stability of groups of odd order. This special case of the Hall-Higman theorem states that in a faithful p-modular representation no p-element has a quadratic minimal polynomial. We developed a new, synthetic proof of this result.
- **Groups of small rank:** The Uniqueness theorem implies that the intersection of two different maximal subgroups has rank at most 2. The second ingredient of Local Analysis is the structure theorem for such groups, which was formalised by Jeremy Avigad and Enrico Tassi. It asserts that these groups factor like their order, with a single factor for each prime divisor.
- **Local analysis:** By combining the above it can now be shown that each maximal group factors into a *kernel* comprising all p-subgroups of large rank, and a complement of small rank. Further analysis of the complement shows that it acts almost regularly on the kernel — technically, that all maximal subgroups are of *Frobenius type*, with at most two exceptions. This is the most complex part of the proof, and it really tested the ability of our proof language and formal library to capture the high-level, often elliptic arguments found in a graduate textbook.

## APPLICATIONS

In addition to our main effort on the Feit-Thompson proof, we have worked on other applications of our library of mathematical components.

- **The FORMATH consortium:** In 2009 we joined with the Universities of Chalmers (Sweden), Nijmegen (the Netherlands), and La Rioja (Spain) in a project to extend the ssreflect mathematical libraries to cover advanced algebra and linear algebra, some analysis, and algebraic topology. The Formath (Formalization of Mathematics) project was accepted in the European Community ICT program, and started in March 2010.
- **Numerical analysis:** In her PhD, Ioana Pasca exploited the basic linear algebra library to formalize convergence conditions for the multivariate Newton method. She proved the Kantorovitch theorem, then refined the formalization to take floating-point rounding errors into account, which had never been done before, and makes the result more relevant to the design and verification of control software in robotics. She then extended this analysis to efficient regularity conditions for matrices with interval coefficients, due to Rex and Rohn.
- **Gene networks:** Genetic networks are an abstraction of the behaviour of living organisms which provide a model of the conditions in which genes are expressed. We used the combinatorial part of our library to formalise some relations of this model to finite-state automata. This work gives new insights on the formal modelling of genetic networks. ■



# PUBLICATIONS & TALKS

## JOURNAL PAPERS AND BOOK CHAPTERS


[1] YVES BERTOT, FRÉDÉRIQUE GUILHOT, AND ASSIA MAHBOUBI. A formal study of Bernstein coefficients and polynomials. Mathematical Structures in Computer Science, 2011.

[2] GEORGES GONTHIER AND ASSIA MAHBOUBI. An introduction to small scale reflection in Coq. Journal of Formalized Reasoning, 3:95–152, 2010.


## CONFERENCE AND WORKSHOP PAPERS


[3] ANDREA ASPERTI AND ENRICO TASSI. Smart matching. In AISC/MKM/Calculemus, pages 263–277, 2010.

[4] CYRIL COHEN. Types quotients en coq. In Hermann, editor, JFLA 2010, Vieux-Port La Ciotat, France, January 2010. INRIA.

[5] CYRIL COHEN AND ASSIA MAHBOUBI. A formal quantifier elimination for algebraically closed fields. In AISC/MKM/Calculemus, volume 6167 of Computer Science, pages 189–203, Paris France, 06 2010. Springer.

[6] MAXIME DÉNÈS, BENJAMIN LESAGE, YVES BERTOT, AND ADRIEN RICHARD. Formal proof of theorems on genetic regulatory networks. In SYNASC'09, Timisoara, Romania, 2009. IEEE.

[7] JEAN-FRANÇOIS DUFOURD AND YVES BERTOT. Formal study of plane delaunay triangulation. In ITP, pages 211–226, 2010.

[8] FRANÇOIS GARILLOT, GEORGES GONTHIER, ASSIA MAHBOUBI, AND LAURENCE RIDEAU. Packaging mathematical structures. In Theorem Proving in Higher-Order Logics, volume 5674 of LNCS, pages 327–342, 2009.

[9] GEORGES GONTHIER. Software Engineering for Mathematics. In Calculemus/MKM, page 27, 2009.

[10] GEORGES GONTHIER. Type design patterns for computer mathematics. In TLDI, pages 1–2, 2011.

[11] NICOLAS JULIEN AND IOANA PASCA. Formal verification of exact computations using Newton's method. In TPHOLs, pages 408–423, 2009.

[12] CHANTAL KELLER AND BENJAMIN WERNER. Importing HOL Light into Coq. In ITP, pages 307–322, 2010.

[13] ASSIA MAHBOUBI. Présentation de ssreflect (cours). In Actes de la conférence JFLA 2009, pages 22–23. INRIA, 2009.

[14] SIDI OULD BIHA. Finite groups representation theory with Coq. In 8th International Conference on Mathematical Knowledge Management, Grand Bend, Ontario, Canada, 2009.

[15] IOANA PASCA. Formally verified conditions for regularity of interval matrices. In AISC/MKM/Calculemus, pages 219–233, 2010.


## THESES


[16] SIDI OULD BIHA. Composants mathématiques pour la théorie des groupes. PhD thesis, Université de Nice Sophia-Antipolis, February 2010.

[17] IOANA PASCA. Vérification formelle pour les méthodes numériques. PhD thesis, Université de Nice Sophia-Antipolis, November 2010.


## TECH REPORTS


[18] GEORGES GONTHIER AND STÉPHANE LE ROUX. An Ssreflect Tutorial. Technical Report RT-0367, INRIA, 2009.

[19] GEORGES GONTHIER, ASSIA MAHBOUBI, AND ENRICO TASSI. A Small Scale Reflection Extension for the Coq system. Research Report RR-6455, INRIA, 2008.

[20] IOANA PASCA. Formal Proofs for Theoretical Properties of Newton's Method. Research Report RR-7228, INRIA, March 2010.


## TALKS


[21] GEORGES GONTHIER. Software Engineering for Mathematics. In Calculemus/MKM, page 27, 2009.

[22] GEORGES GONTHIER. Mechanizing the odd order theorem: Local analysis. In AMS Joint Mathematics Meeting #1067, 2011.

[23] GEORGES GONTHIER. Type design patterns for computer mathematics. In TLDI, pages 1–2, 2011.

[24] JEREMY AVIGAD. Type inference in finite group theory. In AMS Joint Mathematics Meeting #1067, 2011.

[25] ASSIA MAHBOUBI. Présentation de ssreflect (cours). In Actes de la conférence JFLA 2009, pages 22–23. INRIA, 2009.

Track A **This project started in September 2006.**

# SECURE DISTRIBUTED COMPUTATIONS AND THEIR PROOFS

## OVERVIEW

**"We design and prototype formal tools for distributed programming with simple, effective security guarantees"**

We develop and apply formal tools for programming distributed computation with effective security guarantees. Our goal is to enable programmers to express and prove high-level security properties with a reasonable amount of effort—sometimes automatically, sometimes with mechanical assistance—as part of the development process. These properties should hold in a hostile environment, with realistic (partial) trust assumptions on the principals and the machines involved in the computation, and with realistic cryptographic assumptions on the underlying implementation mechanisms.

**Cedric Fournet** graduated from École Polytechnique and École Nationale des Ponts et Chaussées, did a PhD on distributed programming at INRIA, then joined Microsoft Research, Cambridge in 1998. He is interested in security, distributed systems, and concurrent programming. His recent research subjects include verifications of cryptographic protocols, web services security, concurrency in C#, secure implementations of communication abstractions, authorization policies, and access control for mobile code.

## RESEARCH

### COMPILER SUPPORT FOR SECURE MULTI-PARTY INTERACTIONS [41, 46, 49]

We designed and implemented compilers that, given high-level multiparty session descriptions, generate high performance custom cryptographic protocols. Our sessions specify pre-arranged patterns of message exchanges and data accesses between distributed participants. They provide each participant with strong security guarantees for all their messages and integrity and secrecy support for a virtual global store and dynamic principal selection, thus permitting simple, abstract reasoning on global control and data flows.

Our compiler generates code for sending and receiving these messages, with cryptographic operations and checks, to enforce these guarantees against any adversary that may control both the network and coalitions of session participants. As part of the compilation process, we automatically verify that the generated code is secure by relying on F7, a recent type system for security. Most of the proof is performed by mechanized type checking and does not rely on the correctness of our compiler.

As a result of evaluating our first compiler, our approach to secure sessions evolved. Initially, we supported session graphs with limited expressivity for which we constructed a hand-crafted proof of correctness of the compiler:

this theorem states that for all session graphs satisfying certain well-formedness conditions, the cryptographic implementation generated by the compiler guaranteed the security goals. As we extended the expressiveness of sessions for the second version of our compiler, we were confronted with the difficulty of reworking long and delicate hand-constructed proofs for an increasingly complicated compiler, as well as the desire to reduce to a minimum all non-mechanised proof. This motivated us to consider a different, more robust approach whereby we no longer verify the compiler by hand but instead modify the compiler to automatically instrument the generated code with refinement type annotations permitting a mechanised verification of security invariants directly by F7.

We also developed a more general theory of secure multiparty sessions, and obtained sufficient implementability conditions for a large class of sessions expressed as concurrent processes. We finally prototyped modular session libraries with extended support for forks, joins, and nested sessions using variants of F7 and Fine, leading to a simpler and more uniform programming interface.

## LANGUAGE SUPPORT AND VERIFICATION FOR SECURE AUDIT LOGS [45, 42, 32, 50]

In an optimistic approach to security, one can often simplify protocol design by relying on audit logs, which can be analyzed a posteriori, in case there is a conflict. Such audit logs are widely used in practice, but no formal studies guaranteed that the logged data suffices to reconstruct past runs of the protocol, to reliably detect malicious behavior, and to provide strong evidence of such behavior.

Building on process language techniques, we first formalized audit logs first for a sample optimistic scheme, the value commitment, then for a more sophisticated e-cash protocol. In both cases we used standard cryptographic mechanisms to implement secure logs and we showed that our distributed implementations either respects the semantics of commitments or, using the information stored in the logs, proves that one participant has cheated [50, 32]. We then studied a general scheme to generate audit trails. Given an F# (a dialect of OCaml) program that implements some protocol, we discovered that the expected auditable properties can be expressed and verified using the F7 type system. We have thus developed a first formal framework for auditability, with automated verification by typechecking [45].

## TEAM

| Team leader | FOURNET | Cédric | *Microsoft Research Cambridge* |
|---|---|---|---|
| Researcher | BHARGAVAN | Karthik | *INRIA Paris-Rocquencourt* |
| Researcher | BARTHE | Gilles | *INRIA Sophia Antipolis-Méditerranée* |
| Researcher | CORIN | Ricardo | *INRIA Paris-Rocquencourt* |
| Researcher | GREGOIRE | Benjamin | *INRIA Sophia Antipolis-Méditerranée* |
| Researcher | LEIFER | James | *INRIA Paris-Rocquencourt* |
| Researcher | REZK | Tamara | *INRIA Sophia Antipolis-Méditerranée* |
| Researcher | ZAPPA NARDELLI | Francesco | *INRIA Paris-Rocquencourt* |
| Post doc | LE GUERNIC | Gurvan | *MSR-INRIA Joint Centre* |
| Post doc | ZALINESCU | Eugen | *MSR-INRIA Joint Centre* |
| Post doc | REZK | Tamara | *MSR-INRIA Joint Centre* |
| Post doc | PIRONTI | Alfredo | *MSR-INRIA Joint Centre* |
| Post doc | STRUB | Pierre-Yves | *MSR-INRIA Joint Centre* |
| PHD student | GUTS | Nataliya | *Université Paris 6* |
| PHD student | PAIOLO | Miriam | *Ecole Normale Supérieure de Paris* |
| PHD student | CADE | David | *Ecole Normale Supérieure de Cachan* |
| PHD student | ZANELLA | Santiago | *Université Nice Sophia Antipolis* |
| PHD student | DENIELOU | Pierre-Malo | *Ecole Normale Supérieure de Cachan* |
| PHD student | ZHENGQIN | Luo | *Université Nice Sophia Antipolis* |
| PHD student | PLANUL | Jérémy | *Ecole Normale Supérieure de Lyon* |

**Visitors:** *Cosimo Laneve (University of Bologna, Italy) and Pedro Adao (Instituto de Telecomunicações Lisboa, Portugal) as regular visitors. Carl Gunter (University of Illinois, USA) in 2010.*

## VERIFICATION TOOLS FOR CRYPTOGRAPHIC PROTOCOL IMPLEMENTATIONS

In the past decade, an increasing amount of sensitive data is being generated, manipulated, and accessed through web applications, from bank accounts, to health information, to government records. The security of these applications relies on a combination of access control policies, secure databases, and cryptographic protocols. Even a single design flaw or implementation bug in a cryptographic protocol implementation may allow a malicious attacker to bypass all the security protections and steal, modify, or forge sensitive data. Recent research in the security community has yielded excellent tools for analyzing *models* of cryptographic protocols. We aim to build upon these tools to develop verification tools for *implementations* of cryptographic protocols, and to demonstrate their use on implementations of industrial standards such as TLS.

When verifying protocols, there are two widely-accepted styles of modeling cryptographic libraries. In the *symbolic* or *formal* style, cryptographic constructions are treated as idealized black-boxes; for example, a hash function produces a unique hash for each argument, but the hash is an opaque value that cannot be inverted. The adversary is modeled as an arbitrary program that has full control over the network and can access all functions and values publicly exported by the protocol, but cannot break the symbolic abstraction of cryptography. In the *computational* style, cryptographic constructions are treated as mathematical functions over bitstrings and their properties are described in terms of their probabilistic behavior. For example, a hash function only produces a unique hash with high probability (i.e. collisions are allowed with low probability). Moreover, hash functions are not assumed to preserve the secrecy of the argument. In general, the symbolic style is more popular amongst protocol analysts and is considered better suited for automated verification, whereas the computational style is used by cryptographers when they mathematically prove properties about cryptographic constructions. One of the goals of this project is to verify protocol implementations in both these styles, and hence experimentally assess their relative advantages.

## REFINEMENT TYPES FOR SECURITY LIBRARIES
[28, 42, 40]

Specialized provers, such as ProVerif and CryptoVerif, rely on a number of cryptography-specific heuristics to enable verification of protocol models that use complex cryptographic constructions. The cost of using these tools is that the whole protocol implementation has to be faithfully translated to a process model and then verified; the generated processes can grow quite large and the time and memory required for verification does not scale very well. On the other hand, the program verification community has made great strides in verifying general properties of programs by combining program analysis techniques, such as dependent type systems, with rapidly improving logical provers, such as SMT solvers. We have been investigating the use of these general-purpose verification techniques for analyzing cryptographic protocol implementations.

In collaboration with Microsoft Research, we are developing a refinement typechecker for F#, called F7. This typechecker implements a dependent type system, called RCF, which enables programmers to annotate their functions with pre- and post-conditions written in a general first-order logic. Hence, typechecking results in first-order logic

proof obligations, that are passed to the SMT solver Z3. We have developed a series of typed cryptographic libraries in RCF and used them to build and verify (by typechecking) cryptographic protocol implementations for a series of protocols for web services security. These libraries are now part of the latest release of F7.

The RCF type system enables the verification of higher-order functions, such as functions for mapping and folding over lists. However, using such functions imposes a heavy annotation burden on the programmer. In many cases, the programmer must locally copy the corresponding list-processing function and typecheck it for its specific use in his program. To alleviate this burden, we have developed an extension of F7 with special predicates representing the pre- and post-conditions of functions, and used them to write typed implementations of library modules, such as lists. We demonstrate that using these new libraries greatly reduces the annotations needed to typecheck programs. We use these libraries to implement cryptographic protocols that heavily use list processing, such as those for XML digital signatures, and X.509 public-key certificate chains.

## CERTICRYPT

As cryptographic proofs have become essentially unverifiable, cryptographers have argued in favor of developing techniques that help tame the complexity of their proofs. Game-based techniques provide a popular approach in which proofs are structured as sequences of games, and in which proof steps establish the validity of transitions between successive games. Code-based techniques form an instance of this approach that takes a code-centric view of games, and that relies on programming language theory to justify proof steps. While code-based techniques contribute to formalize the security statements precisely and to carry out proofs systematically, typical proofs are so long and involved that formal verification is necessary to achieve a high degree of confidence.

CertiCrypt [36] is a general framework built on top on the Coq proof assistant to certify the security of game-based cryptographic schemes using a code-based approach. The adoption of programming idioms allows giving precise definitions of games, and allows justifying rigorously proof steps using programming language methods like relational Hoare logic, static analysis, and program transformations. Additionally, CertiCrypt implements many specific techniques that arise commonly in cryptographic proofs, in particular failure events and eager and lazy sampling—an inter-procedural optimization that moves random assignments across procedures, and embeds several existing developments of complex mathematical components, in particular arithmetic, group theory and elliptic curves. One specificity of CertiCrypt is that proofs can be verified independently and automatically by a small trustworthy checker; it has been successfully applied to verify prominent cryptographic constructions, including OAEP [35], FDH [47], and zero-knowledge protocols [38].

## OTHERS ACTIVITIES

**Organization of Conferences and Seminars** Fournet organized the 6th workshop on Formal and Computational Cryptography, as part of FLOC'10 in Edinburgh. Rezk co-organized in September 2009 a one-day workshop supported by the Joint Lab on verification at INRIA (The SAFA workshop), Sophia Antipolis.

We also organized several informal one-day seminars in Orsay on language-based security and the formalization of cryptography.

**Teaching** Blanchet and Fournet gave lectures on "Cryptographic protocols: formal and computational proofs" at the Parisian Master of Research in Computer Science (MPRI) in 2009, 2010, and 2011. Blanchet gave lectures on the verification of security protocols at Padova University in 2009. Fournet gave lectures at the Cosyproof spring school and at the FOSAD summer school in 2010, and gave an invited seminar at Collège de France. Grégoire and Rezk gave lectures on "Security and Verification: Provable Cryptography" at U. Nice Sophia-Antipolis in 2009, 2010, and 2011.

## CRYPTOGRAPHIC ENFORCEMENT OF INFORMATION-FLOW SECURITY [43, 44]

Information security in distributed systems usually entails the implementation of protection mechanisms based on cryptography to ensure confidentiality and integrity. This involves expert knowledge, as well as attention to many implementation details. Our goal is to let developers focus on high-level policies and properties of their programs, and use a compiler to generate low-level protection mechanisms that ensure that the distributed implementation is at least as secure as the source program.

In language-based security, confidentiality and integrity policies specify the permitted flows of information between parts of a system with different levels of trust. These policies enable a simple treatment of security, but their enforcement is delicate. We designed and implemented CFLOW, a compiler [43] from an imperative language with locality and security annotations (with selective declassification of information), down to efficient cryptographic distributed implementations coded in F#. In source programs, security depends on a policy for reading and writing the shared variables. In their implementations, shared memory is unprotected, and security depends instead on encryption and signing. Attackers are in charge of the global control flow, as well as the untrusted part of the computation. Our most precise theorem is that, for any computational attack that may succeed against our compiled distributed code, there exists another, computational attack that succeeds against the source program, with the same result and at least the same probability of success; this enables the programmer to securely reason just on the source program.

We have also explored applications of information-flow techniques to the verification of hardware-supported multi-level programs, relying on trusted platform modules for securing the higher-level parts of their execution [44]. To minimize trust assumptions, we rely on cryptographic protection, and we exploit hardware and software mechanisms available on modern architectures, such as virtualization, secure boots, trusted platform modules, and remote attestation. We developed a precise security model for these mechanisms in an imperative language with dynamic code loading. We define program transformations to generate trusted virtual hosts and to run them on untrusted machines. As before, we obtain confidentiality and integrity theorems under realistic assumptions on cryptography, showing that the compiled distributed system is at least as secure as the source program.

We are currently experimenting with a more flexible information-flow type system for a variety of encryptions primitives, reflecting their diverse functional and security features. In particular we want to feature homomorphic encryption primitives. We aim at providing a uniform framework for understanding their properties, and for automatically checking their usage in cryptographic programs and compilers.

In summary, we show how to compile high-level programs with information flow policies to distributed systems, with adequate cryptographic protection to preserve their confidentiality and integrity properties. We believe this approach provides a safer, more reliable alternative to custom cryptographic protocol design. ■

### TOOLS AND SOFTWARE

We mention only new tools developed at MSR-INRIA for the last two years; we also contributed to other tools, such as F7, Fine, and F⋆ (primarily developed at Microsoft Research) and ProVerif and CryptoVerif (primarily developed at ENS Ulm). Earlier tools also developed in our project include a verified reference implementation in F# of a subset of the TLS 1.0 Internet Standard and preliminary prototypes of our session compiler.

**CFLOW:** a compiler from global, imperative programs with information-flow security policies down to networked, distributed cryptographic implementations in ML and in C.

**S2ML:** a verifying compiler from multiparty session specifications to distributed cryptographic protocol implementations in ML.

**CertiCrypt:** a Coq library formalizing the operational semantics and game-based proof techniques for security proofs in the computational model of cryptography.

# PUBLICATIONS & TALKS

## JOURNAL PAPERS AND BOOK CHAPTERS


[26] GILLES BARTHE, TAMARA REZK, ALEJANDRO RUSSO, AND ANDREI SABELFELD. Security of multithreaded programs by compilation. ACM Trans. Inf. Syst. Secur., 13(3), 2010.

[27] MORITZ Y. BECKER, CÉDRIC FOURNET, AND ANDREW D. GORDON. SecPAL: Design and semantics of a decentralized authorization language. Journal of Computer Security (Special issue for CSF'07), 2009.

[28] JESPER BENGTSON, KARTHIKEYAN BHARGAVAN, CÉDRIC FOURNET, ANDREW D. GORDON, AND SERGIO MAFFEIS. Refinement types for secure implementations. ACM Transactions on Programming Languages and Systems, 2010. To appear. An short version appears in CSF'08. See also Microsoft Research Technical Report MSR-TR-2008-118 SP1.

[29] BRUNO BLANCHET. Automatic verification of correspondences for security protocols. Journal of Computer Security, 17(4):363–434, July 2009.

[30] BRUNO BLANCHET. Using Horn clauses for analyzing security protocols. In Véronique Cortier and Steve Kremer, editors, Formal Models and Techniques for Analyzing Security Protocols, volume 5 of Cryptology and Information Security Series. IOS Press, March 2011.

[31] ANDREW D. GORDON AND CÉDRIC FOURNET. Principles and applications of refinement types. In Javier Esparza, Bernd Spanfelner, and Orna Grumberg, editors, International Summer School Logics and Languages for Reliability and Security, Marktoberdorf, August 2009. IOS Press, October 2010. Also Technical Report MSR-TR-2009-147.


## CONFERENCE AND WORKSHOP PAPERS


[32] CÉDRIC FOURNET PEDRO ADÃO, NATALIYA GUTS, AND FRANCESCO ZAPPA NARDELLI. High-level programming for E-cash (extended abstract). In 1st Computational and Symbolic Proofs of Security Workshop, 2009.

[33] BRUNO BARRAS, JEAN-PIERRE JOUANNAUD, PIERRE-YVES STRUB, AND QIAN WANG. COQ MTU: a higher-order type theory with a predicative hierachy of universes parameterized by a decidable first-order theory. In Logics in Computer Science (LICS'11), 2011.

[34] GILLES BARTHE, BENJAMIN GRÉGOIRE, SYLVAIN HERAUD, AND SANTIAGO ZANELLA BÉGUELIN. Formal certification of ElGamal encryption. A gentle introduction to CertiCrypt. In 5th International workshop on Formal Aspects in Security and Trust, FAST 2008, volume 5491 of Lecture Notes in Computer Science, pages 1–19. Springer, 2009.

[35] GILLES BARTHE, BENJAMIN GRÉGOIRE, YASSINE LAKHNECH, AND SANTIAGO ZANELLA BÉGUELIN. Beyond provable security. Verifiable IND-CCA security of OAEP. In Topics in Cryptology – CT-RSA 2011, volume 6558 of Lecture Notes in Computer Science, pages 180–196. Springer, 2011.

[36] GILLES BARTHE, BENJAMIN GRÉGOIRE, AND SANTIAGO ZANELLA BÉGUELIN. Formal certification of code-based cryptographic proofs. In 36th ACM SIGPLAN-SIGACT symposium on Principles of Programming Languages, POPL 2009, pages 90–101. ACM, 2009.

[37] GILLES BARTHE, BENJAMIN GRÉGOIRE, AND SANTIAGO ZANELLA BÉGUELIN. Programming language techniques for cryptographic proofs. In 1st International conference on Interactive Theorem Proving, ITP 2010, volume 6172 of Lecture Notes in Computer Science, pages 115–130. Springer, 2010.

[38] GILLES BARTHE, DANIEL HEDIN, SANTIAGO ZANELLA BÉGUELIN, BENJAMIN GRÉGOIRE, AND SYLVAIN HERAUD. A machine-checked formalization of Sigma-protocols. In 23rd IEEE Computer Security Foundations symposium, CSF 2010, pages 246–260. IEEE Computer Society, 2010.

[39] GILLES BARTHE, ALEJANDRO HEVIA, ZHENGQIN LUO, TAMARA REZK, AND BOGDAN WARINSCHI. Robustness guarantees for anonymity. In The 23rd IEEE Computer Security Foundations Symposium, CSF 2010, Edinburgh, United Kingdom, July 17-19, 2010, pages 91–106, 2010.

[40] K. BHARGAVAN, C. FOURNET, AND A. D. GORDON. Modular verification of security protocol code by typing. In ACM Symposium on Principles of Programming Languages (POPL'10), pages 445–456. Association for Computing Machinery, January 2010.

[41] KARTHIKEYAN BHARGAVAN, RICARDO CORIN, PIERRE-MALO DENIÉLOU, CÉDRIC FOURNET, AND JAMES J. LEIFER. Cryptographic protocol synthesis and verification for multiparty sessions. In 22nd IEEE Computer Security Foundations Symposium (CSF'09), pages 124–140, July 2009.

[42] KARTHIKEYAN BHARGAVAN, CÉDRIC FOURNET, AND NATALIYA GUTS. Typechecking higher-order security libraries. In APLAS, pages 47–62, 2010.

[43] CÉDRIC FOURNET, GURVAN LE GUERNIC, AND TAMARA REZK. A security-preserving compiler for distributed programs: from information-flow policies to cryptographic mechanisms. In ACM Conference on Computer and Communications Security (CCS'09), pages 432–441, Chicago, Illinois, USA, November 2009. ACM.

[44] CÉDRIC FOURNET AND JÉRÉMY PLANUL. Compiling information-flow security to minimal trusted computing bases. In Gilles Barthe, editor, Programming Languages and Systems (ESOP'11), volume 6602 of Lecture Notes in Computer Science, pages 216–235, March 2011.

[45] NATALIYA GUTS, CÉDRIC FOURNET, AND FRANCESCO ZAPPA NARDELLI. Reliable evidence: Auditability by typing. In M. Backes and P. Ning, editors, 14th European Symposium on Research in Computer Security (ESORICS 2009), volume 5789 of Lecture Notes in Computer Science, pages 168–183. Springer-Verlag, September 2009.

[46] JÉRÉMY PLANUL, RICARDO CORIN, AND CÉDRIC FOURNET. Secure enforcement for global process specifications. In M. Bravetti and G. Zavattaro, editors, 20th International Conference on Concurrency Theory (CONCUR'09), volume 5710 of Lecture Notes in Computer Science, pages 511–526, 2009.

[47] SANTIAGO ZANELLA BÉGUELIN, BENJAMIN GRÉGOIRE, GILLES BARTHE, AND FEDERICO OLMEDO. Formally certifying the security of digital signature schemes. In 30th IEEE symposium on Security and Privacy, S&P 2009, pages 237–250. IEEE Computer Society, 2009.

## THESES

[48] DAVID CADÉ. Traduction de spécifications en implémentations protocoles. Master's thesis, Université Paris VII, September 2009.

[49] PIERRE-MALO DENIÉLOU. Sûreté des abstractions et sessions sécurisées dans les langages distribués. PhD thesis, Université Paris 7, 2010. *http://moscova.inria.fr/~denielou/these/*.

[50] NATALIYA GUTS. Auditability for Security Protocols. PhD thesis, Université Paris Diderot - Paris 7, 2011.

[51] MIRIAM PAIOLA. Extending ProVerif's resolution algorithm for verifying group protocols. Master's thesis, University of Padova, May 2010.

[52] SANTIAGO ZANELLA BÉGUELIN. Formal Certification of Game-Based Cryptographic Proofs. PhD thesis, Ecole Nationale Supérieure des Mines de Paris – Mines ParisTech, 2010.

## TALKS

[53] PEDRO ADÃO. High-level programming for E-cash. CoSyProofs, AIST, Japan, April 2009.

[54] KARTHIKEYAN BHARGAVAN. Formal security analysis of cryptographic protocol code. Four week module, Indian Institute of Technology, Delhi, October 2010.

[55] KARTHIKEYAN BHARGAVAN. Modular verification of security protocol code by typing. LSV Seminar, ENS Cachan, January 2010.

[56] KARTHIKEYAN BHARGAVAN. Scalable verification of security protocol code by typechecking. Gallium Seminar, INRIA Paris-Rocquencourt, January 2010.

[57] KARTHIKEYAN BHARGAVAN. Scalable verification of security protocol code by typechecking. SECSI working group, ENS Cachan, March 2010.

[58] KARTHIKEYAN BHARGAVAN. Typechecking higher-order security libraries. Asian Symposium on Programming Languages and Systems (APLAS), Shanghai, November 2010.

[59] BRUNO BLANCHET. The automatic protocol verifier ProVerif. SecRet workshop, Valencia, Spain (invited talk), June 2010.

[60] BRUNO BLANCHET. CryptoVerif: A computationally sound mechanized prover for cryptographic protocols. CryptoForma workshop, IHP, Paris, France (invited talk), May 2010.

[61] BRUNO BLANCHET. From a concurrency course to automatic verification of process equivalences. Anniversary workshop in honour of Gérard Berry and Jean-Jacques Lévy, Gérardmer, France, February 2011.

[62] BRUNO BLANCHET AND DAVID POINTCHEVAL. Automatic, computational proof of EKE using CryptoVerif. Computational and Symbolic Proofs of Security, Spring School and French-Japanese collaboration workshop, Barbizon, France, April 2010.

[63] BRUNO BLANCHET AND DAVID POINTCHEVAL. Automatic, computational proof of EKE using CryptoVerif. Seminar, University of Padova, Italy, May 2010.

[64] BRUNO BLANCHET AND DAVID POINTCHEVAL. Automatically verified mechanized proof of one-encryption key exchange. Seminar, Munich, Germany, December 2010.

[65] BRUNO BLANCHET AND DAVID POINTCHEVAL. The computational and decisional Diffie-Hellman assumptions in CryptoVerif. Workshop on formal and computational cryptography (FCC'10), Edinburgh, UK, July 2010.

[66] DAVID CADÉ. From CryptoVerif specifications to computationally secure implementations of protocols (work in progress). Workshop on Formal and Computational Cryptography (FCC 2009), Port Jefferson, NY, USA, July 2009.

[67] CÉDRIC FOURNET. Computational soundness for cryptographic typechecking. CoSyProof: computational and symbolic proofs of security, AIST, Japan, April 2009.

[68] CÉDRIC FOURNET. A cryptographic compiler for information-flow security. Dagstuhl Seminar 09141: Web Application Security, March 2009.

[69] CÉDRIC FOURNET. A cryptographic protocol compiler for multiparty sessions. Invited talk. ACM SIGPLAN Workshop on ML, August 2009.

[70] CÉDRIC FOURNET. On the computational soundness of cryptographic verification by typing. Workshop on Formal and Computational Cryptography (FCC'09), July 2009.

[71] CÉDRIC FOURNET. Cryptographic verification of protocol implementations by typing. Invited talk. Twenty-sixth Conference on the Mathematical Foundations of Programming Semantics (MFPS 26), Ottawa, Canada, May 2010.

[72] CÉDRIC FOURNET. Refinement types for cryptography. Invited talk. International Workshop on Relations and Data Integrity Constraints and Languages, Cambridge, May 2010.

[73] CÉDRIC FOURNET. Verifying private authentication (by programming and typing). MSR Privacy Workshop, Redmond, October 2010.

[74] CÉDRIC FOURNET. Compiling information-flow security to small trusted computing bases. Microsoft Research, Redmond, March 2011.

[75] CÉDRIC FOURNET. Vérification automatique et cryptographie. Microsoft Techdays, Paris, February 2011.

[76] CÉDRIC FOURNET AND JÉRÉMY PLANUL. Compiling information-flow security to small TCBs. Anniversary workshop in honour of Gérard Berry and Jean-Jacques Lévy, Gérardmer, France, February 2011.

[77] GURVAN LE GUERNIC. A security-preserving compiler for distributed programs. FormaCrypt project meeting, June 2009.

[78] GURVAN LE GUERNIC. Cflow: A security-preserving cryptography-implicit compiler for distributed programs. DIWALL seminar, March 2010.

[79] GURVAN LE GUERNIC. Cflow: A security-preserving cryptography-implicit compiler for distributed programs. Theoretical Computer Science Group, CSC, KTH, June 2010.

[80] GURVAN LE GUERNIC. Cryptographically secured information flows of distributed programs. ParSec project meeting, January 2010.

[81] GURVAN LE GUERNIC. A security-preserving compiler for distributed programs. Highly Adaptable and Trustworthy Software using Formal Models project (HATS), September 2010.

[82] NATALIYA GUTS. Reliable evidence: Auditability by typing. Semantics Lunch seminar, Cambridge University, June 2009.

[83] NATALIYA GUTS. Pre- and postconditions for security typechecking. CryptoForma workshop, Paris, France, May 2010.

[84] NATALIYA GUTS. Pre- and postconditions for security typechecking. FCS PrivMod workshop, Edinburgh, UK, July 2010.

[85] NATALIYA GUTS. Type inference for f7. ANR ParSec meeting, Sophia-Antipolis, France, June 2010.

[86] MIRIAM PAIOLA. Extending ProVerif's resolution algorithm for verifying group protocols. Seminar, MSR-INRIA, June 2010.

[87] JÉRÉMY PLANUL. Compiling applications with information-flow policies to systems with trusted modules. Analysis of Security APIs workshop, Port Jefferson, New York, USA, July 2009.

[88] TAMARA REZK. A compiler for security. PARSEC meeting, Paris, January 2009.

[89] TAMARA REZK. A compiler for security. Invited talk, University of Bristol, February 2009.

[90] TAMARA REZK. A compiler for security. Invited talk, Queen Mary, February 2009.

[91] TAMARA REZK. Security by compilation in distributed computations. Seminaire Croiseé INDES-OASIS, January 2009.

[92] TAMARA REZK. Code injection. PARSEC meeting, Paris, June 2010.

[93] TAMARA REZK. Security in hop web applications. PARSEC meeting, Paris, January 2010.

[94] EUGEN ZALINESCU. Cryptographic verification of protocol implementations. CoSyProof: computational and symbolic proofs of security, AIST, Japan, April 2009.

[95] SANTIAGO ZANELLA BÉGUELIN. Formally certifying the security of digital signature schemes. ANR SCALP Meeting, Paris, France, February 2009.

[96] SANTIAGO ZANELLA BÉGUELIN. Language-based cryptographic proofs in Coq. Marelle Seminar, INRIA, Sophia Antipolis, France, January 2009.

[97] SANTIAGO ZANELLA BÉGUELIN. Cryptography and verification of probabilistic programs. VII Jornadas de Ciencias de la Computación, Universidad Nacional de Rosario, Argentina, October 2010.

[98] SANTIAGO ZANELLA BÉGUELIN. A machine-checked formalization of zero-knowledge proofs. IMDEA Software Theory Lunch, Madrid, Spain, November 2010.

Track A **This project started on summer 2006.**

# TOOLS AND METHODOLOGIES FOR FORMAL SPECIFICATIONS AND FOR PROOFS

## OVERVIEW

**"We aim in this project at building proof checker for high-level system specifications written in the TLA+ language especially specifications of concurrent and distributed systems."**

Systems are designed in a top-down fashion, from a high-level design to the final code. High level designs are usually informal, described in natural language. As a result, problems that could be caught in the design phase are often discovered only in debugging the code, when it is more difficult and expensive to correct them. The first step in catching errors at the design level is to write precise specifications of the high-level design. TLA+ is a formal language for writing such specifications

that is particularly well-suited for specifications of concurrent and distributed systems. TLA+ is based on ordinary math--the math taught in secondary schools or introductory university classes. TLA+ specifications can now be debugged using a model checker. However, many specifications are too complex to allow model checking to detect subtle errors. Writing a mathematical proof is the only way to detect these errors. To avoid errors in the proofs, they must be checked mechanically. This goal of this project is to make it practical to write and mechanically check of proofs of TLA+ specifications of real systems.

**Damien Doligez** graduated from Ecole Normale Supérieure in Paris and received his PhD at the University of Paris 7 in 1995. He is researcher at INRIA (Paris-Rocquencourt), he has been working in the Theory and Implementation of Programming Languages, on Garbage Collection and Theorem Proving. He is co-implementer of the Objective Caml system.

**Leslie Lamport** received his PhD. at Brandeis University in 1972, worked as a computer scientist at Massachusetts Computer Associates, SRI International, Digital Equipment Corporation, and Compaq. In 2001 he joined Microsoft Research at Mountain View, California. Leslie Lamport's research has been centered on concurrency and fault-tolerance. He is the inventor of several well-known concurrent and distributed algorithms, including early algorithms for tolerating ``Byzantine'' faults. He did seminal work on the theory of cache coherence and distributed systems. He has also developed methods for formally specifying and verifying concurrent systems.

## RESEARCH

Many system crashes are caused by what Jim Gray called "heisenbugs"–unreproducible errors whose cause is never discovered. We believe that many of those errors are due to errors in the algorithms used to synchronize concurrent activity. TLA+ and its model checker are effective tools for designing correct concurrent algorithms. However, the huge number of possible states caused by the highly nondeterministic nature of these algorithms limits the ability to eliminate errors by conventional model checking. For some algorithms, nothing short of mathematical proof can guarantee correctness, and mechanical checking is required to avoid errors in proofs. We hope that the TLA+ prover will enable engineers to avoid errors in highly critical synchronization algorithms.

TLA+ is a specification and proof language based on temporal logic, where first-order logic and set theory are used to describe a set of states and the possible transitions between these states. TLA+ also includes a module system for manipulating large-scale specifications.

There are a number of existing tools for working on TLA+ specifications, the most important of which is the TLC model-checker. TLA+ has already proved its worth in significant projects in hardware design (Alpha and Itanium processors), protocols (PCI-X), and software (Doligez-Leroy-Gonthier garbage collector).

In this project, we are defining an extension of the TLA+ language, called TLA$^{+2}$, for writing mathematical proofs, and we are complementing the existing tools so that users can develop, debug, and check proofs about algorithm and system specifications. In this way, TLA+ becomes a complete solution for writing, debugging, and proving specifications. More precisely, we are refining the proof language, building a development environment for TLA+ specifications and proofs, developing and adapting automatic tools that help to prove TLA+ theorems (based on the Zenon and veriT provers), and translating TLA+ proofs into a machine-checkable format for verification by an independent checker (Isabelle/TLA+).

We validate and enhance our tools by finding examples of real-world projects where formal specifications bring real improvements over other methodologies. Feedback from these examples helps us to improve the proof language and the tools and develop methods and design patterns for using TLA+.

### OTHER ACTIVITIES

Zenon participated in the CASC-J5 competition for automated theorem provers (Edinburgh, August 2010).

## TEAM

| Team leader | DOLIGEZ | Damien | *INRIA Paris-Rocquencourt* |
|---|---|---|---|
| Researcher | LAMPORT | Leslie | *Microsoft Research Silicon Valley* |
| Researcher | MERZ | Stephan | *INRIA Lorraine* |
| researcher | CHAUDHURI | Kaustuv | *INRIA Saclay-Île-de-France* |
| Post doc | COUSINEAU | Denis | *MSR-INRIA Joint Centre* |
| Post doc | KUPPE | Markus | *MSR-INRIA Joint Centre* |
| Post doc | RICKETTS | Daniel | *MSR-INRIA Joint Centre* |
| Post doc | TRISTAN | Jean-Baptiste | *MSR-INRIA Joint Centre* |
| PHD student | VANZETTO | Hernan Pablo | *Université Nancy* |
| PHD student | ZAMBROVSKY | Simon | *Hamburg University* |

*Kaustuv Chaudhuri left in November 2009, Denis Cousineau started as post-doc in November 2009. Jean-Baptiste Tristan did a 3-month internship at the end of 2009, Dan Ricketts did a 1-year internship from september 2009 to august 2010. Hernán Vanzetto from the University of Rosario visited the team for a six-month internship in 2009 when he worked on encoding arithmetic over the naturals and the integers in Isabelle/TLA+, and again for five months in 2010 for working on the SMT translation. Since december 2010 he is working as a PhD student on the construction of (counter-)models for TLA$^{+2}$ formulas.*

## TOOLS AND SOFTWARE

The Eclipse-based ToolBox is a graphical user interface that serves as the front-end for all TLA+ tools. It is interfaced with TLC, SANY, the PlusCal translator, and the PM. It includes an editor with syntax coloring, hiding and showing of subtrees of the proof, and allows the user to check parts of the proof independently of each other, reporting the proof obligations that fail.

The non-interactive proof manager (PM) is a tool for checking proofs written in TLA$^{+2}$. This PM takes as input a TLA$^{+2}$ specification and attempts to check some or all of its proofs (depending on command-line arguments) in batch mode. It reports if it fails to check any step of a proof, either because the step is incorrect or because there is insufficient detail in the proof for the PM to automatically check the proof.

## DESIGN

The PM consists principally of two stages: a front-end elaborator and a verifier. The verifier stage interacts with automated theorem provers (Zenon, CVC3, veriT) by sending them proof obligations and with a back-end framework (in our case, Isabelle) by sending it the proof obligations and the proofs generated by Zenon. These proof obligations are expressed in an encoding of TLA$^{+2}$ called Isabelle/TLA+.

**Elaboration:** The front-end elaborator is the first stage of the PM and consists of resolving names and substitutions in the input specification. At the end of this stage, the input proofs are annotated with fully elaborated usable elements at every step.

The semantics of TLA$^{+2}$ modules inlines instanced modules in their host module mediated by a substitution. A proof is allowed to use definitions and theorems from these instanced modules, so the PM must perform the substitutions and flatten all instances. For some TLA$^{+2}$ operators such as ENABLED that quantify implicitly over certain sub-expressions (in particular, primed variables), substitution can only be performed if the quantification is made explicit; therefore, the PM eliminates these operators entirely.

In addition to resolving names and substitutions, further elaborations are carried out to reduce the supported TLA$^{+2}$ language to a core elaborated form, called TLA$^{+2e}$. The primary benefit of this elaboration is to reduce the complexity of the trusted theorems base for the final certification stage. The following are some examples of elaboration:

- $\exists \langle x, y \rangle \in S : P(x, y)$
  elaborates to
  $\exists x : \exists y : (\langle x, y \rangle \in S) \wedge P(x, y)$.
  All bounded quantifiers are similarly replaced with unbounded quantifiers over single variables. ("Variable" is used here with its meaning in ordinary logic, rather than its meaning in TLA+as a flexible variable of temporal logic.)

- $[f \text{ EXCEPT } ![x_1] = e_1, ![x_2] = e_2]$
  elaborates to
  $[[f \text{ EXCEPT } ![x_1] = e_1] \text{ EXCEPT } ![x_2] = e2]$.
  This reduces EXCEPT to the status of a binary operator.

This elaboration can generate fresh bound variables, so it is necessary to ensure that elaborating the same TLA$^{+2}$ expression at two distinct points produces equal elaborated forms. The PM therefore internally represents the TLA$^{+2}$ syntax using de Bruijn indexes, which is nameless and generalizes syntactic equality to $\alpha$-equivalence. Names from the source syntax are maintained as "hints" that are used to produce named representations for output.

**Verification:** The second key stage of the proof manager is to extract proof obligations from the elaborated proof steps and interact with back-end provers and the back-end logical framework (Isabelle) to check that the obligations are true. Before discussing how verification is performed, we note that the PM is allowed to fail to verify a BY directive even though it is obviously true mathematically; in this case the user must simplify the proof further for the PM.

A proof obligation is a TLA$^{+2}$ statement of the form ASSUME $e_1, \ldots, e_m$ PROVE $g_1 \vee \ldots \vee g_n$ where $e_1, \ldots, e_m$ is the list of assumptions and usable facts at that point of the proof, and $g_1, \ldots, g_n$ are the current goals. Given the interpretation $[[-]]$ of TLA$^{+2}$ in a logical framework such as Isabelle, this proof obligation amounts to proving the sequent $[[e_1]], \ldots, [[e_m]] \vdash [[g_1]], \ldots, [[g_n]]$ in the target framework. When Isabelle accepts the generated proof script as correct, we get high assurance that the theorem is true. We also get good assurance that the source proof is correct, although in theory the PM might turn an incorrect TLA$^{+2}$ proof into a correct Isabelle proof.

## IMPLEMENTATION PROGRESS

Compared to the version of 2009, we have enhanced the treatment of modules of the PM, interfaced it with the ToolBox, and added a system of fingerprints for proof obligations that acts as a cache for proofs. Thanks to fingerprints, the user can freely tell the ToolBox to check and recheck any part of his proof without wasting time because the PM automatically avoids re-proving obligations that have already been proved. The PM, together with Zenon and the Isabelle/TLA+ theory, form a unit called TLAPS. In 2010, we have made two releases of TLAPS, synchronized with two releases of the ToolBox.

The current version of TLAPS includes a translation of proof obligations into SMT-lib format, the generic input language of SMT solvers, which mainly handles the fragment of linear arithmetic. It also includes a dedicated decision procedure for Presburger arithmetic. We are currently extending the SMT translation for handling a larger fragment of TLA$^{+2}$, including elementary set theory, functions, tuples, and records. Because SMT-lib is a multi-sorted language and overloaded symbols such as equality should be resolved differently depending on the sorts of their arguments, this translation requires a form of sort inference for a given proof obligation. Preliminary experiments indicate that SMT solvers are able to successfully discharge low-level proof obligations, complementary to automated reasoners for first-order logic such as Zenon.

TLAPS was presented in a tool paper at IJCAR 2010 [104], at an invited talk at ICTAC 2010 [103], and at a tutorial presentation at IFM 2010 [106].

## INTERFACES WITH ISABELLE AND ZENON

**TLA$^{+2}$ encoding in Isabelle** The core of TLA$^{+2}$ has been encoded as a new object logic in the generic interactive proof assistant Isabelle. More precisely, the encoding includes propositional and first-order logic, elementary set theory, functions, fixed-point constructions, and the construction of the natural numbers. The main automatic proof methods (such as rewriting, case-based reasoning, and the tableau prover) available in Isabelle have been instantiated for TLA$^{+2}$. This encoding provides the basis for the verification stage of the proof manager: it receives the proof obligations and attempts to discharge them based on the 10 available proof methods. We have chosen to define a new object logic rather than encode TLA$^{+2}$ in one of the existing logics (such as HOL or ZF); this minimizes the overhead of the translation and makes it easier to understand the error messages when Isabelle fails to prove formulas.

We have implemented support for strings, records, tuples, and natural numbers. Integer arithmetic has been defined in Isabelle, but little automation is currently available to users of TLAPS; support for reals and for the temporal logic TLA is still missing. Nevertheless, the current fragment can already be used to prove the correctness of some complex algorithms.

**Zenon** The Zenon prover initially produced proofs only in Coq format. A new back-end was written to produce proofs in Isabelle format (as Isar scripts). Zenon was extended with TLA-specific inference rules to make it more efficient on the kind of proof obligations that are produced when checking a TLA$^{+2}$ proof.

Zenon was also extended to handle the special features of the TLA+ base logic: the CHOOSE operator, n-ary CASE constructs, strings, records, sequences, etc.

## PLUSCAL

PlusCal is a high-level language for describing algorithms. The PlusCal compiler generates a TLA+ model from a PlusCal algorithm, and many of our case studies are based on models that originate from algorithms written in PlusCal. The language was presented in an invited paper at ICTAC 2009 [105]. The version 1.5, about to be released, contains several changes, including one that simplifies the translation to TLA+ by removing the variable representing the control state when it is not need. This will simplify correctness proofs of distributed algorithms written in PlusCal, such as Byzantine Paxos.

## PROOFS OF ALGORITHMS

Using TLAPS and the ToolBox, we have written and checked a safety proof of a Byzantine Paxos consensus algorithm, which abstracts and generalizes the one at the heart of the well-known Castro-Liskov algorithm. We proved that the Byzantine algorithm implements a variant of the ordinary Paxos consensus algorithm. The entire proof was checked except for a few lines of trivial temporal-logic reasoning. (TLAPS does not yet check temporal formulas.) A few very simple mathematical facts about finite sets were also assumed without proof. Writing and checking the proof revealed that the Byzantine algorithm does not implement the version of the Paxos algorithm we originally thought that it did.

We have also specified a complete chain of implementations linking the Byzantine algorithm to a simple specification

of consensus: the ordinary Paxos algorithm implements an abstract, non-distributed voting algorithm, which implements the consensus specification. A proof of both the safety and liveness parts of the final implementation has been written, and it has all been checked by TLAPS except for the temporal-logic steps and seven steps that required reasoning about the ENABLED predicate (not yet supported by TLAPS). The remaining refinement (of the voting algorithm by the ordinary Paxos algorithm) is understood well enough, and has been model-checked with a large enough model, so we are confident that it too is correct.

In cooperation with Bernadette Charron-Bost from the LIX laboratory, we studied proofs of Consensus algorithms in the Heard-Of model for distributed computing [99, 102]. In this model, one can pretend that all nodes of a distributed system act synchronously, which substantially simplifies the proof. In another case study [100], we addressed the verification of programmable logic controllers in TLA+. ■

## PUBLICATIONS & TALKS

### JOURNAL PAPERS AND BOOK CHAPTERS


[99] BERNADETTE CHARRON-BOST AND STEPHAN MERZ. Formal verification of a Consensus algorithm in the Heard-Of model. Intl. J. Software and Informatics, 3(2-3):273–204, 2009.

[100] HEHUA ZHANG, STEPHAN MERZ, AND MING GU. Specifying and verifying PLC systems with TLA+: a case study. Computers & Mathematics with Applications, 60(3):695–705, August 2010.


### CONFERENCE AND WORKSHOP PAPERS


[101] SABINA AKHTAR, STEPHAN MERZ, AND MARTIN QUINSON. Extending PlusCal: A language for describing concurrent and distributed algorithms. In Eric Cariou, Laurence Duchien, and Yves Ledru, editors, Actes des deuxièmes journées nationales du Groupement De Recherche CNRS du Génie de la Programmation et du Logiciel, France Pau, March 2010.

[102] MOUNA CHAOUCH-SAAD, BERNADETTE CHARRON-BOST, AND STEPHAN MERZ. A reduction theorem for the verification of round-based distributed algorithms. In Olivier Bournez and Igor Potapov, editors, Reachability Problems '09, volume 5797 of Lecture Notes in Computer Science, pages 93–106, Palaiseau, France, 2009. Springer.

[103] KAUSTUV CHAUDHURI, DAMIEN DOLIGEZ, LESLIE LAMPORT, AND STEPHAN MERZ. The TLA+ proof system: Building a heterogeneous verification platform. In Ana Cavalcanti, David Déharbe, Marie-Claude Gaudel, and Jim Woodcock, editors, International Conference on Theoretical Aspects of Computing - ICTAC 2010, volume 6255 of Lecture Notes in Computer Science, page 44, Brazil Natal, 2010. Springer. The original publication is available at www.springerlink.com.

[104] KAUSTUV CHAUDHURI, DAMIEN DOLIGEZ, LESLIE LAMPORT, AND STEPHAN MERZ. Verifying safety properties with the TLA+ proof system. In Jürgen Giesl and Reiner Hähnle, editors, 5th Intl. Joint Conf. Automated Reasoning (IJCAR 2010), volume 6173 of Lecture Notes in Computer Science, pages 142–148, Edinburgh, UK, 2010. Springer.

[105] LESLIE LAMPORT. The pluscal algorithm language. In Theoretical Aspects of Computing—ICTAC 2009, volume 5684 of Lecture Notes in Computer Science, pages 36–60. Springer-Verlag, 2009.


### TALKS


[106] DENIS COUSINEAU AND STEPHAN MERZ. The TLA+ proof system. Tutorial at International Conference on Integrated Formal Methods, October 2010.

[107] LESLIE LAMPORT. The PlusCal algorithm language. Keynote address at theInternational Colloquium on Theoretical Aspects of Computing, August 2009.

[108] LESLIE LAMPORT. What is computation. Invited talk at The CNRS Summer School on Communication Networks, June 2009.

[109] LESLIE LAMPORT. The PlusCal algorithm language. Keynote address at the International Conference of the Chilean Computer Society, February 2010.

[110] LESLIE LAMPORT. Preuves et prouveur TLA+. Invited Talk at Journées Francophones des Langages Applicatifs, February 2010.

[111] LESLIE LAMPORT. What is computation. Invited talk at the European Computer Science Summit, October 2010.

[112] STEPHAN MERZ. The TLA+ proof system. Institut für Informatik, Universität Augsburg, February 2011.

[113] STEPHAN MERZ. The TLA+ proof system. FB Softwaretechnik, Technische Universität Berlin, February 2011.

**Track B** **This project started at fall 2007.**

# DIGITAL DICTIONARY OF MATHEMATICAL FUNCTIONS

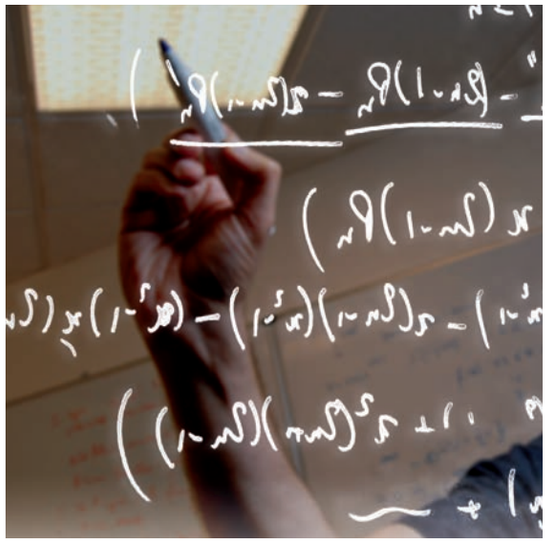
## OVERVIEW

**"Our ambition with the DDMF is to develop an authoritative interactive web site on the special functions of mathematics."**

A great deal of functions from mathematical analysis are involved in a recurrent manner in diverse domains of applied mathematics. Their properties and the mathematical identities that they satisfy have been considerably studied and documented in classical works since the 19th century. These properties and mathematical identities have recently become amenable to computer algebra, the branch of computer science that concerns itself with exact and efficient computations with general mathematical objects. Therefore, it has become natural to ask for generating books on special functions, rather than compiling them from diverse sources.

In view of this, our goal is to make the results of computations with the special functions of mathematics available to an audience that is not expert in computer algebra. To this end, we provide users with a dynamical presentation of them on the web, in the form of an online encyclopedic dictionary (*http://ddmf.msr-inria.inria.fr*). This can be viewed as a modern version of the textbooks and handbooks of the 19th and 20th centuries. For each function, our current encyclopedia shows its essential properties and mathematical objects attached to it, which are often infinite in nature (numerical evaluations, asymptotic expansions). This way of disseminating has the advantage of allowing for a presentation that adapts interactively to the user's actual needs.

**Bruno Salvy** graduated from École Polytechnique. He is a researcher at Inria since 1991, where he headed the Algorithms project from 2000 to 2008. His research centers on the interface between computer algebra and analysis of algorithms. He authored more than 50 articles in the journals and conferences of his field. He is also member of the editorial boards of the Journal of Symbolic Computation and of the Journal of Algebra (section Computational Algebra).



## RESEARCH

The originality of our work in computer algebra lies in the systematic use of linear operators as a data-structure from which various information such as identities for special functions can be extracted. Our work proceeds along three lines: design of fast algorithms either based on using linear operators or giving better complexity for operations on these operators; new algorithmic applications of linear operators; new algorithms extending the class of functions to which our methods apply.

### SPECIAL FUNCTIONS

**Interactive Online Dictionary:** The main features of the Dynamic Dictionary of Mathematical Functions (version 1.5) have been described in [128]. The web site consists of interactive tables of mathematical formulas on elementary and special functions. The formulas are automatically generated by computer algebra routines. The user can ask for more terms of the expansions, more digits of the numerical values, or proofs of some of the formulas.

**Approximation and Asymptotic Series:** A Chebyshev expansion is a series in the basis of Chebyshev polynomials of the first kind. When such a series solves a linear differential equation, its coefficients satisfy a linear recurrence equation. In [129], we interpret this equation as the numerator of a fraction of linear recurrence operators. This interpretation lets us give a simple view of previous algorithms, analyze their complexity, and design a faster one for large orders. Simple proofs of several results and conjectures formulated by Stolarsky and Tran concerning generating functions of some families of Chebyshev-like polynomials are given in [122].

The classification of possible asymptotic behaviours of solutions of linear differential equations is classical. Together with a Lindelf integral representation, this is used in [125] to study various sequences that possess explicit analytic expressions. One of the outcomes of such analyses concerns the non-existence of linear recurrences with polynomial coefficients annihilating these sequences, and, accordingly, the non-existence of linear differential equations with polynomial coefficients annihilating their generating functions. In particular, the corresponding generating functions are transcendental. Asymptotics of certain finite difference sequences come out as a byproduct of our approach.

**Guaranteed Numerical Calculations:** In [126], we describe an algorithm that takes as input a complex sequence

## TEAM

| Team leader | SALVY | Bruno | *INRIA Paris-Rocquencourt* |
|---|---|---|---|
| Researcher | BOSTAN | Alin | *INRIA Paris-Rocquencourt* |
| Researcher | CHYZAK | Frédéric | *INRIA Paris-Rocquencourt* |
| Post doc | DARRASSE | Alexis | *MSR-INRIA Joint Centre* |
| Post doc | GERHOLD | Stefan | *MSR-INRIA Joint Centre* |
| Post doc | KOUTSCHAN | Christoph | *MSR-INRIA Joint Centre* |
| Post doc | STAN | Flavia | *MSR-INRIA Joint Centre* |
| PHD student | BENOIT | Alexandre | *Ecole Polytechnique* |
| PHD student | CHEN | Shaoshi | *Ecole Polytechnique & Chinese Academy of Sciences* |
| PHD student | MEZZAROBBA | Marc | *Ecole Polytechnique* |

*Marc Mezzarobba started his PhD in October 2007 and Alexandre Benoit in September 2008. They are finishing next year.*

*Shaoshi Chen spent 6 months with us at the end of 2010 to finish his PhD, codirected with the Chinese Academy of Sciences in Beijing (China). He is now a post-doc at the Research Institute for Symbolic Computation (Linz, Austria).*

*Alexis Darrasse was a post-doc with us during 9 months in 2010 and is now post-doc at the Universit Pierre et Marie Curie (Paris).*

*Stefan Gerhold is an assistant professor at the Vienna University of Technology (Vienna, Austria). He was a post-doc with us during 6 months in the first half of 2009.*

*Flavia Stan started to be a post-doc with us in February 2011 and Christoph Koutschan in April 2011.*

given by a linear recurrence relation with polynomial coefficients along with initial values, and outputs a simple explicit upper bound on its modulus, valid for all values of the index. Generically, the bound is tight, in the sense that its asymptotic behaviour matches that of the bounded sequence. We discuss applications to the evaluation of power series with guaranteed precision. The previous algorithm, and more classical algorithms that operate on solutions of linear differential equations or recurrence relations with polynomial coefficients, have been incorporated into the software package NumGfun. The article [134] describes this implementation, including what seems to be the first general implementation of the fast high-precision numerical evaluation algorithms of Chudnovsky and Chudnovsky. In some cases, our descriptions contain small improvements over existing algorithms.

**Summation and Integration:** A very important result is in [133], where we extend Zeilberger's approach to special function identities to cases that are not holonomic. The method of creative telescoping is thus applied to definite sums or integrals involving Stirling or Bernoulli numbers, incomplete Gamma function or polylogarithms, which are not covered by the holonomic framework. The basic idea is to take into account the dimension of appropriate ideals in Ore algebras. This unifies several earlier extensions and provides algorithms for summation and integration in classes that had not been accessible to computer algebra before. Algorithms of creative telescoping are being developed for specific classes, where added efficiency is expected. For rational functions, this is related to the classical Hermite reduction and was studied in [130, 139, 141]. A structure theorem for the solutions of first-order linear differential-difference partial systems, namely multivariate hyperexponential-hypergeometric functions, is obtained in [135]. It allows us to decompose such a function as the product of a rational function, several exponential and power functions, and factorial terms, and to derive first criteria for the termination of the continuous-discrete analogue of Zeilberger´s algorithms. On the implementation side, a faster re-implementation of our general algorithms for summation and integration has been obtained. This has been reported on in [137].

## APPLICATIONS

**Application to Combinatorics:** In [120], we propose an experimental mathematics approach leading to the computer-driven discovery of various conjectures about structural properties of generating functions coming from enumeration of 2D and 3D lattice walks. Gessel walks are lattice walks in the quarter plane which start at the origin and consist only of steps chosen from the set. In [131], we prove that the number of Gessel walks of given length ending at any given point has a trivariate generating series that is an algebraic function. This is a solution to an extension of a famous conjecture due to Gessel.

**Application to Physics:** The linear differential equations satisfied by the n-particle contributions to the susceptibility of the square-lattice Ising model have a very important role in physics, but are also very difficult to obtain, as they are objects of very large sizes, which increases dramatically with n. This explains that a lot of calculations are done only modulo a prime number. In [114], we consider the case of the five-particle contribution and show that one can understand the factorization of the corresponding order-29 linear differential operator from calculations modulo several prime numbers. In [115], we study various multiple integrals related to the isotropic square Ising model. These integrals define holonomic, and even G-functions: they satisfy Fuchsian linear differential equations with polynomial coefficients, and share remarkable arithmetic properties with a geometric origin. Using an experimental mathematics approach, we show that almost all these differential operators can be interpreted as explicit modular forms of the elliptic curve of the Ising model. In [117], we exhibit an order-four linear differential operator that is not reducible to this elliptic curve, modular forms scheme. This operator is shown to correspond to a natural generalization of the elliptic setting, with the emergence of a Calabi–Yau equation associated to a selected 4F3 hypergeometric function. We explicitly compute its differential Galois group, the mirror maps and higher order Schwarzian ODEs. In relation to these studies, we give in [116] a concrete example of an infinite-order rational transformation that leaves a linear differential equation covariant. This example can be seen as a non-trivial but still simple illustration of an exact representation of the renormalization group.

## FAST ALGORITHMS

An important special class of solutions of linear operators is given by orthogonal polynomials. In [121], we discuss efficient conversion algorithms. We describe a known conversion algorithm from an arbitrary orthogonal basis to the monomial basis, and deduce a new algorithm of the same complexity for the converse operation.

In [132], we address complexity issues for linear differential equations in positive characteristic p: resolution and computation of the p-curvature. For these tasks, our main focus is on algorithms whose complexity behaves well with respect to p. We prove bounds linear in p on the degree of polynomial solutions and propose algorithms for testing

the existence of polynomial solutions in sublinear time, and for determining a whole basis of the solution space in quasi-linear time. We show that for equations of arbitrary order, the p-curvature can be computed in subquadratic time, and that this can be improved to logarithmic time for first order equations and to quasi-linear time for classes of second order equations.

The cost of multiplication modulo triangular families of polynomials is studied in [118]. Following previous work by Li, Moreno Maza, and Schost, we propose an algorithm that relies on homotopy and fast evaluation-interpolation techniques. We obtain a quasi-linear time complexity for substantial families of examples, for which no such result was known before. Notably, applications are given to addition of algebraic numbers in small characteristic.

A relatively simple algorithm with a low constant factor for exponentials of truncated power series is given in [123].

In [124], we present algorithms to perform modular polynomial multiplication or a modular dot product efficiently in a single machine word. We use a combination of techniques. Polynomials are packed into integers by Kronecker substitution; several modular operations are performed at once with machine integer or floating point arithmetic; normalization of modular images is avoided when possible; some conversions back to polynomial coefficients are avoided; the coefficients are recovered efficiently by preparing them before conversion. We discuss precisely the required control on sizes and degrees. We then present applications to polynomial multiplication, prime field linear algebra and small extension field arithmetic, where these techniques lead to practical gains of quite large constant factors.

The complexity of the cyclic-vector method and the quality of the uncoupled systems it obtains are analysed in [136].

Efficient algorithms on fundamental objects like integers, polynomials, matrices, series, and solutions to linear differential equations and linear recurrences are surveyed in the lecture notes [138]. There, it is shown how numerous questions on these objects admit a solution in quasi-optimal complexity. This includes classical algorithms as well as more recent ones, developed by the team. ■

## PUBLICATIONS & TALKS

### JOURNAL PAPERS AND BOOK CHAPTERS


[114] A. BOSTAN, S. BOUKRAA, A. J. GUTTMANN, S. HASSANI, I. JENSEN, J.-M. MAILLARD, AND N. ZENINE. High order Fuchsian equations for the square lattice Ising model: (5). Journal of Physics A: Mathematical and Theoretical, 42(27):32pp, 2009.

[115] A. BOSTAN, S. BOUKRAA, S. HASSANI, J. M. MAILLARD, J. A. WEIL, AND N. ZENINE. Globally nilpotent differential operators and the square Ising model. Journal of Physics A: Mathematical and Theoretical, 42(12):50pp, 2009.

[116] A. BOSTAN, S. BOUKRAA, S. HASSANI, J.-M. MAILLARD, J.-A. WEIL, N. ZENINE, AND N. ABARENKOVA. Renormalization, isogenies and rational symmetries of differential equations. Advances in Mathematical Physics, 2010:44pp, 2010.

[117] A. BOSTAN, S. BOUKRAA, S. HASSANI, M. VAN HOEIJ, J.-M. MAILLARD, J.-A. WEIL, AND N. ZENINE. The Ising model: from elliptic curves to modular forms and Calabi-Yau equations. Journal of Physics A: Mathematical and Theoretical, 44(4):44pp, 2011.

[118] A. BOSTAN, M.F.I. CHOWDHURY, J. VAN DER HOEVEN, AND É. SCHOST. Homotopy methods for multiplication modulo triangular sets. Journal of Symbolic Computation, To appear.

[119] ALIN BOSTAN AND PHILIPPE DUMAS. Wronskians and linear independence. American Mathematical Monthly, 117(8):722–727, 2010.

[120] ALIN BOSTAN AND MANUEL KAUERS. The complete generating function for Gessel walks is algebraic. Proceedings of the American Mathematical Society, 138(9):3063–3078, September 2010. With an Appendix by Mark van Hoeij.

[121] ALIN BOSTAN, BRUNO SALVY, AND ÉRIC SCHOST. Fast conversion algorithms for orthogonal polynomials. Linear Algebra and its Applications, 432(1):249–258, January 2010.

[122] ALIN BOSTAN, BRUNO SALVY, AND KHANG TRAN. Generating functions of Chebyshev-like polynomials. International Journal of Number Theory, 6(7):1659–1667, 2010.

[123] ALIN BOSTAN AND ÉRIC SCHOST. A simple and fast algorithm for computing exponentials of power series. Information Processing Letters, 109(13):754–756, 2009.

[124] JEAN-GUILLAUME DUMAS, LAURENT FOUSSE, AND BRUNO SALVY. Simultaneous modular reduction and Kronecker substitution for small finite fields. Journal of Symbolic Computation, To appear.

[125] PHILIPPE FLAJOLET, STEFAN GERHOLD, AND BRUNO SALVY. Lindelöf representations and (non-)holonomic sequences. The Electronic Journal of Combinatorics, 17(1):1–28, 2010.

[126] MARC MEZZAROBBA AND BRUNO SALVY. Effective bounds for P-recursive sequences. Journal of Symbolic Computation, 45(10):1075–1096, October 2010.

[127] BRUNO SALVY AND JOHN SHACKELL. Measured limits and multiseries. Journal of the London Mathematical Society, 82(3):747–762, 2010.

## CONFERENCE AND WORKSHOP PAPERS

[128] ALEXANDRE BENOIT, FRÉDÉRIC CHYZAK, ALEXIS DARRASSE, STEFAN GERHOLD, MARC MEZZAROBBA, AND BRUNO SALVY. The Dynamic Dictionary of Mathematical Functions (DDMF). In Komei Fukuda, Joris van der Hoeven, Michael Joswig, and Nobuki Takayama, editors, The Third International Congress on Mathematical Software (ICMS 2010), volume 6327 of Lecture Notes in Computer Science, pages 35–41, 2010.

[129] ALEXANDRE BENOIT AND BRUNO SALVY. Chebyshev expansions for solutions of linear differential equations. In John May, editor, ISSAC '09: Proceedings of the twenty-second international symposium on Symbolic and algebraic computation, pages 23–30, 2009.

[130] ALIN BOSTAN, SHAOSHI CHEN, FRÉDÉRIC CHYZAK, AND ZIMING LI. Complexity of creative telescoping for bivariate rational functions. In ISSAC '10: Proceedings of the 2010 International Symposium on Symbolic and Algebraic Computation, pages 203–210, New York, NY, USA, 2010. ACM.

[131] ALIN BOSTAN AND MANUEL KAUERS. Automatic classification of restricted lattice walks. In DMTCS Proceedings of the 21st International Conference on Formal Power Series and Algebraic Combinatorics (FPSAC'09), Hagenberg, Austria, pages 203–217, 2009.

[132] ALIN BOSTAN AND ÉRIC SCHOST. Fast algorithms for differential equations in positive characteristic. In John May, editor, ISSAC'09, pages 47–54, 2009.

[133] FRÉDÉRIC CHYZAK, MANUEL KAUERS, AND BRUNO SALVY. A non-holonomic systems approach to special function identities. In John May, editor, ISSAC '09: Proceedings of the twenty-second international symposium on Symbolic and algebraic computation, pages 111–118, 2009.

[134] MARC MEZZAROBBA. Numgfun: a package for numerical and analytic computation with D-finite functions. In Proceedings of the 2010 International Symposium on Symbolic and Algebraic Computation (ISSAC 2010), pages 139–145. ACM, 2010.

## THESES

[135] SHAOSHI CHEN. Quelques applications de l'algèbre différentielle et aux différences pour le télescopage créatif. PhD thesis, École polytechnique (Palaiseau, France), February 2011. To be defended on February 16, 2011.

[136] ÉLIE DE PANAFIEU. Complexité et qualité du découplage obtenu par la méthode du vecteur cyclique. Master's thesis, École Normale Supérieure, août 2010. 26 pages.

[137] LUCIEN PECH. Algorithmes pour la sommation et l'intégration symboliques. Master's thesis, École Normale Supérieure, novembre 2009. 22 pages.

## POSTERS, LECTURE NOTES

[138] ALIN BOSTAN. Algorithmes rapides pour les polynômes, séries formelles et matrices. In Actes des Journées Nationales de Calcul Formel, Les cours du CIRM, tome 1, numéro 2, pages 75–262, Luminy, France, 2010.

[139] ALIN BOSTAN, FRÉDÉRIC CHYZAK, SHAOSHI CHEN, AND ZIMING LI. Rational-functions telescopers: Blending creative telescoping with Hermite reduction. Poster at the conference ISSAC'09 (Seoul, South Korea), 2009.

[140] ALEXANDRE CASAMAYOU, GUILLAUME CONNAN, THIERRY DUMONT, LAURENT FOUSSE, FRANÇOIS MALTEY, MATTHIAS MEULIEN, MARC MEZZAROBBA, CLÉMENT PERNET, NICOLAS M. THIÉRY, AND PAUL ZIMMERMANN. Calcul mathématique avec sage. Web publication, 2010.

[141] SHAOSHI CHEN AND ZIMING LI. A speed-up of the Hermite reduction for rational functions. Poster at the conference ICMM'09 (Beijing, China), 2009.

## TALKS

[142] ALIN BOSTAN. Algébricité de la série génératrice complète des chemins de Gessel. Invited talk at the Journées hypergéométriques 2010 (Grenoble, France), April 2010.

[143] ALIN BOSTAN. Algorithmes rapides pour les polynômes, séries formelles et matrices. Tutorial at the Journées Nationales du Calcul Formel 2010 (Luminy, France), May 2010.

[144] FRÉDÉRIC CHYZAK. Algorithmique symbolique pour les fonctions spéciales. Tutorial at Algorithmique et programmation 2010 (Luminy, France), May 2010.

[145] FRÉDÉRIC CHYZAK. Explicit formula for the generating series of diagonal 3D rook paths. Invited talk at The Renaissance of Combinatorics — Advances, Algorithms, Applications (Tianjin, China), August 2010.

[146] BRUNO SALVY. Automatic proofs of identities: Beyond A=B. Invited talk at Formal Power Series and Algebraic Combinatorics FPSAC'09 (Linz, Austria), July 2009.

[147] BRUNO SALVY. The dynamic dictionary of mathematical functions. Invited talk at the Conferences on Intelligent Computer Mathematics, CICM'10 (Paris, France), July 2010.

[148] BRUNO SALVY. Newton iteration: From numerics to combinatorics, and back. Invited talk at Analytic Algorithmics and Combinatorics ANALCO'10 (Austin, Texas), January 2010.

**Track B** **This project started in December 2007.**

# ReActivity

## OVERVIEW

**"The goal of this project is to explore how to capture and visualize a user activity, enabling scientists to reflect upon, interact with and improve their research processes."**

As lead users of networked computing, scientists face an ever increasing flood of data and must master a constantly evolving set of tools for searching and analyzing both physical and on-line data. Scientific exploration often involves successive modifications to a primary experiment, with similar but not identical analysis processes. Through trial and error, scientists discover specific tools and strategies that work. Unfortunately, managing this process remains cumbersome and time consuming, relying upon ad hoc techniques to recreate successful search and analysis patterns. The Reactivity project focuses on the fundamental problem of how to capture researchers' work processes in a form that enables them to observe, reflect upon and improve their own future activity. We must identify the appropriate level of data capture and create sophisticated tools to log and store records of user activity, including their interactions with the physical world and across computer platforms. We must also develop efficient algorithms for visualizing the resulting multi-media temporal data and develop interactive applications that allow scientists to explore, reuse and improve successful strategies. We use participatory design to identify real-world work patterns and needs and we evaluate our work via benchmarks.and field tests with practicing researchers.

## TEAM

| | | | |
|---|---|---|---|
| **Team leader** | **FEKETE** | **Jean-Daniel** | ***INRIA Saclay-Île-de-France*** |
| **Team leader** | **MACKAY** | **Wendy** | ***INRIA Saclay-Île-de-France*** |
| Researcher | BEAUDOUIN-LAFON | Michel | *Université Paris 11* |
| Researcher | BONGSHIN | Lee | *Microsoft Research Redmond* |
| Researcher | CHAPUIS | Olivier | *CNRS* |
| Researcher | CZERWINSKI | Mary | *Microsoft Research Redmond* |
| Researcher | DRAGICEVIC | Pierre | *INRIA Saclay-Île-de-France* |
| Researcher | FISHER | Danyel | *Microsoft Research Redmond* |

**Wendy Mackay** received her Ph.D. from the Massachusetts Institute of Technology in 1990. She joined INRIA in 2000 where she is senior researcher and heads the in situ research team. Her research centers on mixed reality and the participatory design of innovative interactive technologies. A former chair of ACM/SIGCHI, she is active on numerous program committees and is a member of the editorial board for the ACM/Transactions on Human-Computer Interaction. She has authored over 100 research articles in refereed international conferences and journals in the field of human-computer interaction.

**Jean-Daniel Fekete** received his PhD from the University of Paris 11 in 1996. He headed the «Interactive Design and Modeling» group of the Ecole des Mines in Nantes in 2000, was invited at the University of Maryland at College Park in 2001-2002. He is a researcher at INRIA since 2002, where he heads the AVIZ Research team since 2007. His research interests include Human-Computer Interaction, Information Visualization and Visual Analytics. He authored 70 articles in journals and conferences. He is editor of the International Journal of Human-Computer Studies.

## RESEARCH

ReActivity is an abbreviation for "Reflecting on Activity". The goal is to help researchers *capture*, *visualize* and *interact* with their own and their colleagues' activities, over time. We are interested in increasing researchers' awareness of their activities, so as to improve not only productivity, but also their creativity and understanding.

Although based in the domain of e-Science, the project addresses the more general problem of how to help sophisticated users handle increasingly large quantities of heterogeneous data, over different hardware and software platforms. ReActivity is a three-way collaboration between AVIZ (INRIA), INSITU (INRIA) and VIBE (Microsoft Research, Redmond). The ReActivity project builds upon INRIA's and Microsoft Research's experience developing tools for biologists, historians, Wikipedians and software developers. Each of these user communities faces major challenges, not only in dealing with increasingly large quantities of heterogeneous, temporal data, but also because they must manage this data with a mix of open-source and commercial software across a range of hardware platforms.

The challenge is to provide capture and interactive visualization tools that enable users to understand their own activities within this complex environment. The benefit of the collaboration between INRIA and Microsoft is that, like our users, we cannot rely on a single approach, but must develop solutions that take into account the inherent heterogeneity of the hardware and software environments we use. Our approach is to develop common data models and tools that work across platforms, to provide both generic support for awareness when possible, but also specialized support within particular scientific domains when needed.

ReActivity has followed two research paths: one focused on Wikipedia and the other on Biology.

### WIKIPEDIA

The AVIZ group has focused on developing tools to support awareness for Wikipedia active members (Wikipedians) as well as casual Wikipedia users. We consider Wikipedia as an advanced social media for gathering and maintaining knowledge in general and believe that the Sciences will rely more and more on similar organizations

| | | | |
|---|---|---|---|
| Researcher | FISHER | Danyel | *Microsoft Research Redmond* |
| Researcher | MEYERS | Brian | *Microsoft Research Redmond* |
| Researcher | PIETRIGA | Emmanuel | *INRIA Saclay-Île-de-France* |
| Researcher | ROBERTSON | Georges | *Microsoft Research Redmond* |
| Researcher | SMITH | Greg | *Microsoft Research Redmond* |
| Post doc | BOUKHELIFA | Nadia | *MSR-INRIA Joint Centre* |
| Post doc | CHEVALIER | Fanny | *MSR-INRIA Joint Centre* |
| Post doc | ELMQVIST | Niklas | *MSR-INRIA Joint Centre* |
| Post doc | HENRY-RICHE | Nathalie | *MSR-INRIA Joint Centre* |
| Post doc | LETONDAL | Catherine | *MSR-INRIA Joint Centre* |
| Post doc | LI | Xiulum | *MSR-INRIA Joint Centre* |
| Post doc | LICCARDI | Ilaria | *MSR-INRIA Joint Centre* |
| Post doc | MOSCOVICH | Tomer | *MSR-INRIA Joint Centre* |
| Post doc | TSANDILAS | Theophanis | *MSR-INRIA Joint Centre* |
| PHD student | MASSON | Nicolas | *Université Paris 11* |
| PHD student | TABARD | Aurelien | *Université Paris 11* |

**PhD students who have graduated:** *Nathalie Henry-Riche, Univ. Paris-Sud*

**Onward destinations of post docs and PhDs who have left:** *Niklas Elmqvist is Assistant Professor at Purdue University, Nathalie Henry-Riche is a Researcher at MSR, Redmond, Fanny Chevalier is post-doc at OCAD, Toronto, Tomer Moscovich is researcher at Lab126*

**Visitors:** *Scott Klemmer (Stanford University, USA) in 2008. Jérémy Fry, M.C. Schraefel, Paul Andre and Max Wilson from the University of Southampton, UK in 2008. David Karger and Max Von Kleek from MIT, USA, in 2008. James Hollan, Gaston Cangiano and Adam Fouse from University California San Diego, USA, in 2008. Chris Maloney (Brown University, USA), Petra Isenberg (University of Calgary, Canada) and Anastasia Bezerianos (NICTA, Australia) in 2008. Sheelagh Carpendale (University of Calgary, Canada) in 2010.*

in the future: collaborative, non-hierarchical, aiming at complete coverage more than heavily curated contents. The main challenge of these kinds of knowledge gathering and organization systems is and will be to improve the quality level of the information while keeping up with the coverage. The ReActivity project aims at designing and experimenting tools to address this challenge.

The more concrete problem we are addressing is that Wikipedians — heavy Wikipedia contributors — have to split their time between:

1. writing and improving articles,
2. fixing issues in articles caused by changes made by occasional contributors, and
3. performing administrative tasks (e.g. renaming articles or resolving disagreements between two or more contributors.)

The amount of time spent on fixing articles is growing due to the popularity of Wikipedia, leaving Wikipedians with little or no time to work on improvements. Our goal is to explore how to help them improve the quality of Wikipedia by spending less time fixing articles and having power tools for writing better articles.

We have been studying the problems of Wikipedians, their needs for awareness and possible solutions through several participatory design sessions attended by about 30 Wikipedians overall in the past years.

Supporting awareness requires having mechanisms to visualize important information about changes in Wikipedia for monitoring and quality checking, and these visualizations should be included in the standard applications. In addition, these mechanisms should present information about roles of contributors (e.g. this person mainly worked on this specific article, or a small set, or the whole project, or for several other projects, etc.) and their maturity and reliability. From participatory design sessions, we realized this information is either not available at all and needs to be computed (in term of aggregated values computed from user contributions) or available but requiring extensive navigation to access using the current Wikipedia interface. Therefore, we have built the WikiReactive infrastructure to compute the missing information by creating a local copy of the French Wikipedia and we also started to design interfaces to be evaluated with Wikipedia contributors.

Creating a local repository and computing additional information has turned to be very resource consuming due to the size of Wikipedia (1Tb for the French version) and the numerous changes that occurred at each version of Wikipedia pages. We now have a running base that is used for our tools.

## THE WIKIREACTIVE INFRASTRUCTURE

According to our participatory design sessions, Wikipedians need aggregated information to quickly assess the quality of articles or users and spend less time hunting for information than using long series of navigations as in the current Wikipedia interface.

Wikipedia provides a Web API to collect raw information about each revision and each user but collecting all the contributions of a user to compute its quality is not feasible in interactive time. Thus, we have built an infrastructure called WikiReactive ([156], *http://www.aviz.fr/wikireactive/*, and *Figure 1*) that computes aggregated information on Wikipedia and maintain them in real-time while Wikipedia evolves.

WikiReactive is usable as a Web service and can be used by high-level tools such as Diffamation, by Dashboard visualizations such as WikipediaViz or to collect statistics on the activity of Wikipedia. It is currently deployed on the French Wikipedia, serving about 1 million pages with an average of 10 revisions per minute. The database is about 700 GB large and continues to grow.

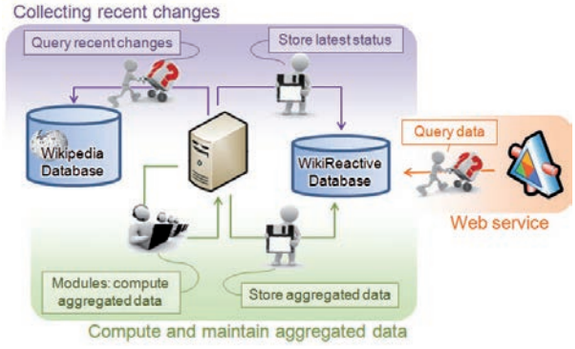



Figure 1: The WikiReactive architecture

## HIGH-LEVEL TOOLS

In addition to the need for a low-level infrastructure that gives Wikipedians direct access to Wikipedia data, higher-level data exploration tools need to be designed to raise their productivity. We have designed, implemented and evaluated two such tools: Diffamation and iChase.

**Diffamation** ([157], *http://www.aviz.fr/diffamation/*, and *Figure 2*) is an animation technique for smoothly transitioning between different text revisions. Diffamation allows rapid exploration of revision histories by combining

text animated transitions with simple navigation and visualization tools.

We performed a controlled user study showing that smooth text animation allows users to track changes in the evolution of textual documents more effectively than flipping pages, allowing Wikipedians to quickly understand the evolution of articles in an entertaining manner.

**iChase** ([163] and *Figure 3*) is a novel interactive visualization tool to provide Wikipedians with better awareness of editing activities on Wikipedia. Unlike the currently used visualizations that provide only page-centric information, iChase visualizes the trend of activities for two entity types, articles and contributors.

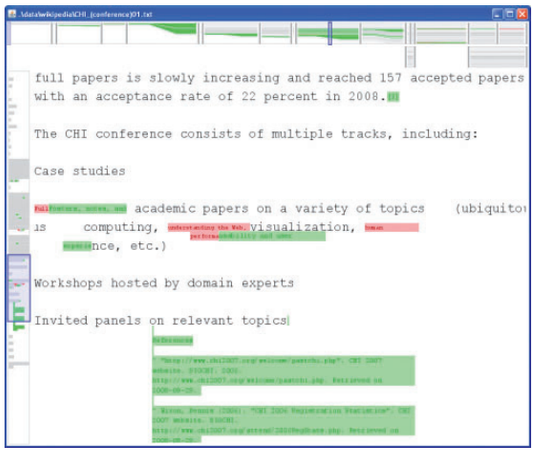



Figure 2: The Diffamation system showing the evolution of the article CHI (Conference) on Wikipedia

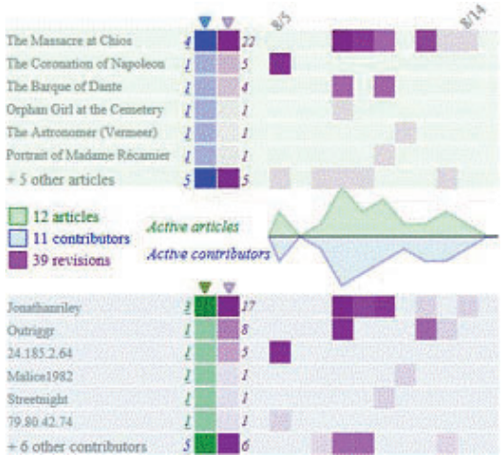



Figure 3: The iChase system showing the activity of several articles watched by a typical Wikipedian, along with the list of contributors on these articles

## EVENTS, WORKSHOPS, CONFERENCES, SEMINARS

- Members of AVIZ and INSITU went to Seattle in January for a 3-day kickoff meeting with members of the VIBE team. Members of AVIZ, INSITU and VIBE also met at the CHI'08 conference in Italy, at the CSCW program committee in San Diego, and at the UIST and CSCW conferences in Monterey and in San Diego. Wendy Mackay, Jean-Daniel Fekete and Michel Beaudouin-Lafon also presented the ReActivity project to the Steering Committee.

- The first ReActivity workshop, entitled Capture, *Visualization and Display of Temporal Data* was held over 3 days in Paris (3-5 June). Co-sponsored by Tony Hey at Microsoft, the workshop attracted researchers from the three ReActivity teams, as well as MIT, University of California, San Diego, Brown University, University of Southampton and Cambridge University, plus additional researchers from other parts of Microsoft. We chose participants with a range of perspectives and complementary interests, and were pleased to see that it resulted in several new research collaborations. Specifically, INSITU just received word that Sirius, our INRIA Associate Lab proposal with UCSD and Stanford, was accepted and will enable us to reinforce our work on the WILD wall-sized display (see News below) and also work on interactive paper. Aurélien Tabard met with researchers at MIT in Boston between the UIST and CSCW conferences to work on the workshop for CHI'09 (below) and to explore time-line based data analysis.

- We submitted a follow-on workshop proposal to CHI'09, which was accepted, and will be held at the CHI'09 conference in Boston (see *http://temporal.csail.mit.edu/*). So far, 34 people, in addition to members from AVIZ, INSITU and VIBE, have registered for the workshop, indicating a great deal of interest in the topic. The goals of the second workshop are to develop new research collaborations with other researchers and to put together a book or special issue of a journal, with original research on the topic.

- Members of AVIZ and VIBE participated in a one-day workshop with Wikipedia senior contributors on the management of Wikipedia Projects in Paris.

## TOOLS AND SOFTWARE

- ScatterDice (*https://engineering.purdue.edu/~elm/projects/scatterdice.html*)
- GraphDice (*http://www.aviz.fr/graphdice*)
- Diffamation (*http://www.aviz.fr/diffamation*)
- WikipediaViz (*http://www.aviz.fr/wikipediaviz*)
- WikiReactive (*http://www.aviz.fr/wikireactive*)

## DASHBOARD VISUALIZATION FOR AWARENESS

While high-level visualization tools are useful to explore in depth some pages and users' activity, they are not meant to be used all the time by Wikipedians since they use too much screen real-estate and require a lot of attention. When working on Wikipedia, most of the attention of Wikipedians should be devoted to articles and text. Therefore, we have designed and implemented several small visualizations, meant to be integrated into the Wikipedia web page or on the Desktop on the border of the screen. These visualizations are meant to convey focused information to allow Wikipedians to quickly assess the context of a page and raise their awareness in case of strange patterns.

**WikipediaViz** ([159, 158], *http://www.aviz.fr/wikipediaviz/*, and *Figure 4*) uses some of these Dashboard visualizations to reveal the profile of an article in term of contributors, evolution and structure. An evaluation of WikipediaViz showed that users could assess the quality of articles faster using the visualization than without while not using much screen real-estate.

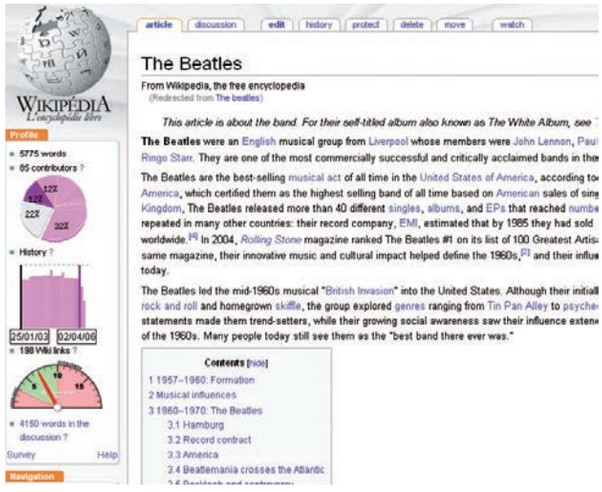



Figure 4: WikipediaViz showing three dashboard visualizations on the left

## CONCLUSION

Providing awareness tools for Wikipedia involves several complementary steps:

- Understanding the high-level questions of Wikipedians, their process and the current bottlenecks,
- Designing and implementing solutions to address the most important issues. Some of the solutions involve Wikipedians oriented applications or Web-page with mixed contents (mashups). Others also require more complex information system with real-time computation of complex measures on the whole dynamic Wikipedia datasets. We have implemented both kinds of solutions.
- Evaluating the impact of the results to assess how the tools improve the process. We have performed several evaluations at the tool level but we have not been able to conduct them at the whole-system level due to time constraints and this is something that needs to be addressed in the future. ■

### LIST OF SCIENTIFIC VISITORS FROM ACADEMIA

- Prof. Scott Klemmer from Stanford University (June 2008)
- Prof. Jeremy Fry, Prof. M.C. Schraefel, Paul Andre and Max Wilson from University Southampton (June 2008)
- Prof. David Karger and Max van Kleek from MIR (June 2008)
- Prof. James Hollan, Gaston Cangiano and Adam Fouse from University of San Diego (June 2008)
- Petra Isenberg from Univ. of Calgary, Canada (May 2008)
- Anastasia Bezerianos from NICTA, Australia (May 2008)
- Prof. Sheelagh Carpendale from University of Calgary (June 2010) Press

### PRESS

- Wendy Mackay was interviewed on radio, including France Bleu and France Inter.
- Wendy Mackay was interviewed for "2.1 Regards sur le numérique."
- Wendy Mackay was a panelist on the VIA Round Table on "Interfaces"
- Nathalie Henry was interviewed on "OKAPI Phosphore" (*http://www.inria.fr/actualites/2008/pdf/inria-okapi.pdf*)

# PUBLICATIONS & TALKS

## JOURNAL PAPERS AND BOOK CHAPTERS

[149] ANASTASIA BEZERIANOS, FANNY CHEVALIER, PIERRE DRAGICEVIC, NIKLAS ELMQVIST, AND JEAN-DANIEL FEKETE. GraphDice: A system for exploring multivariate social networks. Comput. Graph. Forum, 29(3):863–872, 2010.

[150] ANASTASIA BEZERIANOS, PIERRE DRAGICEVIC, JEAN-DANIEL FEKETE, JUHEE BAE, AND BEN WATSON. GeneaQuilts: A system for exploring large genealogies. IEEE Trans. Vis. Comput. Graph., 16(6):1073–1081, 2010.

[151] NIKLAS ELMQVIST AND JEAN-DANIEL FEKETE. Hierarchical aggregation for information visualization: Overview, techniques, and design guidelines. IEEE Trans. Vis. Comput. Graph., 16(3):439–454, 2010.

[152] NIKLAS ELMQVIST, YANN RICHE, NATHALIE HENRY RICHE, AND JEAN-DANIEL FEKETE. Melange: Space folding for visual exploration. IEEE Trans. Vis. Comput. Graph., 16(3):468–483, 2010.

[153] PETRA ISENBERG, ANASTASIA BEZERIANOS, NATHALIE HENRY, M. SHEELAGH T. CARPENDALE, AND JEAN-DANIEL FEKETE. CoCoNutTrix: Collaborative retrofitting for information visualization. IEEE Computer Graphics and Applications, 29(5):44–57, 2009.

[154] NATHALIE HENRY-RICHE AND JEAN-DANIEL FEKETE. Novel visualizations and interactions for social networks exploration. In Borko Furht, editor, Handbook of Social Network Technologies and Applications, pages 611–636. Springer, 2010.

## CONFERENCE AND WORKSHOP PAPERS

[155] ASTERIOS KATSIFODIMOS, JEAN-DANIEL FEKETE, ALAIN CADY, AND CECILE GERMAIN RENAUD. Visualizing the dynamics of e-science social networks. In EGEE User Forum, Uppsala Sweden, April 2010.

[156] NADIA BOUKHELIFA, FANNY CHEVALIER, AND JEAN-DANIEL FEKETE. Real-time aggregation of Wikipedia data for visual analytics. In Proceedings of Visual Analytics Science and Technology (VAST 2010), pages 147–154, Salt Lake City, UT, USA, 2010. IEEE Computer Society.

[157] FANNY CHEVALIER, PIERRE DRAGICEVIC, ANASTASIA BEZERIANOS, AND JEAN-DANIEL FEKETE. Using text animated transitions to support navigation in document histories. In Proceedings of the 28th International Conference on Human Factors in Computing Systems, CHI 2010, Atlanta, Georgia, USA, April 10-15, 2010, pages 683–692. ACM, 2010.

[158] FANNY CHEVALIER, STÉPHANE HUOT, AND JEAN-DANIEL FEKETE. Visualisation de mesures agréegées pour l'estimation de la qualité des articles Wikipedia. In Extraction et gestion des connaissances (EGC'2010), Actes, 26 au 29 janvier 2010, Hammamet, Tunisie, volume RNTI-E-19 of Revue des Nouvelles Technologies de l'Information, pages 351–362. Cépaduès-Éditions, 2010.

[159] FANNY CHEVALIER, STÉPHANE HUOT, AND JEAN-DANIEL FEKETE. WikipediaViz: Conveying article quality for casual Wikipedia readers. In IEEE Pacific Visualization Symposium PacificVis 2010, Taipei, Taiwan, March 2-5, 2010, pages 49–56. IEEE, 2010.

[160] JEAN-DANIEL FEKETE, NIKLAS ELMQVIST, AND YVES GUIARD. Motion-pointing: target selection using elliptical motions. In Proceedings of the 27th International Conference on Human Factors in Computing Systems, CHI 2009, Boston, MA, USA, April 4-9, 2009, pages 289–298. ACM, 2009.

[161] TOMER MOSCOVICH. Contact area interaction with sliding widgets. In Proceedings of the 22nd Annual ACM Symposium on User Interface Software and Technology, Victoria, BC, Canada, October 4-7, 2009, pages 13–22, 2009.

[162] TOMER MOSCOVICH, FANNY CHEVALIER, NATHALIE HENRY, EMMANUEL PIETRIGA, AND JEAN-DANIEL FEKETE. Topology-aware navigation in large networks. In Proceedings of the 27th International Conference on Human Factors in Computing Systems, CHI 2009, Boston, MA, USA, April 4-9, 2009, pages 2319–2328. ACM, 2009.

[163] NATHALIE HENRY-RICHE, BONGSHIN LEE, AND FANNY CHEVALIER. iChase: supporting exploration and awareness of editing activities on Wikipedia. In Proceedings of the International Conference on Advanced Visual Interfaces, AVI 2010, Roma, Italy, May 26-28, 2010, pages 59–66, 2010.

## TALKS

[164] JEAN-DANIEL FEKETE. Advanced interaction for information visualization. Keynote presentation at CHItaly conference in Rome, June 18th, 2009.

[165 JEAN-DANIEL FEKETE. Visualizing networks using adjacency matrices: Progresses and challenges. Keynote presentation at IEEE CAD/Graphics2009, Aug. 20, Yellow Mountains, China, 2009.

[166] JEAN-DANIEL FEKETE. Advanced interaction for information visualization. Keynote presentation at PacificVis, Taipei, Taiwan, 2010.

**Track B** **This project started at fall 2007.**

# ADAPTIVE COMBINATORIAL SEARCH FOR E-SCIENCES

## OVERVIEW

> **The goal of this research is to improve the applicability of constraint-based or heuristic based solvers to computational sciences.**

As demonstrated by a large literature, scientists already benefit from the use of search procedures to tackle a large variety of important problems. For instance, in experimental-design simulation and inference, data interpretation, etc. Unfortunately, these applications suffer from the limits of current solving technologies which appear to be poorly adapted to these new domains.

One solution to increase performance is the fine tuning of the solver parameters. This is a tedious and time-consuming task that often requires knowledge about both the domain and the algorithm of interest. This approach is hardly applicable to computational sciences whose applications fields are constantly growing. We claim that the self-adaptation of a solver to the domain of interest is the only viable solution to this problem. Our goal in this project is to develop tools able to automatically choose the optimal parameter configuration of a given search algorithm. This adaptation would allow us to deploy search techniques on new computational sciences grounds with good expected performance.

## TEAM

| | | | |
|---|---|---|---|
| **Team leader** | **SCHOENAUER** | **Marc** | ***INRIA Saclay-Île-de-France*** |
| **Team leader** | **HAMADI** | **Youssef** | ***Microsoft Research Cambridge*** |
| Researcher | AUGER | Anne | *INRIA Saclay-Île-de-France* |
| Research Software Developer | BORDEAUX | Lucas | *Microsoft Research Cambridge* |
| Researcher | SEBAG | Michèle | *CNRS* |
| Post doc | BOUKHELIFA | Nadia | *INRIA Saclay-Île-de-France* |
| Post-doc | HANSEN | Niklaus | *MSR-INRIA Joint Centre* |
| Post-doc | JABBOUR | Said | *INRIA Saclay-Île-de-France* |
| PHD student | ARBALAEZ | Alexandro | *Université Paris 12* |
| PHD student | FIALHO | Alvaro | *Université Paris 11* |

**Marc Schoenauer** graduated from Ecole Normale Supérieure in Paris and received aPhD in applied mathematics at the University of Paris 6 in 1980. Marc Schoenauer was researcher at CNRS-CMAP (Ecole Polytechnique) and is now senior researcher at INRIA. He has been working in the area of Evolutionary Computation since 1990. He is author of 60 publications in journals and international conferences.

**Youssef Hamadi** holds a PhD from the University of Montpellier 2. He joined Microsoft Research Cambridge in 2003, where he is currently leading the Constraint Reasoning group. His research interests include Constraint Programming and SAT solving. His work has been applied to various domains like software verification, computer graphics and smart workflow engines.

## RESEARCH

Many forefront techniques in both Stochastic and Combinatorial Search have been very successful in solving difficult real-world problems. However, their application to newly encountered problems, or even to new instances of known problems, remains a challenge, even for experienced researchers of the field - not to mention newcomers, be they skilled scientists or engineers from other areas. Theory and/or practical tools are still missing to make them 'Crossing the Chasm' (from Geoffrey A. Moore's 1991 book about the Diffusion of Innovation). The difficulties faced by the users arise mainly from the significant range of algorithm and/or parameter choices involved when using this type of approaches, and the lack of guidance as to how to proceed for selecting them. Moreover, state-of-the-art approaches for real-world problems tend to represent bespoke problem-specific methods which are expensive to develop and maintain.

One longer-term goal, that could be helpful for all of the on-going work described below, is the design of accurate descriptors that would allow the user to accurately describe a given problem (or instance). From thereon, it would be possible to learn from extensive experiments what are the good algorithms/parameters for classes of instances, or even individual instances, like has been done in the SAT domain by Y. Hamadi and co-authors (F. Hutter, Y. Hamadi, H.H. Hoos, and K. Leyton-Brown, Performance Prediction and Automated Tuning of Randomized and Parametric Algorithms, *Constaint Programming 2006*).

### ADAPTIVE OPERATOR SELECTION

Adaptive Operator Selection (AOS) is concerned with the on-line adaptation of the mechanism that chooses among the different variation operators in Evolutionary Algorithms, a series of approaches have been proposed durint Alvaro Fialho's PhD, defended in December 2010 [233]. All are based on the original idea of using Multi-Armed Bandit (MAB) algorithms: each operator is viewed as one arm of a MAB problem. This approach was initially introduced in 2008 (DaCosta, Fialho, Schoenauer, and Sebag. Adaptive Operator Selection with Dynamic Multi-Armed Bandits. In C. Ryan et al., eds.: ACM-GECCO, ACM Press, p. 913-920, 2008), and continuously improved and extended ever since, in particular with respect to the Credit Assignement procedure, for which several procedures were successively designed. Note that a survey paper on AOS techniques has recently been written with our colleagues from Angers [173].

- A first idea was to use Extreme values rather than averages as a reward for operators: It has been advocated in many domains that extremely rare but extremely beneficial events can be much more consequential than average good events. This has been first validated on the OneMax problem, where the optimal strategy for a given fitness level is known (PPSN'08 paper), and later on k-path and Royal Road problems [204] (Best Paper Award at LION'09 conference in Trento, January 09), and thoroughly compared to other state-of-the-art Operator Selection procedures [207].
- MAB ideas for Operator Selection were recombined with the Compass approach of our colleagues from Angers University (Maturana and Saubion. A Compass to Guide Genetic Algorithms. PPSN 2008: 256–265) in order to mix fitness improvements and diversity in the Credit Assignement [225].
- The idea of using a sliding window for assessing the dynamic characteristic of the operator selection problem was introduced in [177] and compared with previous approaches. However, at that time the best technique to-date had bot yet been proposed: the comparison-based reward, based on the Machine-Learning idea of the Area Under the Curve (AUC) led to a much more robust AOS by preserving the invariance with respect to monotonous transformations of the fitness whenever the underlying algorithm was comparison-based. It was first investigated in the binary framework [209], and was later demonstrated to apply to a much wider context than pure EA, such as Differential Evolution [206, 208], following up a collaboration with colleagues from Wuhan Geoscience University [211, 205]. A journal version of this on-going work has been recently submitted [178].

### ADAPTATION FOR CONTINUOUS OPTIMIZATION

Building on the well-known Covariance Matrix Adaptation Evolution Strategy (CMA-ES) algorithm, that adapts the covariance matrix of the Gaussian mutation of an Evolution Strategy based on the path followed by the evolution, several improvements and generalizations have been proposed. A general overview of CMA-ES present state, as well as a brief description for most of the recent developments can be found in the HDR dissertation of CMA-ES main author, Nikolaus Hansen [234], defended in February 2010.

- The underlying principles of the CMA-ES way have been carefully analyzed in [172], while theoretical principles of ES at large have been gathered in [168] (part of a book co-edited by Anne Auger summarizing the recent theoretical advances in the theory of stochastic search [169]). In particular, the benefits of the rotational invariance of CMA-ES are becoming very clear, even more so after their generalization to any search heuristic through "Adaptive Encoding" (see the seminal paper by N. Hansen, *Adaptive Encoding: How to Render Search Coordinate System Invariant*. In G. Rudolph et al., Eds, Proc. PPSN X, LNCS, Vol. 5199, p. 205-214. Springer Verlag 2008). This can be of immense practical use for all algorithms that actually take advantage of the separability of the objective function, as the well-known Particle Swarm Optimization (PSO) [184].

- A new variant of CMA-ES has been proposed [201]. It is based on mirrored sampling and sequential selection, and was demonstrated to outperform the standard CMA-ES on most of BBOB-comparisons [240, 241, 242, 243, 244, 246, 247, 248, 249, 238, 239, 245].

- In the specific context of multi-objective optimization, the multi-objective variant of CMA-ES has been improved [230, 229] by adding more information exchange among the solutions that are spread along the Pareto front. This algorithm has also been applied to solve a parameter identification problem for a system of differential equations describing the dynamics of a Gene Regulatory Network for the drosophila embryo [188] (paper submitted, in collaboration our partners from the NonLinear Dynamics Group at IST, Lisboa in the framework of the GENNETEC European project). Recently, the introduction of a non-uniform selection in the steady-state variant resulted in performance improvement on classical benchmark functions [224].

- Several applications have been tackled: within the ANR OMD project, the optimization of a complete launcher has been addressed by CMA-ES in collaboration with EADS [176, 203]. The on-line optimization of the parameters of H∞ feedback controllers of thermo-acoustic instabilities of gas turbine combustors has been addressed using a specific version of CMA-ES introducing a generic way to handle uncertainty [183].

- A great deal of effort has been devoted to the benchmarking of optimizers for continuous problems. Though not directly linked to parameter tuning, it is an important step toward enabling a sound comparison of different parameter settings, thus impacting any work regarding parameter setting. The powerful and flexible platform COCO (*http://coco.gforge.inria.fr/*)has been developed, allows an easy test and comparison of any continuous optimizer on a test-bench of 24 base functions and their noisy versions (see Section *Tools and software*). Furthermore, two workshops have been organized during the ACM-GECCO conference in 2009 and 2010 where more than 30 entries could be compared [259, 260], including classical methods that are being used daily by applied mathematicians, like BFGS (Matlab version), Fletcher-Reeves, NEWUOA, …[270, 271, 272, 273] and now-standard stochastic methods frequently used in the EC community [236, 237, 252, 253, 255, 254]. Furthermore, independently of the BBOB workshops, COCO has been used to assess Adaptive Operator Selection methods when it has been applied to continuous optimizers [206, 208, 205, 211, 256, 220].

- A new thread of research is concerned with using surrogate models in the case of expensive fitness functions. In the single-objective context, previous proposal for local quadratic models have been greatly improved in the case of large populations [199], and has been successfully applied to a real-world problem [200]. A global approach that preserves the comparison-based property of CMA-ES has also been proposed, using rank-based SVM as a surrogate [222]. In the multi-objective framework, an original approach trying to build a single-objective surrogate model for multi-objective Pareto optimization using a specifically-tailored SVM algorithm has been introduced [221], and improved by the use of rank-based SVMs [223].

## AUTONOMOUS SEARCH WITH CONTEXT

Autonomous Search is a new emerging area in Constraint Programming motivated by the demonstrated importance of the application of Machine Learning techniques to the Algorithm Selection Problem, and with potential applications ranging from planning, configuring to scheduling. This area aims at developing automatic tools to improve the performance of search algorithms to solve combinatorial problems, e.g., selecting the best parameter settings for a constraint solver to solve a particular problem instance. During A. Arbelaez' PhD (to be defended in May 2011 [232]), three different points of view are studied to automatically solve combinatorial problems: Constraint Satisfaction Problems, Constraint Optimization Problems, and SAT problems.

- DomFD is a new Variable Selection Heuristic whose objective is to heuristically compute a simplified form of functional dependencies called weak dependencies. These week dependencies are then used to guide the search at each decision point [186]. Furthermore, this new technique has been embedded in the Open Source constraint solver Gecode. It is now freely accessible to the community (see Section *Tools and software*).
- The Algorithm Selection Problem has been studied from two different angles. On the one hand, a traditional portfolio algorithm is used to offline learn a heuristics model for the Protein Structure Prediction Problem [190]. On the other hand, we present the Continuous Search paradigm, which objective is to allow any user to eventually get their constraint solver achieving a top performance on their problems [191]. Continuous Search comes in two modes: the functioning mode solves the user's problem instances using the current heuristics model; the exploration mode reuses these instances to training and improve the heuristics model through Machine Learning during the computer idle time. The most recent work along this line of research is summarized in a book chapter [167].
- Finally, recent work as part of Alejandro Arbelaez' thesis considers the question of adding a knowledge-sharing layer to current portfolio-based parallel local search solvers for SAT [187]. It has been shown that the overall performance can be greatly improved by sharing the best configuration of each candidate in the parallel portfolio on regular basis and aggregating this information in special ways [189] (an extended journal version is under submission [175]). The result of this work is the Cooperative Stochastic Parallel Local Search solver which is currently engaged in the 2011 SAT-Competition.

## SAT, SMT, AND PARALLELISM

This line of work has been published in several journals [182, 180, 181, 179] and conferences [212, 228, 231, 226, 196, 213]. Some of this works made it through the last SAT Races and Competitions, especially in the parallel tracks. Overall our approach won several tracks since 2010 (several gold, and silver medals). Our parallel portfolio-based is now state-the-art in parallel SAT, and all the solvers presented during the 2010 SAT Race are based on this approach [195, 219, 227]. ■

### MICROSOFT BENEFITS

**Microsoft Solver Foundation (MSF)** is designed to help businesses make optimal strategic decisions. The possible applications cover a vast range: real-time supply chain optimization, data center energy profile management, on-line advertising profit maximization, logistics of large conference scheduling, transportation network flows, and risk analysis of investment portfolios. There are also direct applications to graphics and machine learning. All these problems are NP-difficult and MSF allows the programmer to choose between different solvers in order to quickly compute optimal or approximate solutions for their applications.

Unfortunately, if MSF provides several solvers, it does not provide the expertise required to decide between them or even to pick up the right parameters that should be used with a particular algorithm. In other words, it does not guarantee that a standard programmer will reach an acceptable level of performance for the resolution of its problem.

In this context, the "Adapt" project which aims at improving the usability of modern optimization methods (targeting e-Scientists users) clearly meets the actual limitations of MSF. We are sharing these results with the MSF SD Leads, and we believe that the the results of our project will greatly influence the future versions of the Microsoft Solver Foundation line of products.

## TOOLS AND SOFTWARE

- **The Covariance Matrix Adaptation Evolution Strategy (CMA-ES)** is considered state-of-the-art in continuous domain evolutionary computation. (See H.-G. Beyer (2007). *Evolution Strategies, Scholarpedia*, p. 1965.) It has been shown to be highly competitive on different problem classes. The algorithm is widely used in research and industry as witnessed by more than a hundred published applications. We provide source code for the CMA-ES in C, Java, Matlab, Octave, Python, and Scilab [170, 171] including the latest variants of the algorithm [192, 185, 201] (see Section *Adaptation for Continuous Optimization*).
  Check all details and download source code at *http://www.lri.fr/~hansen/cmaes_inmatlab.html*.
- **Extension to the Gecode Constraint Solver** (open source) to implement the new DomFD, a new variable ordering heuristic which heuristically discovers a simplified form of functional dependencies. Freely accessible to the community
  Link: *http://www.msr-inria.inria.fr/~arbelaez/domFD/domFD.html*
- **COCO** (COmparing Continuous Optimizers, *http://coco.gforge.inria.fr/*) is a platform for systematic and sound comparisons of real-parameter global optimizers. COCO provides benchmark function testbeds and tools for processing and visualizing data generated by one or several optimizers [267]. Several classes of benchmark functions have been thoroughly designed, including noiseless [268, 257] and noisy [269, 258] functions, where great care has been taken to perturb the well-known test functions (e.g. moving the optimum around, rotating the coordinate system, ...). Several post-processing procedures have also been defined, with standard outputs leading to fair and sound comparisons between different optimizers. An API allows the users to easily interface their own optimizer with the test set, and the graphics and tables [250, 251] that are automatically generated give instantly a clear picture of pairwise or more global comparisons. Further work include the handling of constraints and that of mixed-integer problems. However, such extension also require the definition of both new performance measures and new sets of benchmark functions.

## EVENTS, WORKSHOPS, CONFERENCES, SEMINARS

- **Best Paper Award** at LION'09 conference in January 2009 for [204].
- **Best Paper Award** at the EvoBIO'09 conference in April 2009 for [188] (paper with colleagues from European project GENNETEC).
- **Best Paper Award** at the ICTAI'09 conference in November 2009 for [214].
- Two **BBOB workshops** (Black-Box Optimization Benchmarking) were mainly co-organized by Anne Auger, Nikolaus Hansen, and Raymond Ros during the 2009 and 2010 editions of ACM-GECCO (Genetic and Evolutionary Computation COnference), the main conference in the Evolutionary Computation domain. Both workshops gathered more than 30 entries to the comparison challenge: all participants submitted the best version of their favorite continuous optimizer that were globally compared on the 24 noiseless functions and/or on their noisy counterparts – including entries for the classical optimizers (see Section *Adaptation for Continuous Optimization*). Global analyses of the results have been published after the workshops [259, 260].
- A follow-up of the two BBOB workshops is the **Special Issue** of *Evolutionary Computation* journal around benchmarking issues in continuous optimization. Guest Editors are Anne Auger, Nikolaus Hansen, and Marc Schoenauer. Twelve papers are presently being reviewed, publication is planned for end 2011.
- Marc Schoenauer has co-organized with Gabriela Ochoa and Darrell Whitley the **Self-* workshop** at PPSN conference (Parallel Problem Solving from Nature, the oldest European event on Evolutionary Computation) in Krakow, September 2010. This workshop was dedicated to "self-tuning, self-configuring and self-generating search heuristics". See *http://www.cs.nott.ac.uk/~gxo/ppsn2010-selfstar.html*.
- Marc Schoenauer has co-organized with Gabriela Ochoa the **Self-* Special Session** at LION-5 conference in Rome (January 2011), dedicated to "Self-Tuning, Self-Configuring and Self-Generating Evolutionary Algorithms". See the conference Web site at *http://www.intelligent-optimization.org/LION5/*. Post-proceedings of the conference will be published as a LNCS volume (Springer Verlag).
- A follow-up of the two Self-* events above is the **Special Issue** of *Evolutionary Computation* journal dedicated to parameter tuning in Evolutionary Computation at large. Guest Editors are Thomas Bartz-Beielstein , Gabriela Ochoa, Mike Preuss, and Marc Schoenauer. Eight papers have been submitted at the moment, publication is scheduled for mid-2012.
- Youssef Hamadi and Marc Schoenauer will be **co-chairing LION-6** conference in January 2012 in Paris, in Microsoft Conference Center as Issy-les-Moulineaux. See preliminary Web site at http://www.intelligent-optimization.org/LION6/.
- **Award:** Marc Schoenauer and his PhD student Jacques Bibaï, together with their co-authors Pierre Savéant and Vincent Vidal, were awarded the **Silver Medal** at the ACM-GECCO 2010 Humies Award, for their Human-Competitive results obtained using Evolutionary Computation in the domain of AI Planning [217] – including some automatic parameter tuning procedure [218].
- **Best paper award** for [192] in the ES-EP track at the ACM-GECCO 2010.

## PRESS

- Interview of Michèle Sebag by scientific magazine La Recherche about Watson performance in popular TV game Jeopardy!, March 2011.

# PUBLICATIONS & TALKS

## BOOKS


[167] ALEJANDRO ARBELAEZ, YOUSSEF HAMADI, AND MICHELE SÈBAG. Continuous Search in Constraint Programming. In Youssef Hamadi, Eric Monfroy, and Frédéric Saubion, editors, Autonomous Search. Springer-Verlag, 2011.

[168] A. AUGER AND N. HANSEN. Theory of evolution strategies: a new perspective. In Anne Auger and Benjamin Doerr, editors, Theory of Randomized Search Heuristics: Foundations and Recent Developments, pages 289–325. World Scientific Publishing, 2010. In press.

[169] ANNE AUGER AND BENJAMIN DOERR, editors. Theory of Randomized Seach Heuristics–Foundations and Recent Developments, volume 1 of Theoretical Computer Science. World Scientific, 2010.

[170] Y. COLETTE, NIKOLAUS HANSEN, AND G. PUJOL. Vers une programmation orientée objet des optimiseurs. In Opimisation multidisciplinaire en mécanique 2. Réduction de modèles, robustesse, fiabilité, réalisations logicielles, volume 2 of Méthodes Numériques en Mécanique, chapter 7. Hermes Science, Lavoisier, Paris, London, Chippenham, April 2009.

[171] YANN COLLETTE, NIKOLAUS HANSEN, GILLES PUJOL, DANIEL SALAZAR APONTE, AND RODOLPHE LE RICHE. On object-oriented programming of optimizers – examples in scilab. In P. Breitkopf and R. F. Coelho, editors, Multidisciplinary Design Optimization in Computational Mechanics, chapter 14, pages 527–565. Wiley, 2010.

[172] NIKOLAUS HANSEN AND ANNE AUGER. Principled design of continuous stochastic search in practice: the CMA evolution strategy. In Y. Borenstein and A. Moraglio, editors, Theory and Principled Methods for the Design of Metaheuristics. Springer, 2010. submitted.

[173] JORGE MATURANA, ALVARO FIALHO, FRÉDERIC SAUBION, MARC SCHOENAUER, FRÉDERIC LARDEUX, AND MICHÈLE SEBAG. Adaptive operator selection and management in evolutionary algorithms. In Hamadi, Y. et al, editor, Autonomous Search. Springer Verlag, 2010. (To appear).


## JOURNAL PAPERS AND BOOK CHAPTERS


[174] SARA ALOUF, GIOVANNI NEGLIA, IACOPO CARRERAS, DANIELE MIORANDI, AND ALVARO FIALHO. Fitting genetic algorithms to distributed on-line evolution of network protocols. Computer Networks, 54(18):3402 – 3420, December 2010.

[175] ALEJANDRO ARBELAEZ AND YOUSSEF HAMADI. Efficient Parallel Local Search for SAT. Submitted, 2011.

[176] GUILLAUME COLLANGE, STÉPHANE REYNAUD, AND NIKOLAUS HANSEN. Covariance matrix adaptation evolution strategy for multidisciplinary optimization of expendable launcher families. In 13th AIAA/ISSMO Multidisciplinary Analysis Optimization Conference, Proceedings, 2010.

[177] ALVARO FIALHO, LUIS DA COSTA, MARC SCHOENAUER, AND MICHÈLE SEBAG. Analyzing bandit-based adaptive operator selection mechanisms. Annals of Mathematics and Artificial Intelligence – Special Issue on Learning and Intelligent Optimization, September 2010.

[178 W. GONG, Á. FIALHO, Z. CAI, AND H. LI. Adaptive strategy selection in differential evolution for numerical optimization. Information Sciences, 2010. submitted.

[179] YOUSSEF HAMADI. Conclusion to the special issue on parallel sat solving. JSAT, 6(4):263, 2009.

[180] YOUSSEF HAMADI, SAÏD JABBOUR, AND LAKHDAR SAIS. Learning for dynamic subsumption. CoRR, abs/0904.0029, 2009.

[181] YOUSSEF HAMADI, SAÏD JABBOUR, AND LAKHDAR SAIS. Manysat: a parallel sat solver. JSAT, 6(4):245–262, 2009.

[182] YOUSSEF HAMADI, SAÏD JABBOUR, AND LAKHDAR SAIS. Learning for dynamic subsumption. International Journal on Artificial Intelligence Tools, 19(4):511–529, 2010.

[183] NIKOLAUS HANSEN, ANDRE NIEDERBERGER, LINO GUZZELLA, AND PETROS KOUMOUTSAKOS. A method for handling uncertainty in evolutionary optimization with an application to feedback control of combustion. IEEE Transactions on Evolutionary Computation, 13(1):180–197, 2009.

[184] NIKOLAUS HANSEN, RAYMOND ROS, NIKOLAS MAUNY, MARC SCHOENAUER, AND ANNE AUGER. Impacts of invariance in search: when CMA-ES and pso face ill-conditioned problems. Applied Soft Computing, page to be resubmitted with minor revisions, 2009. to be resubmitted with minor revisions.

[185] THORSTEN SUTTORP, NIKOLAUS HANSEN, AND CHRISTIAN IGEL. Efficient covariance matrix update for variable metric evolution strategies. Machine Learning, 72(2):167–197, 2009.


## CONFERENCE AND WORKSHOP PAPERS


[186] ALEJANDRO ARBELAEZ AND YOUSSEF HAMADI. Exploiting Weak Dependencies in Tree-based Search. In Proceedings of the 24th Annual ACM Symposium on Applied Computing, pages 1385–1391. ACM, ACM Press, 2009.

[187] ALEJANDRO ARBELAEZ, YOUSSEF HAMADI, AND MICHÈLE SEBAG. Online Heuristic Selection in Constraint Programing. In International Symposium on Combinatorial Search, 2009.

[188 RUI DIL AO, DANIELE MURARO, MIGUEL NICOLAU, AND MARC SCHOENAUER. Validation of a morphogenesis model of drosophila early development by a multi-objective evolutionary optimization algorithm. In Clara Pizzuti, Marylyn D. Ritchie, and Mario Giacobini, editors, European Conference on Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics, number 5483 in LNCS, pages 176–190. Springer Verlag, April 2009. Best Paper Award.

[189] ALEJANDRO ARBELAEZ AND YOUSSEF HAMADI. Improving parallel local search for SAT. In Learning and Intelligent Optimization, Fifth International Conference, LION 2011 (to appear), 2011.

[190] ALEJANDRO ARBELAEZ, YOUSSEF HAMADI, AND MICHÈLE SEBAG. Building portfolios for the protein structure prediction problem. In Workshop on Constraint Based Methods for Bioinformatics, July 2010.

[191] ALEJANDRO ARBELAEZ, YOUSSEF HAMADI, AND MICHÈLE SEBAG. Continuous search in constraint programming. In IEEE Press, editor, Proc. 22nd ICTAI, pages 53–60, 2010.

[192] D. V. ARNOLD AND N. HANSEN. Active covariance matrix adaptation for the (1+1)-CMA-ES. In ACM GECCO 2010 Proceedings, pages 385–392, 2010.

[193] ANNE AUGER, DIMO BROCKHOFF, AND NIKOLAUS HANSEN. Analyzing the impact of mirrored sampling and sequential selection in elitist evolution strategies. In Foundations of Genetic Algorithms (FOGA 2011). ACM, 2011. to appear.

[194] ANNE AUGER, NIKOLAUS HANSEN, JORGE PEREZ ZERPA, RAYMOND ROS, AND MARC SCHOENAUER. Experimental comparisons of derivative free optimization algorithms. In Jan Vahrenhold, editor, 8th International Symposium on Experimental Algorithms, number 5526 in LNCS, pages 3–15, Dortmund, 2009. Springer Verlag.

[195] A. BIERE. Lingeling, plingeling, picoSAT and precoSAT at SAT race 2010. Technical Report 10/1, FMV Reports Series, 2010.

[196] LUCAS BORDEAUX, YOUSSEF HAMADI, AND HORST SAMULOWITZ. Experiments with massively parallel constraint solving. In IJCAI, pages 443–448, 2009.

[197] AHMED BOUAJJANI AND ODED MALER, editors. Computer Aided Verification, 21st International Conference, CAV 2009, Grenoble, France, June 26 - July 2, 2009. Proceedings, volume 5643 of Lecture Notes in Computer Science. Springer, 2009.

[198] CRAIG BOUTILIER, editor. IJCAI 2009, Proceedings of the 21st International Joint Conference on Artificial Intelligence, Pasadena, California, USA, July 11-17, 2009, 2009.

[199] ZYED BOUZARKOUNA, ANNE AUGER, AND DIDIER YU DING. Investigating the local-meta-model CMA-ES for large population sizes. In 3rd European event on Bio-inspired algorithms for continuous parameter optimisation (EvoNUM'10). Springer-Verlag, 2010.

[200] ZYED BOUZARKOUNA, DIDIER YU DING, AND ANNE AUGER. Using evolution strategy with meta-models for well placement optimization. In 12th European Conference on the Mathematics of Oil Recovery (ECMOR 2010), Oxford, UK, 2010. EAGE.

[201] DIMO BROCKHOFF, ANNE AUGER, NIKOLAUS HANSEN, DIRK V. ARNOLD, AND TIM HOHM. Mirrored sampling and sequential selection for evolution strategies. In R. Schaefer et al., editor, Parallel Problem Solving from Nature (PPSN XI), volume 6238 of LNCS, pages 11–20. Springer, 2010.

[202] DAVID COHEN, editor. Principles and Practice of Constraint Programming - CP 2010 - 16th International Conference, CP 2010, St. Andrews, Scotland, UK, September 6-10, 2010. Proceedings, volume 6308 of Lecture Notes in Computer Science. Springer, 2010.

[203] GUILLAUME COLLANGE, NATHALIE DELATTRE, NIKOLAUS HANSEN, ISABELLE QUINQUIS, AND MARC SCHOENAUER. Multidisciplinary optimisation in the design of future space launchers. In P. Breitkopf and R. F. Coelho, editors, Multidisciplinary Design Optimization in Computational Mechanics, chapter 12, pages 487–496. Wiley, 2010.

[204] ALVARO FIALHO, LUIS DA COSTA, MARC SCHOENAUER, AND MICHÈLE SEBAG. Dynamic multi-armed bandits and extreme value-based rewards for adaptive operator selection in evolutionary algorithms. In T. Stuetzle et al., editor, LION'09: Proceedings of the 3rd International Conference on Learning and Intelligent OptimizatioN, volume 5851, pages 176–190. Springer Verlag, January 2009. Best Paper Award.

[205] ALVARO FIALHO, WENYIN GONG, AND ZHIHUA CAI. Probability matching-based adaptive strategy selection vs. uniform strategy selection within differential evolution. In Workshop Proceedings of the Genetic and Evolutionary Computation Conference (GECCO), pages 1527–1534, 2010.

[206] ALVARO FIALHO, RAYMOND ROS, MARC SCHOENAUER, AND MICHELE SEBAG. Comparison-based adaptive strategy selection with bandits in differential evolution. In R. Schaefer et al., editor, Parallel Problem Solving from Nature (PPSN XI), volume 6238 of LNCS, pages 194–203. Springer, 2010.

[207] ALVARO FIALHO, MARC SCHOENAUER, AND MICHÈLE SEBAG. Analysis of adaptive operator selection techniques on the royal road and long K-path problems. In G. Raidl et al., editor, Genetic and Evolutionary Computation Conference (GECCO), pages 779–786. ACM Press, July 2009.

[208] ALVARO FIALHO, MARC SCHOENAUER, AND MICHELE SEBAG. Fitness-AUC bandit adaptive strategy selection vs. the probability matching one within differential evolution. In Workshop Proceedings of the Genetic and Evolutionary Computation Conference (GECCO), pages 1535–1542, 2010.

[209] ALVARO FIALHO, MARC SCHOENAUER, AND MICHELE SEBAG. Toward comparison-based adaptive operator selection. In J. Branke et al., editor, Genetic and Evolutionary Computation Conference (GECCO), pages 767–774. ACM Press, July 2010.

[210] SILVIO GHILARDI AND ROBERTO SEBASTIANI, editors. Frontiers of Combining Systems, 7th International Symposium, FroCoS 2009, Trento, Italy, September 16-18, 2009. Proceedings, volume 5749 of Lecture Notes in Computer Science. Springer, 2009.

[211] WENYIN GONG, ALVARO FIALHO, AND ZHIHUA CAI. Adaptive strategy selection in differential evolution. In J. Branke et al., editor, Genetic and Evolutionary Computation Conference (GECCO), pages 409–416. ACM Press, July 2010.

[212] Long Guo, Youssef Hamadi, Saïd Jabbour, and Lakhdar Sais. Diversification and intensification in parallel SAT solving. In CP, pages 252–265, 2010.

[213] YOUSSEF HAMADI, SAÏD JABBOUR, AND LAKHDAR SAIS. Control-based clause sharing in parallel SAT solving. In IJCAI, pages 499–504, 2009.

[214] YOUSSEF HAMADI, SAÏD JABBOUR, AND LAKHDAR SAIS. Learning for dynamic subsumption. In ICTAI, pages 328–335, 2009.

[215] IEEE Computer Society. ICTAI 2009, 21st IEEE International Conference on Tools with Artificial Intelligence, Newark, New Jersey, USA, 2-4 November 2009, 2009.

[216] IEEE Computer Society. 22nd IEEE International Conference on Tools with Artificial Intelligence, ICTAI 2010, Arras, France, 27-29 October 2010 - Volume 1, 2010.

[217] JACQUES BIBAI, PIERRE SAVÉANT, MARC SCHOENAUER, AND VINCENT VIDAL. An evolutionary metaheuristic based on state decomposition for domain-independent satisficing planning. In Ronen Brafman, Héctor Geffner, Jörg Hoffmann, and Henry Kautz, editors, ICAPS 2010, pages 15–25. AAAI Press, May 2010.

[218] JACQUES BIBAI, PIERRE SAVÉANT, MARC SCHOENAUER, AND VINCENT VIDAL. On the generality of parameter tuning in evolutionary planning. In Genetic and Evolutionary Computation Conference (GECCO), pages 241–248. ACM Press, July 2010.

[219] S. KOTTLER. SArTagnan: solver description. Technical report, SAT Race 2010, July 2010.

[220] KE LI, ALVARO FIALHO, AND SAM KWONG. Multi-objective differential evolution with adaptive control of parameters and operators. In X. Yao et al., editor, LION'11: Proceedings of the 5th International Conference on Learning and Intelligent OptimizatioN. Springer Verlag, January 2011. (to appear).

[221] ILYA LOSHCHILOV, MARC SCHOENAUER, AND MICHÈLE SEBAG. A Mono Surrogate for Multiobjective Optimization. In J. Branke et al., editor, Genetic and Evolutionary Computation Conference (GECCO), pages 471–478. ACM Press, July 2010.

[222] ILYA LOSHCHILOV, MARC SCHOENAUER, AND MICHÈLE SEBAG. Comparison-Based Optimizers Need Comparison-Based Surrogates. In R. Schaefer et al., editor, Parallel Problem Solving from Nature (PPSN XI), volume 6238 of LNCS, pages 364–373. Springer, September 2010.

[223] ILYA LOSHCHILOV, MARC SCHOENAUER, AND MICHÈLE SEBAG. Dominance-Based Pareto-Surrogate for Multi-Objective Optimization. In Simulated Evolution and Learning (SEAL 2010), pages 230–239. LNCS 6457, Springer Verlag, December 2010.

[224] ILYA LOSHCHILOV, MARC SCHOENAUER, AND MICHÈLE SEBAG. Not all parents are equal for MO-CMA-ES. In Evolutionary Multi-Criterion Optimization 2011 (EMO 2011). Springer, April 2011. to appear.

[225] JORGE MATURANA, ALVARO FIALHO, FRÉDERIC SAUBION, MARC SCHOENAUER, AND MICHÈLE SEBAG. Extreme compass and dynamic multi-armed bandits for adaptive operator selection. In CEC'09: Proceedings of the IEEE International Conference on Evolutionary Computation, pages 365–372. IEEE, May 2009.

[226] CÉDRIC PIETTE, YOUSSEF HAMADI, AND LAKHDAR SAIS. Efficient combination of decision procedures for MUS computation. In FroCos, pages 335–349, 2009.

[227] T. SCHUBERT, M. LEWIS, AND B. BECKER. Antom: solver description. Technical report, SAT Race, 2010.

[228] JÉRÉMIE VAUTARD, ARNAUD LALLOUET, AND YOUSSEF HAMADI. A parallel solving algorithm for quantified constraints problems. In ICTAI (1), pages 271–274, 2010.

[229] T. VOSS, N. HANSEN, AND C. IGEL. Improved step size adaptation for the MO-CMA-ES. In ACM GECCO 2010 Proceedings, pages 487–494, 2010.

[230] THOMAS VOSS, NIKOLAUS HANSEN, AND CHRISTIAN IGEL. Recombination for learning strategy parameters in the MO-CMA-ES. In Fifth International Conference on Evolutionary Multi-Criterion Optimization (EMO 2009), pages 155–168. Springer-Verlag, 2009.

[231] CHRISTOPH M. WINTERSTEIGER, YOUSSEF HAMADI, AND LEONARDO MENDONÇA DE MOURA. A concurrent portfolio approach to SMT solving. In CAV, pages 715–720, 2009.

## THESES

[232] ALEJANDRO ARBELAEZ. Search with the Context. PhD thesis, Université Paris-Sud, to be defended in June 2011.

[233] ALVARO FIALHO. Adaptive Operator Selection for Optimization. PhD thesis, Université Paris-Sud, December 2010.

[234] NIKOLAUS HANSEN. Variable Metrics in Evolutionary Computation. PhD thesis, HDR – Habilitation à Diriger des Recherches, Université Paris-Sud, February 2010.

[235] RAYMOND ROS. Real-Parameter Black-Box Optimisation: Benchmarking and Designing Algorithms. PhD thesis, Université Paris-Sud, Orsay, France, December 2009.

## TECH REPORTS

[236] ANNE AUGER. Benchmarking the (1+1)-ES with One-Fifth Success Rule on the BBOB-2009 noisy Function Testbed. In G. Raidl et al., editor, Workshop Proceedings of the GECCO Genetic and Evolutionary Computation Conference, pages 2453–2458. ACM, July 2009.

[237] ANNE AUGER. Benchmarking the (1+1) evolution strategy with one-fifth success rule on the BBOB-2009 function testbed. In G. Raidl et al., editor, Workshop Proceedings of the GECCO Genetic and Evolutionary Computation Conference, pages 2447–2452. ACM, July 2009.

[238] ANNE AUGER, DIMO BROCKHOFF, AND NIKOLAUS HANSEN. Benchmarking the (1,4)-CMA-ES with mirrored sampling and sequential selection on the noiseless BBOB-2010 testbed. In Workshop Proceedings of the Genetic and Evolutionary Computation Conference (GECCO), pages 1617–1624, 2010.

[239] ANNE AUGER, DIMO BROCKHOFF, AND NIKOLAUS HANSEN. Benchmarking the (1,4)-CMA-ES with mirrored sampling and sequential selection on the noisy BBOB-2010 testbed. In Workshop Proceedings of the Genetic and Evolutionary Computation Conference (GECCO), pages 1625–1632, 2010.

[240] ANNE AUGER, DIMO BROCKHOFF, AND NIKOLAUS HANSEN. Comparing the (1+1)-CMA-ES with a mirrored (1+2)-CMA-ES with sequential selection on the noiseless bbob-2010 testbed. In Workshop Proceedings of the Genetic and Evolutionary Computation Conference (GECCO), pages 1543–1550, 2010.

[241] ANNE AUGER, DIMO BROCKHOFF, AND NIKOLAUS HANSEN. Investigating the impact of sequential selection in the (1,2)-CMA-ES on the noiseless bbob-2010 testbed. In Workshop Proceedings of the Genetic and Evolutionary Computation Conference (GECCO), pages 1591–1596, 2010.

[242] ANNE AUGER, DIMO BROCKHOFF, AND NIKOLAUS HANSEN. Investigating the impact of sequential selection in the (1,2)-CMA-ES on the noisy bbob-2010 testbed. In Workshop Proceedings of the Genetic and Evolutionary Computation Conference (GECCO), pages 1605–1610, 2010.

[243] ANNE AUGER, DIMO BROCKHOFF, AND NIKOLAUS HANSEN. Investigating the impact of sequential selection in the (1,4)-CMA-ES on the noiseless BBOB-2010 testbed. In Workshop Proceedings of the Genetic and Evolutionary Computation Conference (GECCO), pages 1597–1604, 2010.

[244] ANNE AUGER, DIMO BROCKHOFF, AND NIKOLAUS HANSEN. Investigating the impact of sequential selection in the (1,4)-CMA-ES on the noisy BBOB-2010 testbed. In Workshop Proceedings of the Genetic and Evolutionary Computation Conference (GECCO), pages 1611–1616, 2010.

[245] ANNE AUGER, DIMO BROCKHOFF, AND NIKOLAUS HANSEN. Mirrored sampling and sequential selection for evolution strategies. Rapport de Recherche RR-7249, INRIA Saclay—Ile-de-France, April 2010.

[246] ANNE AUGER, DIMO BROCKHOFF, AND NIKOLAUS HANSEN. Mirrored variants of the (1,2)-CMA-ES compared on the noiseless BBOB-2010 testbed. In Workshop Proceedings of the Genetic and Evolutionary Computation Conference (GECCO), pages 1551–1558, 2010.

[247] ANNE AUGER, DIMO BROCKHOFF, AND NIKOLAUS HANSEN. Mirrored variants of the (1,2)-CMA-ES compared on the noisy BBOB-2010 testbed. In Workshop Proceedings of the Genetic and Evolutionary Computation Conference (GECCO), pages 1575–1582, 2010.

[248] ANNE AUGER, DIMO BROCKHOFF, AND NIKOLAUS HANSEN. Mirrored variants of the (1,4)-CMA-ES compared on the noiseless BBOB-2010 testbed. In Workshop Proceedings of the Genetic and Evolutionary Computation Conference (GECCO), pages 1559–1566, 2010.

[249] ANNE AUGER, DIMO BROCKHOFF, AND NIKOLAUS HANSEN. Mirrored variants of the (1,4)-CMA-ES compared on the noisy BBOB-2010 testbed. In Workshop Proceedings of the Genetic and Evolutionary Computation Conference (GECCO), pages 1583–1590, 2010.

[250] ANNE AUGER, STEFFEN FINCK, NIKOLAUS HANSEN, AND RAYMOND ROS. BBOB 2009: Comparison Tables of All Algorithms on All Noiseless Functions. Technical Report RT-0383, INRIA, 04 2010.

[251] ANNE AUGER, STEFFEN FINCK, NIKOLAUS HANSEN, AND RAYMOND ROS. BBOB 2009: Comparison Tables of All Algorithms on All Noisy Functions. Technical Report RT-0384, INRIA, 04 2010.

[252] ANNE AUGER AND NIKOLAUS HANSEN. Benchmarking the (1+1)-CMA-ES on the BBOB-2009 function testbed. In G. Raidl et al., editor, Workshop Proceedings of the GECCO Genetic and Evolutionary Computation Conference, pages 2459–2466. ACM, July 2009.

[253] ANNE AUGER AND NIKOLAUS HANSEN. Benchmarking the (1+1)-CMA-ES on the BBOB-2009 noisy testbed. In G. Raidl et al., editor, Workshop Proceedings of the GECCO Genetic and Evolutionary Computation Conference, pages 2467–2472. ACM, July 2009.

[254] ANNE AUGER AND RAYMOND ROS. Benchmarking the pure random search on the BBOB-2009 noisy testbed. In G. Raidl et al., editor, Workshop Proceedings of the GECCO Genetic and Evolutionary Computation Conference, pages 2485–2490. ACM, July 2009.

[255] ANNE AUGER AND RAYMOND ROS. Benchmarking the pure random search on the BBOB-2009 testbed. In G. Raidl et al., editor, Workshop Proceedings of the GECCO Genetic and Evolutionary Computation Conference, pages 2479–2484. ACM, July 2009.

[256] ALVARO FIALHO AND RAYMOND ROS. Analysis of adaptive strategy selection within differential evolution on the BBOB-2010 noiseless benchmark. INRIA Research Report RR-7259, INRIA Saclay - Ile-de-France, April 2010.

[257] STEFFEN FINCK, NIKOLAUS HANSEN, RAYMOND ROS, AND ANNE AUGER. Real-parameter black-box optimization benchmarking 2009: Presentation of the noiseless functions. Technical Report 2009/20, Research Center PPE, 2009.

[258] STEFFEN FINCK, NIKOLAUS HANSEN, RAYMOND ROS, AND ANNE AUGER. Real-parameter black-box optimization benchmarking 2009: Presentation of the noisy functions. Technical Report 2009/21, Research Center PPE, 2009.

[259] N. HANSEN, A. AUGER, S. FINCK, AND R. ROs. Real-parameter black-box optimization benchmarking 2010: Experimental setup. Technical Report RR-7215, INRIA, 2010.

[260] N. HANSEN, A. AUGER, R. ROS, S. FINCK, AND P. POSÍK. Comparing results of 31 algorithms from the black-box optimization benchmarking BBOB-2009. In J. Branke et al., editor, GECCO (Companion), pages 1689–1696. ACM, July 2010.

[261] N. HANSEN AND R. ROS. Benchmarking a weighted negative covariance matrix update on the BBOB-2010 noiseless testbed. In J. Branke et al., editor, GECCO (Companion), pages 1673–1680. ACM, July 2010.

[262] N. HANSEN AND R. ROS. Benchmarking a weighted negative covariance matrix update on the BBOB-2010 noisy testbed. In J. Branke et al., editor, GECCO (Companion), pages 1681–1688. ACM, July 2010.

[263] N. HANSEN AND R. ROS. Black-box optimization benchmarking of NEWUOA compared to BIPOP-CMA-ES: on the BBOB noiseless testbed. In GECCO (Companion), pages 1519–1526, July 2010.

[264] NIKOLAUS HANSEN. Benchmarking a BI-population CMA-ES on the BBOB-2009 function testbed. In G. Raidl et al., editor, Workshop Proceedings of the GECCO Genetic and Evolutionary Computation Conference, pages 2389–2395. ACM, July 2009.

[265] NIKOLAUS HANSEN. Benchmarking a BI-population CMA-ES on the BBOB-2009 noisy testbed. In G. Raidl et al., editor, Workshop Proceedings of the GECCO Genetic and Evolutionary Computation Conference, pages 2397–2402. ACM, July 2009.

[266] NIKOLAUS HANSEN. Benchmarking the Nelder-Mead downhill simplex algorithm with many local restarts. In G. Raidl et al., editor, Workshop Proceedings of the GECCO Genetic and Evolutionary Computation Conference, pages 2403–2408. ACM, July 2009.

[267] NIKOLAUS HANSEN, ANNE AUGER, STEFFEN FINCK, AND RAYMOND ROS. Real-parameter black-box optimization benchmarking 2009: Experimental setup. Research Report RR-6828, INRIA, 2009.

[268] NIKOLAUS HANSEN, STEFFEN FINCK, RAYMOND ROS, AND ANNE AUGER. Real-parameter black-box optimization benchmarking 2009: Noiseless functions definitions. Technical Report RR-6829, INRIA, 2009.

[269] NIKOLAUS HANSEN, STEFFEN FINCK, RAYMOND ROS, AND ANNE AUGER. Real-parameter black-box optimization benchmarking 2009: Noisy functions definitions. Technical Report RR-6869, INRIA, 2009.

[270] RAYMOND ROS. Benchmarking the BFGS algorithm on the BBOB-2009 function testbed. In G. Raidl et al., editor, Workshop Proceedings of the GECCO Genetic and Evolutionary Computation Conference, pages 2409–2414. ACM, July 2009.

[271] RAYMOND ROS. Benchmarking the BFGS algorithm on the BBOB-2009 noisy testbed. In G. Raidl et al., editor, Workshop Proceedings of the GECCO Genetic and Evolutionary Computation Conference, pages 2415–2420. ACM, July 2009.

[272] RAYMOND ROS. Benchmarking the NEWUOA on the BBOB-2009 function testbed. In G. Raidl et al., editor, Workshop Proceedings of the GECCO Genetic and Evolutionary Computation Conference, pages 2421–2428. ACM, July 2009.

[273] RAYMOND ROS. Benchmarking the NEWUOA on the BBOB-2009 noisy testbed. In G. Raidl et al., editor, Workshop Proceedings of the GECCO Genetic and Evolutionary Computation Conference, pages 2429–2434. ACM, July 2009.

## TALKS

[274] ANNE AUGER. New perspectives in stochastic derivative free optimization. Workshop on Advanced Methods and Perspective in Nonlinear Optimization and Control, Toulouse, France, February 2010.

[275] YOUSSEF HAMADI. From SAT to parallel SAT, calais, france. 22th International Conference on Tools with Artificial Intelligence (ICTAI 2010), October 2010.

[276] YOUSSEF HAMADI. Parallelism in SAT, twente, netherlands. 9th International Workshop on Parallel and Distributed Methods in verifiCation (PDMC 2010), October 2010.

[277] YOUSSEF HAMADI. Control-based clause-sharing, lisbon, portugal. INESC-ID, Lisbon, February 2011.

[278] NIKOLAUS HANSEN. Invariance properties of CMA-ES. PPSN Workshop on Self-tuning, self-configuring and self-generating Search Heuristics (Self* 2010), Krakow, Poland, September 2010.

[279] MARC SCHOENAUER. Experimental comparisons of derivative free optimization algorithms. SEA, 8th International Symposium on Experimental Algorithms, Dortmund, Germany, June 2009.

[280] MARC SCHOENAUER. Les algorithmes évolutionnaires, outils de créativité artificielle ? ArchiLab, Orléans, France, November 2009.

[281] MICHÈLE SEBAG. Toward autonomic computational systems. COE Program for Next Generation Information Technology, Sapporo, Japan, January 2009.

[282] MICHÈLE SEBAG. Planlearn, planning to learn. Workshop at ECAI 2010, Lisboa, Portugal, August 2010.

[283] MICHÈLE SEBAG. Self-driven rewards for an autonomous robot: An information theoretic approach, calais, france. 22th International Conference on Tools with Artificial Intelligence (ICTAI 2010), October 2010.

**Track B** **This project started in September 2008.**

# IMAGE AND VIDEO MINING FOR SCIENCE AND HUMANITIES

## OVERVIEW

> **This e-science project builds on several ideas articulated in the "2020 Science[5]" report, including the importance of data mining and machine learning in computationals sciences.**

Our project involves fundamental computer science research in computer vision and machine learning, applied to archaeology, cultural heritage preservation, environmental science, and sociology, and validated by collaborations with researchers and practitioners in these fields. Concretely, we propose to address: (i) Mining historical collections of photographs and paintings with applications to archaeology and cultural heritage preservation; (ii) Mining and analysis of TV broadcasts with applications to sociology; (iii) the problem of detection, identification and tracking of dynamical geophysical events, with application to risk assessment and weather forecast.

**Jean Ponce** graduated from ENSET in 1982, received a PhD at University of Paris 11 in 1988. He has been a visiting scientist at MIT during 2 years and at Stanford University in 1985-1989, and professor at University of Illinois (Urbana-Champaign) during 15 years. He is now a computer science professor at École Normale Supérieure (ENS) in Paris and heads the joint ENS/INRIA/project-team WILLOW. Jean Ponce is the author of 120 technical publications in computer vision and robotics, including the textbook ``Computer Vision: A Modern Approach". He is editor-in-chief for the International Journal of Computer Vision.

## TEAM

| Team leader | PONCE | Jean | *INRIA Paris-Rocquencourt* |
|---|---|---|---|
| Researcher | BLAKE | Andrew | *Microsoft Research Cambridge* |
| Researcher | DESSALES | Hélène | *Ecole Normale Supérieure de Paris* |
| Researcher | HARCHAOUI | Zaid | *INRIA Grenoble-Rhône-Alpes* |
| Researcher | JEGOU | Hervé | *INRIA Grenoble-Rhône-Alpes* |
| Researcher | LABORELLI | Louis | *Institut National de l'Audiovisuel (INA)* |
| Researcher | LAPTEV | Ivan | *INRIA Rennes Bretagne Atlantique* |
| Researcher | MEMIN | Etienne | *INRIA Rennes Bretagne Atlantique* |
| Researcher | PEREZ | Patrick | *INRIA Rennes Bretagne Atlantique* |
| Researcher | SCHMID | Cordelia | *INRIA Grenoble-Rhône-Alpes* |

**Left:**

- *Neva Cherniavsky (post-doc, PhD University of Washington) was hired from March 1st 2009 until August 31st 2010. (Currently post-doc at MIT, Cambridge, USA).*
- *Bryan Russell (post-doc, Phd MIT) was hired June 1st 2009 until November 30, 2010. (Currently at Intel Research, Seattle, USA).*

## RESEARCH

Concretely, we propose to address the following problems:

- *Mining historical collections of photographs and paintings with applications to archeology and cultural heritage preservation*. This includes the use of image-based modeling technology to facilitate the metrology side of field work, but also the quantitative analysis of environmental damage on wall paintings or mosaics over time, and the cross-indexing of XIXth Century paintings of Pompeii with modern photographs. This part of our research is done in collaboration with Helene Dessales at the ENS Archeology Laboratory.
- *Mining TV broadcasts with applications to sociology*. This includes automating the analysis and annotation of human actions and interactions in video segments to assist - and provide data for - studies of consumer trends in commercials, political event coverage in newscasts, and class- and gender-related behavior patterns in situation comedies, for example. This part of our research is done in collaboration with Louis Laborelli and Daniel Teruggi at Institut National de l'Audiovisuel (INA).
- *Detection, identification and tracking of dynamical geophysical events*, with application to risk assessment and weather forecast.

For every one of the problems we have in mind, indexing, searching and analyzing photo and video collections is a key issue. Recent advances in image analysis, computer vision, and machine learning promise an opportunity to automate, partly or completely, these tasks (e.g., annotation of photos and videos), as well as to access information whose extraction from images is simply beyond human capabilities (e.g., indexing of very large image archives). To fulfill this promise, we propose to conduct fundamental research in object, scene, and activity modeling, learning, and recognition, and to validate it with the development of computerized image and video mining tools at the service of sciences and humanities.

## ORGANIZATION

Our project brings together three INRIA teams, LEAR, FLUMINANCE and WILLOW (the latter is a joint venture between INRIA, Ecole Normale Superieure [ENS], and CNRS, and is hosted by ENS in Paris), with complementary strengths in computer vision (image-based modeling, dynamic image analysis, object recognition) and machine learning. It involves MSR researchers from Cambridge, under the leadership of A. Blake, as well as others from Bangalore and Redmond for example. It also involves external partners including Helene Dessales at the ENS Archaeology Laboratory, as well as Louis Laborelli and Daniel Teruggi at Institut National de l'Audiovisuel (INA). The Collaborative Research Agreement (CRA) between Microsoft Research, INRIA, and ENS has been signed on September 15, 2008.

We have hired two Phd students, W. Harchaoui (WILLOW) and A. Gaidon (LEAR), and post-doc S. Gorthi (FLUMINANCE). Oliver Whyte (Phd student at WILLOW) is funded by the project since September 2010 working on a collaborative project with R. Szeliski (MSR Redmond), J. Sivic and A. Zisserman (WILLOW). In 2010-2011, we have also hired three interns (J. Lezama, U. Jardonnet, P. Gronat) to work on video analysis and large scale image matching. Bryan Russell and Neva Cherniavsky, hired as post-docs in 2009, left WILLOW in 2010. We are planning to hire two new post-docs in 2011. classes of activities in video.

| | | | |
|---|---|---|---|
| Researcher | SIVIC | Josef | INRIA Paris-Rocquencourt |
| Researcher | SZELISKI | Rick | *Microsoft Research Redmond* |
| Researcher | ZISSERMAN | Andrew | *Oxford University - Ecole Normale Supérieure de Paris* |
| Post doc | CHERMIAVSKY | Neva | *MSR-INRIA Joint Centre* |
| Post-doc | GORTHI | Rama Krishna Sai Subrah-manyam | *INRIA Rennes Bretagne Atlantique* |
| Post-doc | RUSSELL | Bryan | *INRIA Paris-Rocquencourt* |
| PHD student | GAIDON | Adrien | *INP Grenoble* |
| PHD student | HARCHAOUI | Warith | *INRIA Paris-Rocquencourt* |
| PHD student | WHYTE | Oliver | *INRIA Paris-Rocquencourt* |

**Visitors:** *Alexei Efros (Professor, Carnegie Mellon University, USA) visited WILLOW for 2 months in Summer 2010. Fredo Durand (Professor, Massachusetts Institute of Technology, USA) spent the academic year 2009-2010 as a visiting professor at WILLOW. Short term visitors to WILLOW include (among others): Ram Nevatia (University of Southern California, USA), Rick Szeliski (Microsoft Research, USA), Leon Bottou (Microsoft Research, USA) and Tomas Pajdla (Czech Technical University in Prague).*

# RESULTS

## QUANTITATIVE IMAGE ANALYSIS FOR ARCHEOLOGY

The goal of this project is to enable fully automatic matching and alignment of paintings and drawings to photographs depicting a complex 3D scene. This is an extremely difficult task due to various distortions that can arise such as perspective or caricature distortion as well as inaccuracies due to drawing by hand. Progress on this topic is of interest to archaeologists, artists or curators. For this, B. Russell and H. Dessales visited the archaeological site at Pompeii to photograph the Championnet house, which is the focus of our study. From this set of images (in total, 585 photographs were taken), we were able to produce a large-scale dense 3D reconstruction of the Championnet house using existing photometric multi-view stereo methods. *Figure 5* shows some of the captured photographs and snapshots of the 3D reconstruction of the site. Notice that the 3D reconstruction captures much detail of the walls and structures.

Next, we have obtained initial results on coarse alignment of paintings with the 3D model of the site. This is achieved by matching to virtual viewpoints that are uniformly sampled across the 3D model. The result is a viewpoint that is sufficiently close to the painting viewpoint, where the depicted scene objects in the painting are close to their 3D model projection.

Currently we are exploring different techniques to refine the obtained viewpoints and align the paintings, and drawings with the 3D model.

**Large scale image matching**. Building large scale 3D models from photographs involves establishing correspondences between large amounts of images, which is one of the most time consuming steps in the 3D reconstruction pipeline. We have developed [328] an approach for detecting and removing non-informative image features (such as trees or road markings), which often confuse the image matching process and are responsible for many mismatches. The proposed technique has demonstrated significant gains in matching accuracy for place recognition and we plan to investigate its benefits for speeding-up large scale 3D reconstruction. In addition, we have developed a method for learning descriptors for large scale matching [334], which improves matching accuracy, while maintaining efficiency.

**A geometric model for non-uniform deblurring of camera shake**. The blur caused by camera-shake spoils many photos taken in low-light conditions. In addition, within the context of quantitative image analysis for archaeology, if the source images are blurred, de-blurring is likely to improve matching and registration accuracy.

While significant progress has been made recently towards removing this blur from images, almost all approaches model the blurred image as a convolution of a sharp image with a spatially uniform filter. However, blur from camera shake is mostly due to the 3D rotation of the camera, resulting in a blur kernel that can be significantly non-uniform across the image. In [339], we have proposed a new parametrized geometric model of the blurring process in terms of the

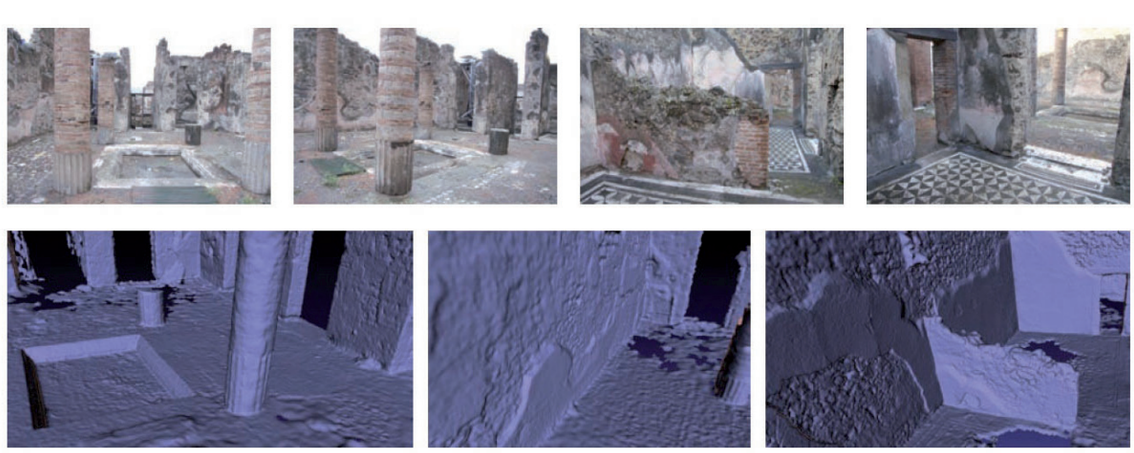



Figure 5: (a) Example photographs captured of the Pompeii site (563 photographs are used in total). (b) Rendered viewpoints of the recovered 3D model. Notice the fine-level details that are captured by the model.

rotational velocity of the camera during exposure. We have shown that our approach makes it possible to model and remove a wider class of blurs than previous approaches, including uniform blur as a special case, and demonstrate its effectiveness with experiments on real images.

## ACTION MODELING AND RECOGNITION

Based on our earlier promising results on human action classification reported in [329], we address a more challenging problem and propose to localize actions in time and space, i.e. answering the questions "where" and "when" actions (e.g. Open Door or Drinking) appear in the video.

Our first contribution, published in [316], addresses the problem of automatic temporal annotation of realistic human actions in video using minimal manual supervision. To this end we consider two associated problems: (a) weakly-supervised learning of action models from readily available annotations, and (b) temporal localization of human actions in test videos. To avoid the prohibitive cost of manual annotation for training, we use movie scripts as a means of weak supervision. Scripts, however, provide only implicit, sometimes noisy, and imprecise information about the type and location of actions in video (cf. *Figure 6(a)*). We address this problem with a kernel-based discriminative clustering algorithm that locates actions in the weakly-labeled training data (cf. *Figure 6(b)*). Using the obtained action samples, we train temporal action detectors and apply them to locate actions in the raw video data. Our experiments demonstrate that the proposed method for weakly-supervised learning of action models leads to significant improvement in action detection. We present detection results for three action classes in four feature length movies with challenging and realistic video data.

In our more recent work [326], we proposes a novel human-centric approach to localize human actions in time and in space, We show that splitting the action localization task into spatial and temporal search leads to an efficient localization algorithm where generic human tracks can be reused to recognize multiple human actions. We develop a generic human detector and tracker which is able to cope with a wide range of postures, articulations, motions and camera viewpoints and propose a track-aligned 3D-HOG action representation. Results are presented on a number of real-world movies with crowded, dynamic environment, partial occlusion and cluttered background. We significantly improve over state-of-the-art, including improvements on a new dataset based on Hollywood movies that we have introduced (see *Figure 7*).

Besides the work on action localization, we have addressed several other directions aiming to improve recognition of


Figure 6: Video clips with *OpenDoor* actions provided by automatic script-based annotation. Selected frames illustrate both the variability of action samples within a class as well as the imprecise localization of actions in video clips. (b): In feature space, positive samples are constrained to be located on temporal feature tracks corresponding to consequent temporal windows in video clips. Background (non-action) samples provide further constrains on the clustering.


Figure 7: The five highest ranked phoning (top) and standing-up (bottom) actions detected on the Hollywood-Localization dataset.

actions and human traits in video and still images.

**Actions in context**. In [332] we exploit the context of natural dynamic scenes for human action recognition in video. We (a) automatically discover relevant scene classes and their correlation with human actions, (b) show how to learn selected scene classes from video without manual supervision and (c) develop a joint framework for action and scene recognition and demonstrate improved recognition of both in challenging video data.

**Action re-ranking**. In [318] we have proposed an extension of the text-based action retrieval approach of [329]. We remove "outliers" generated by the purely text-based approach by ranking the retrieved videos by visual consistency. We present a new iterative ranking algorithm, called iter-SVR, that improves over both text-based results and state-of-the-art outlier detection approaches on actions in challenging videos.

**Bag of features with non-local cues**. In [338] we improve local space-time feature video representations by integrating additional non-local cues in the bag-of-features model. We decompose video into regions and use pre-learned region properties together with local features to improve description of human actions in video. We show significant improvements on the challenging Hollywood-2 human actions dataset.

**Action recognition in still images**. In [315] we address the under-studied problem of human action recognition in still images. We construct a new dataset from user-generated images and study action recognition using the state-of-the-art generic methods for still image object categorization and detection. We demonstrate an improved accuracy compared to previous results in the literature.

**Semi-supervised learning of facial attributes**. Building on our work on annotating TV videos [336], in [314] we investigate a weakly-supervised approach to learning facial attributes of humans in video. First, we show that training on video data improves classification performance over training on images alone. Second, and more significantly, we show that tracks in video provide a natural mechanism for expanding the still-image training data. Improved results are demonstrated for classification of gender and age attributes in challenging videos.

**Modeling temporal structure of human actions and long-range temporal interactions**. Our most recent work [317] (to appear at CVPR 2011) pertains to the modeling of the temporal structure of actions, in order to improve temporal action detection of short actions in long video streams. In a parallel work [330] (to appear at the same venue), we consider long-range temporal interactions and address low-level spatio-temporal video segmentation. Our future work will focus on several directions including (i) the design of novel action unit representations and the unsupervised learning of action units at training time and (ii) unified understanding of dynamic scenes by exploring high-level interactions between objects, people and scene structure.

## ANALYSIS OF DYNAMICAL GEOPHYSICAL EVENTS

We have addressed the three following problems within the context of analyzing dynamical physical events.

**Variational Data Assimilation for Convective Cells Tracking** [304]. This work focuses on the tracking and analysis of convective clouds systems from Meteosat Second Generation images. The highly deformable nature of convective clouds, the complexity of the physical processes involved, but also the partially hidden

measurements available from image data make difficult a direct use of conventional image analysis techniques for tasks of detection, tracking and characterization. We face these issues using variational optimal control techniques. Such techniques enable to perform the estimation of an unknown state function according to a given dynamical model and to noisy and incomplete measurements. The system state we are setting in this study for the system clouds representation is composed of two nested curves corresponding to the exterior frontiers of the clouds and to the interior coldest parts (heart) of the convective clouds. Since no reliable simple dynamical model exists for such phenomena at the image grid scale, the dynamics on which we are relying has been directly defined from image based motion measurements and takes into account an uncertainty modeling of the curves dynamics along time. In addition to this assimilation technique, we also investigate how each cell of the recovered clouds system can be labeled and associated to characteristic parameters (birth or death time, mean temperature, velocity, growth, etc.) of great interest for meteorlogists.

**Level set tracking through particle filtering** [306, 307]. In this work we have defined a proper continuous stochastic dynamics of a level sets surface for the tracking of highly deformable objects from a discrete sequence of images. This dynamics is defined from a low dimensional noise and a second vectorial level sets function allowing to maintain

## TOOLS AND SOFTWARE

- **PMVS** (Patch-based Multi-view Stereo Software) package was developed in collaboration with Y. Furukawa at the University of Illinois at Urbana-Champaign. The software is released as open source and is available at *http://grail.cs.washington.edu/software/pmvs/*.
- **Accurate Calibration Software**. This package, developed once again in collaboration with Y. Furukawa at UIUC, provides a toolkit for high-accuracy camera calibration and object and scene modeling. It is available for free to academics at *http://www.cs.washington.edu/homes/furukawa/research/pba/*.
- **Non-uniform Deblurring for Shaken Images**. This Matlab package contains code to perform blind deblurring of non-uniform / spatially-varying blur caused by camera shake. It is publicly available at http://www.di.ens.fr/willow/research/deblurring/.
- **Dense local space-time features STIP-2.0** *http://www.irisa.fr/vista/Equipe/People/Laptev/download.html\#stip*
- **Software for large scale search**. This software finds similar images in a database of 10 million images in about one second. Its performance is demonstrated on-line, see *http://bigimbaz.inrialpes.fr*. The code has been transferred to Milpix and research licenses have been granted to Stanford University, University of California at San Diego and the California Institute of Technology.

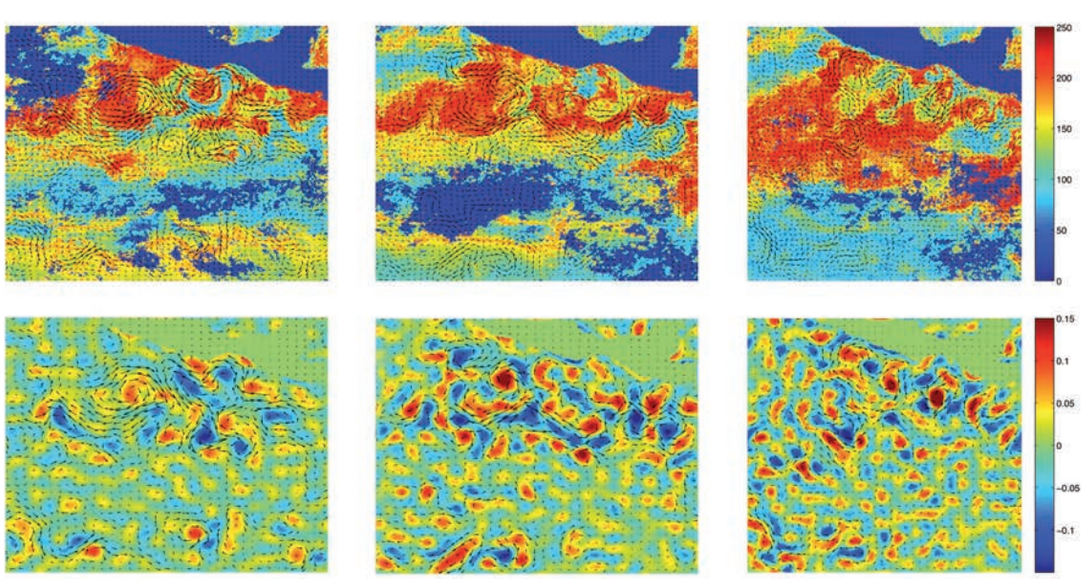



Figure 8: Real satellite sequence of Sea Surface Temperature images (48 days) off the Panama isthmus in the Pacific ocean shots during an El Nino- Southern Oscillation. First row: SST images superimposed with the velocity fields estimated through the proposed filtering method; second row: corresponding vorticity maps.

a backward correspondence of the curve of interest. The tracker, implemented through a Bayesian recursive filter, has been successfully validated for the tracking of convective cloud systems from satellite images. In the future we aim at assessing the performance of the dynamics for very short-time (20-30 mn) forecasting of convective cloud systems. Such an issue is for instance of major interest for aircraft routing problem. This technique has been successfully applied for meteorological radar images, Infrared satellite meteorological images and and ice-density satellite images. It allows to track along time isolines of a scalar transported by the geophysical fluid flow and associated to temperature or density.

**Ensemble filter strategy for the tracking of geophysical flows** [301]. We have investigated the use of so-called ensemble Kalman filtering for fluid tracking problems. This kind of filter 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 Kalman gain, rely on an ensemble of forecasts. Such a process gives rise to a Monte Carlo approximation for a family of stochastic non-linear 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 ensemble kalman filter. This technique leads to ensemble Kalman filter with an improved efficiency. The technique has been recently extended to introduce nonlinear image reconstruction errors and has been successfully applied on Sea Surface Temperature images. The dynamics considered corresponds to a velocity-vorticity stochastic formulation of the Navier-Stokes equation coupled to nonlinear image measurements. Apart from the initial instant, the technique does not require any motion estimates. It provides very promising results. In future works we aim at extending this framework to rely on more complex models of the oceanic dynamics. ■

## PUBLICATIONS & TALKS

### JOURNAL PAPERS AND BOOK CHAPTERS


[284] SYLVAIN ARLOT AND PETER L. BARTLETT. Margin adaptive model selection in statistical learning. Bernoulli, 2010. Accepted. arXiv:0804.2937.

[285] SYLVAIN ARLOT, GILLES BLANCHARD, AND ETIENNE ROQUAIN. Some non-asymptotic results on resampling in high dimension, II: Multiple tests. The Annals of Statistics, 38(1):83–99, 2010.

[286] SYLVAIN ARLOT AND ALAIN CELISSE. Segmentation of the mean of heteroscedastic data via cross-validation. Statistics and Computing, pages 1–20, 2010.

[287] SYLVAIN ARLOT AND ALAIN CELISSE. A survey of cross-validation procedures for model selection. Statist. Surv., 4:40–79, 2010.

[288] JEAN-YVES AUDIBERT AND SÉBASTIEN BUBECK. Regret bounds and minimax policies under partial monitoring. Journal of Machine Learning Research, 11:2635–2686, October 2010.

[289] FRANCIS BACH. Self-concordant analysis for logistic regression. Electronic Journal of Statistics, 4:384–414, 2010.

[290] OLIVIER DUCHENNE, FRANCIS BACH, INSO KWEON, AND JEAN PONCE. A tensor-based algorithm for high-order graph matching. PAMI, 2011. to appear.

[291] Y. FURUKAWA AND JEAN PONCE. Accurate, dense, and robust multi-view stereopsis. IEEE Trans. Patt. Anal. Mach. Intell., 32(8), 2010.

[292] RODOLPHE JENATTON, JULIEN MAIRAL, GUILLAUME OBOZINSKI, AND FRANCIS BACH. Proximal methods for hierarchical sparse coding. Journal Machine Learning Research, 2010. to appear.

[293] MICHEL JOURNÉE, FRANCIS BACH, P.-A. ABSIL, AND RODOLPHE SEPULCHRE. Low-rank optimization on the cone of positive semidefinite matrices. SIAM Journal on Optimization, 20(5):2327–2351, 2010.

[294] I. JUNEJO, E. DEXTER, I. LAPTEV, AND P. PEREZ. View-independent action recognition from temporal self-similarities. IEEE Transactions on Pattern Analysis and Machine Intelligence, 33(1):172–185, 2011.

[295] I. JUNEJO, E. DEXTER, IVAN LAPTEV, AND PATRICK PÉREZ. View-independent action recognition from temporal self-similarities. IEEE Trans. Patt. Anal. Mach. Intell., in press, 2010.

[296] BILIANA KANEVA, JOSEF SIVIC, A. TORRALBA, SHAI AVIDAN, AND WILLIAM T. FREEMAN. Infinite images: Creating and exploring a large photorealistic virtual space. Proceedings of the IEEE, 98(8):1391–1407, 2010.

[297] HUI KONG, JEAN-YVES AUDIBERT, AND JEAN PONCE. Detecting abandoned objects with a moving camera. Image Processing, IEEE Transactions on, 19(8):2201–2210, 2010.

[298] HUI KONG, JEAN-YVES AUDIBERT, AND JEAN PONCE. General road detection from a single image. Image Processing, IEEE Transactions on, 19(8):2211–2220, 2010.

[299] JULIEN MAIRAL, FRANCIS BACH, JEAN PONCE, AND GUILLERMO SAPIRO. Online learning for matrix factorization and sparse coding. Journal of Machine Learning Research, 11(1):19–60, January 2010.

[300] JULIEN MAIRAL, RODOLPHE JENATTON, GUILLAUME OBOZINSKI, AND FRANCIS BACH. Network flow algorithms for structured sparsity. Journal Machine Learning Research, 2010. to appear.

[301] N. PAPADAKIS, E. MEMIN, N. CUZOL, AND N. GENGEMBRE. Data assimilation with the weighted ensemble kalman filter. Tellus A: Dynamic Meteorology and Oceanography, 62(5):673–2697, 2010.

[302] JAMES PHILBIN, JOSEF SIVIC, AND ANDREW ZISSERMAN. Geometric latent dirichlet allocation on a matching graph for large-scale image datasets. International Journal of Computer Vision, 2010.

[303] H. SAHBI, JEAN-YVES AUDIBERT, AND RENAUD KERIVEN. Context-dependent kernels for object classification. IEEE Transactions on Pattern Analysis and Machine Intelligence, November 2010.

[304] C. THOMAS, T. CORPETTI, AND E MEMIN. Data assimilation for convective cells tracking on meteorological image sequences. IEEE trans. on Geoscience and Remote sensing, 48(8):3162–3177, 2010.

## CONFERENCE AND WORKSHOP PAPERS

[305] JEAN-YVES AUDIBERT, SÉBASTIEN BUBECK, AND RÉMI MUNOS. Best Arm Identification in Multi-Armed Bandits. In Proceedings of the 23th annual conference on Computational Learning Theory (COLT), 2010.

[306] C. AVENEL, E. MEMIN, AND P. PEREZ. Stochastic filtering of level sets for curve tracking. In International conference on Pattern Recognition, 2010.

[307] C. AVENEL, E. MEMIN, AND P. PEREZ. Tracking closed curves with non-linear stochastic filters. In Curves and Surfaces, 2010.

[308] FRANCIS BACH. Structured sparsity-inducing norms through submodular functions. In NIPS 2010: Twenty-Fourth Annual Conference on Neural Information Processing Systems, pages 118–126, Canada Vancouver, 2010.

[309] FRANCIS BACH. Structured sparsity-inducing norms through submodular functions. In Adv. Neural Info. Proc. Systems, 2010.

[310] GUILLAUME BATOG, XAVIER GOAOC, AND JEAN PONCE. Admissible linear map models of linear cameras. In 23rd IEEE Conference on Computer Vision and Pattern Recognition - CVPR 2010, pages 1578–1585, United States San Francisco, June 2010. IEEE.

[311] GUILLAUME BATOG, XAVIER GOAOC, AND JEAN PONCE. Admissible linear map models of linear cameras. In Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, pages 1578–1585. IEEE, 2010.

[312] Y-LAN BOUREAU, FRANCIS BACH, YANN LECUN, AND JEAN PONCE. Learning mid-level features for recognition. In Proc. International Conference on Computer Vision and Pattern Recognition (CVPR'10). IEEE, 2010.

[313] Y-LAN BOUREAU, JEAN PONCE, AND YANN LECUN. A theoretical analysis of feature pooling in vision algorithms. In Proc. International Conference on Machine learning (ICML'10), 2010.

[314] NEVA CHERNIAVSKY, IVAN LAPTEV, JOSEF SIVIC, AND ANDREW ZISSERMAN. Semi-supervised learning of facial attributes in video. In The first international workshop on parts and attributes (in conjunction with ECCV 2010), 2010.

[315] VINCENT DELAITRE, IVAN LAPTEV, AND JOSEF SIVIC. Recognizing human actions in still images: a study of bag-of-features and part-based representations. In Proceedings of the British Machine Vision Conference, 2010.

[316] O. DUCHENNE, I. LAPTEV, J. SIVIC, F. BACH, AND J. PONCE. Automatic annotation of human actions in video. In Proceedings of the IEEE International Conference on Conference on Computer Vision, 2009.

[317] ADRIEN GAIDON, ZAID HARCHAOUI, AND CORDELIA SCHMID. Actom sequence models for efficient action detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2011.

[318] ADRIEN GAIDON, MARCIN MARSZAŁEK, AND CORDELIA SCHMID. Mining visual actions from movies. In Proceedings of the British Machine Vision Conference, 2009.

[319] S. GRÜNEWÄLDER, JEAN-YVES AUDIBERT, M. OPPER, AND J. SHAWE-TAYLOR. Regret bounds for gaussian process bandit problems. In Proceedings of the 14th International Conference on Artificial Intelligence and Statistics, Chia (Italy), May 2010.

[320] MATT HOFFMAN, FRANCIS BACH, AND DAVID BLEI. Online learning for latent dirichlet allocation. In Adv. Neural Info. Proc. Systems, 2010.

[321] RODOLPHE JENATTON, JULIEN MAIRAL, GUILLAUME OBOZINSKI, AND FRANCIS BACH. Proximal methods for sparse hierarchical dictionary learning. In Proceedings of the International Conference on Machine Learning (ICML), 2010.

[322] RODOLPHE JENATTON, GUILLAUME OBOZINSKI, AND FRANCIS BACH. Structured sparse principal component analysis. In International Conference on Artificial Intelligence and Statistics (AISTATS), 2010.

[323] ARMAND JOULIN, FRANCIS BACH, AND JEAN PONCE. Discriminative clustering for image co-segmentation. In Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), 2010.

[324] ARMAND JOULIN, FRANCIS BACH, AND JEAN PONCE. Efficient optimization for discriminative latent class models. In Advances in Neural Information Processing Systems (NIPS), 2010.

[325] BILIANA KANEVA, JOSEF SIVIC, A. TORRALBA, SHAI AVIDAN, AND WILLIAM T. FREEMAN. Matching and predicting street level images. In ECCV 2010 Workshop on Vision for Cognitive Tasks, 2010.

[326] ALEXANDER KLÄSER, MARCIN MARSZAŁEK, CORDELIA SCHMID, AND ANDREW ZISSERMAN. Human focused action localization in video. In Workshop on Sign, Gesture and Activity (in conjunction with ECCV), 2010.

[327] ALEXANDER KLÄSER, MARCIN MARSZAŁEK, CORDELIA SCHMID, AND ANDREW ZISSERMAN. Human focused action localization in video. In International Workshop on Sign, Gesture, Activity, 2010.

[328] JAN KNOPP, JOSEF SIVIC, AND THOMAS PAJDLA. Avoiding confusing features in place recognition. In Proceedings of the European Conference on Computer Vision, September 2010.

[329] I. LAPTEV, M. MARSZALEK, C. SCHMID, AND B. ROZENFELD. Learning realistic human actions from movies. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2008.

[330] JOSE LEZAMA, KARTEEK ALAHARI, IVAN LAPTEV, AND JOSEF SIVIC. Track to the future: Spatio-temporal video segmentation with long-range motion cues. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2011. to appear.

[331] JULIEN MAIRAL, RODOLPHE JENATTON, GUILLAUME OBOZINSKI, AND FRANCIS BACH. Network flow algorithms for structured sparsity. In Advances in Neural Information Processing Systems, 2010.

[332] M. MARSZALEK, I. LAPTEV, AND C. SCHMID. Actions in contexts. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2009.

[333] A. PATRON-PEREZ, MARCIN MARSZAŁEK, ANDREW ZISSERMAN, AND I. D. REID. High five: Recognising human interactions in TV shows. In British Machine Vision Conference, 2010.

[334] JAMES PHILBIN, MICHAEL ISARD, JOSEF SIVIC, AND ANDREW ZISSERMAN. Descriptor learning for efficient retrieval. In Proceedings of the European Conference on Computer Vision, September 2010.

[335] M. RODRIGUEZ, JEAN-YVES AUDIBERT, IVAN LAPTEV, AND JOSEF SIVIC. Data-driven crowd analysis in videos. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2011. to appear.

[336] J. SIVIC, M. EVERINGHAM, AND A. ZISSERMAN. "Who are you?": Learning person specific classifiers from video. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2009.

[337] A. TORII, THOMAS PAJDLA, AND JOSEF SIVIC. Read between the views: image-based localization by matching linear combinations of views. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2011. to appear.

[338] M. MUNEEB ULLAH, S. PARIZI, AND IVAN LAPTEV. Improving bag-of-features action recognition with non-local cues. In Proceedings of the British Machine Vision Conference, 2010.

[339] OLIVER WHYTE, JOSEF SIVIC, ANDREW ZISSERMAN, AND JEAN PONCE. Non-uniform deblurring for shaken images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2010.

[340] MIKHAIL ZASLAVSKIY, FRANCIS BACH, AND JEAN-PHILIPPE VERT. Many-to-many graph matching: a continuous relaxation approach. In Proc. Europ. Conf. on Machine Learning, 2010.

## THESES

[341] JEAN-YVES AUDIBERT. PAC-Bayesian aggregation and multi-armed bandits. Habilitation à Diriger des Recherches, Université Paris Est, 2010.

[342] JULIEN MAIRAL. Représentations parcimonieuses en apprentissage statistique, traitement d'image et vision par ordinateur. PhD thesis, Ecole Normale Supérieure de Cachan, 2010.

## TECH REPORTS

[343] JEAN-YVES AUDIBERT AND OLIVIER CATONI. Risk bounds in linear regression through PAC-bayesian truncation. Technical report, HAL, 2010. 78 pages.

[344] JEAN-YVES AUDIBERT AND OLIVIER CATONI. Robust linear least squares regression. Technical report, HAL, 2010. 48 pages.

[345] JEAN-YVES AUDIBERT AND OLIVIER CATONI. Robust linear regression through PAC-bayesian truncation. Technical report, HAL, 2010.

[346] FRANCIS BACH. Shaping level sets with submodular functions. Technical report, HAL, 2010.

[347] RODOLPHE JENATTON, JEAN-YVES AUDIBERT, AND FRANCIS BACH. Structured variable selection with sparsity-inducing norms. Research report, INRIA, 2010.

[348] RODOLPHE JENATTON, JULIEN MAIRAL, GUILLAUME OBOZINSKI, AND FRANCIS BACH. Proximal methods for hierarchical sparse coding. Research report, INRIA, 2010.

[349] ALEXANDER KLÄSER, MARCIN MARSZAŁEK, IVAN LAPTEV, AND CORDELIA SCHMID. Will person detection help bag-of-features action recognition? Research Report RR-7373, INRIA, September 2010.

[350] JULIEN MAIRAL, FRANCIS BACH, AND JEAN PONCE. Task-driven dictionary learning. Research Report RR-7400, INRIA, September 2010.

## OTHER

[351] SYLVAIN ARLOT. Sélection de modèles. Type : Conference digest, August 2010.

[352] FRANCIS BACH. Convex analysis and optimization with submodular functions: a tutorial. Tutorial.

[353] SÉBASTIEN BUBECK, JEAN-YVES AUDIBERT, AND RÉMI MUNOS. Bandit view on noisy optimization. In S. Sra, S. Nowozin, and S. J. Wright, editors, Optimization for Machine Learning. MIT Press, 2010.

## INVITED TALKS.

- A. Gaidon, Reunion du comite scientifique du LIMA, St-Etienne, July 2009
- A. Gaidon, MSR-INRIA Computer Vision and Machine Learning Workshop, Paris, January 2010
- A. Gaidon, VGG, University of Oxford, March 2010
- A. Gaidon, ENS-Lyon Computer Vision Seminar, Sept-Laux, January 2011
- C. Schmid, Discriminative metric learning in nearest neighbor models for image auto-annotation , Seminar at CMU, Pittsburgh, September 2009.
- C. Schmid, Large scale image search , International Workshop on Recent Trends in Computer Vision, Kyoto, Japan, June 2009.
- C. Schmid, Learning classes and context of human actions from movies , International Workshop on Video, Barcelona, Spain, May 2009.
- C. Schmid, Large scale image search , Keynote speaker at the Conference on Machine Vision Applications, Yokohama, Japan, May 2009.
- C. Schmid, Learning visual human actions from movies, Seminar at University of Texas at Austin, April 2009.
- C. Schmid, Burstiness for large scale image search, Seminar at Oxford University, March 2009.
- C. Schmid, Learning visual human actions from movies, Seminar at UCL, London, March 2009.
- C. Schmid, Large scale image search, Seminar at ETH, Zurich, February 2009.
- J. Ponce, Laboratoire d'Analyse et d'Architecture des Systèmes, 2009.
- J. Ponce, Télécom Paristech, 2009
- J. Ponce, Beckman Institute 20th Anniversary Symposium, 2009
- J. Ponce, ICCV Area Chair Symposium, 2009
- J. Ponce, University of Southern California, 2009
- J. Sivic, INRIA Rennes seminar, S. Malo, France, 2009
- J. Sivic, INRIA Grenoble seminar, Grenoble, France, 2009
- J. Sivic, International workshop on video, Barcelona, 2009
- J. Sivic, BIRS Workshop on Computer Vision and the Internet, Banff, Canada, 2009
- I. Laptev, Keynote at Sibgrapi 2009 Digital Video Journey, Rio de Janeiro, Brazil, October 2009.
- I. Laptev, Workshop on Trends in Computer Vision, Prague, July 2009.
- I. Laptev, International Workshop on Video, Barcelona, May 2009.
- A. Zisserman, Key note speaker at the International Workshop on Image Analysis, 2009. for Multimedia Interactive Services (WIAMIS), London, 2009.
- A. Zisserman, Invited presentation at BIRS Workshop on Computer Vision and the Internet, Banff, Canada, 2009.
- I. Laptev, GDR-ISIS scientific meeting, Paris, France, December 2010.
- I. Laptev, University Luxembourg, Data Mining Applications, Luxembourg, November 2010.
- I. Laptev, ECCV2010 Workshop on Sign, Gesture and Activity, Grece, September 2010.
- I. Laptev, ICPR2010 Workshop on Analysis and Evaluation of Large-Scale Multimedia Collections, Istanbul, Turkey, August 2010.
- I. Laptev, ICPR2010 Workshop on Human Behaviour Understanding, Istanbul, Turkey, August 2010.
- I. Laptev, Joint Oxford-KTH-INRIA workshop, Oxford, UK, July 2010.
- I. Laptev, International Workshop on Frontiers of Activity Recognition, Los Angeles, USA, June 2010.
- J. Ponce, Keynote speaker, British Machine Vision Conference, Aberystwyth, Wales, 2010.
- J. Ponce, Distinguished speaker, Dept. of Computer Science, University of Delaware, 2010.
- J. Ponce, Janelia Conference on What Can Computer Vision Do for Neuroscience and Vice Versa, VA, 2010.
- J. Ponce, Department of Computer Science, New York University, New York City, 2010.
- J. Ponce, Laboratoire d'informatique Gaspard-Monge, Paris, 2010.
- J. Ponce, Laboratoire Jacques-Louis Lions, Paris, 2010.
- C. Schmid, CVPR area chair meeting workshop, University of Maryland, February 2010.
- C. Schmid, seminar at CMLA, ENS Cachan, Paris, April 2010.
- C. Schmid, ECCV area chair colloquium, Paris, June 2010.
- C. Schmid, seminar at Oxford University, July 2010.
- C. Schmid, seminar at New York University, July 2010.
- C. Schmid, journee d'echanges et de formation, LERTI, INRIA, Grenoble, September 2010.
- C. Schmid, keynote speaker at Coresa 2010, Lyon, October 2010.
- J. Sivic, Joint Oxford-KTH-INRIA workshop, Oxford, UK, July 2010.
- J. Sivic, New York University, USA, September 2010.
- J. Sivic, DARPA/ARO Workshop on Interactive query refinement for image/video retrieval, September 2010, Columbia University, USA.
- A. Zisserman, Willow MSR-INRIA workshop, 2010.
- A. Zisserman, Keynote speaker at DAGM 2010.
- E. Memin, Seminar INSU Action Lefe assim at ENS 2010.

- E. Memin, Seminar MeteoFrance, Toulouse 2010.
- E. Memin, Seminar ENSPS Strasbourg 2010.
- E. Memin, Seminar LIMSI 2011.
- Popular science:
  J. Ponce et L. Benoit, intervention in Emission spéciale en public et en direct de l'ENS Ulm : "Cryptographie et vision artificielle", Place de la Toile, France Culture, Sep. 25, 2009.
  J. Ponce, "Comment donner un sens á l'image numérique?", Les cahiers de l'INRIA, La Recherche, Nov. 2009, No. 435.
  P. Heas, D. Heitz, E. Memin, "Dynamique des fluides : la turbulence par l'image" Les cahiers de l'INRIA, La Recherche, Sep. 2010, No. 444.

## EVENTS, WORKSHOPS, CONFERENCES, SEMINARS, EDITORIAL BOARDS

- C. Schmid, I. Laptev, J. Sivic : Co-organized the INRIA Visual Recognition and Machine Learning Summer School, Grenoble, July 2010. The school has attracted 137 participants from 26 countries.
- I. Laptev, J. Sivic, J. Ponce : Co-organized the MSR-INRIA workshop on Computer Vision and Machine Learning, Paris, January 2010.
- C. Schmid: Co-organizer of CVPR'09 Workshop on Feature Detectors and Descriptors, 2009.
- C. Schmid, A. Zisserman: Co-organizers of International Workshop on Video, Barcelona, Spain, 2009.
- C. Schmid: Program chair, European Conference on Computer Vision, 2012.
- C. Schmid: Area chair, IEEE Conference on Computer Vision and Pattern Recognition, 2010.
- C. Schmid: Area chair, European Conference on Computer Vision, 2010.
- C. Schmid: Area chair, IEEE International Conference on Computer Vision, 2009.
- C. Schmid: Area chair, RFIA 2010.
- A. Zisserman: Area chair, Neural Information and Processing Systems (NIPS) Conference, 2009.
- I. Laptev: Area chair, IEEE Conference on Computer Vision and Pattern Recognition, 2010.
- J. Ponce: Area chair, International Conference on Computer Vision, 2009.
- A. Zisserman, Co-organizer of the Pascal VOC 2009 Workshop at the IEEE International Conference on Computer Vision, Kyoto, Japan, 2009
- I. Laptev: Area chair, IEEE Conference on Computer Vision and Pattern Recognition, 2010.
- J. Sivic: Area chair, IEEE Conference on Computer Vision and Pattern Recognition, 2011.
- I. Laptev: Area chair, IEEE International Conference on Automatic Face and Gesture Recognition, 2011.
- I. Laptev, J. Sivic: Area chairs, IEEE International Conference on Computer Vision, 2011.
- I. Laptev, J. Ponce, C. Schmid, J. Sivic, A. Zisserman: Editorial board, International Journal of Computer Vision.
- C. Schmid, J. Ponce: Editorial board, Foundations and Trends in Computer Graphics and Vision, since 2005.
- J. Ponce: editorial board, SIAM Journal on Imaging Sciences.
- I. Laptev, E. Memin: editorial board, Image and Vision Computing Journal.
- J. Sivic, I. Laptev, L. Laborelli: Seminar with Frank R. Baumgartner (Professor of Political Science at UNC), 2010
- J. Sivic, I. Laptev, L. Laborelli: Meeting with Francois Jost (Professor at Sorbonne, Sciences de l'information et de la communication), 2010

## ACHIEVEMENTS

- The paper A tensor-based algorithm for high-order graph matching by O. Duchenne, F. Bach, I. Kweon, and J. Ponce, has been awarded a Best Student Paper award (Honorable Mention) at CVPR 2009.
- The paper Tracking closed curves with non-linear stochastic filters by C. Avenel, E. Memin and P. Perez received a best poster award at the conference Curves and Surfaces, 2010.
- A. Zisserman received the British Machine Vision Association Distinguished Fellow award, 2009.
- LEAR's submissions to the ImageCLEF evaluation campaign for the "Photo Annotation" and "Photo Retrieval" tracks obtained a second place among 19 participants for each track, see *http://imageclef.org/2009*
- LEAR: In the PASCAL visual object classes challenge 2010 LEAR/INRIA's work on human action recognition achieved best results on three out of nine action classes, see *http://pascallin.ecs.soton.ac.uk/challenges/VOC/voc2010/* for more details.
- LEAR: The best paper prize at the ECCV'10 International Workshop on Sign, Gesture, and Activity.
- LEAR: Submissions to the ImageCLEF evaluation campaign for the "Photo Annotation" track obtained the first place among 16 participants, *http://imageclef.org/2010*.
- LEAR: The best demonstration award At RFIA 2010 for the demonstration "10 million images on my laptop".
- I. Laptev, J. Sivic: INRIA Prime d'excellence scientifique.
- I. Laptev: Outstanding reviewer award at ECCV 2010.
- Best industrial paper prize for Patron-Perez, A., Marszalek, M., Reid, I. and Zisserman, A. High Five: Recognizing Human Interactions in TV Shows, British Machine Vision Conference (2010).
- J. Ponce and A. Zisserman were awarded Advanced ERC Grants.

**Track B** **This project started on winter 2010.**

# A-Brain

## OVERVIEW

> **In this project, we plan to explore cloud computing techniques to address the challenge of optimized storage for joint genetics and neuroimaging analysis.**

This project will bring together researchers from algorithmic and statistical analysis domain on the one hand, and researchers involved on the organization of data management in intensive computation on the other hand, to work on the Microsoft Windows Azure platform in order to unveil the relationships between genes and brain characteristics.

## TEAM

| | | | |
|---|---|---|---|
| **Team leader** | **THIRION** | **Bertrand** | ***INRIA Saclay-Île-de-France*** |
| **Team leader** | **ANTONIU** | **Gabriel** | ***INRIA Rennes Bretagne Atlantique*** |

## RESEARCH

Joint acquisition of neuroimaging and genetic data on large cohorts of subjects is a new approach used to assess and understand the variability that exists between individuals, and that has remained poorly understood so far. As both neuroimaging- and genetic-domain observations represent a huge amount of variables (of the order of 106), performing statistically rigorous analyses on such amounts of data represents a computational challenge that cannot be addressed with conventional computational techniques. On one hand, sophisticated regression techniques need to be used in order to perform sensitive analysis on these large datasets; on the other hand, the cost entailed by parameter optimization and statistical validation procedures (e.g. permutation tests). However, the computational framework can easily run in parallel.

**Bertrand Thirion** graduated from École Polytechnique in 1998. He specialised in applied mathematics, with applications to computer vision. He did his PhD in 2000-2003 with the Odyssée team (Sophia-Antipolis, France) on the statistical analysis of functional brain images, under the direction of Olivier Faugeras. His main research interests are the modeling of brain variability in group studies, the mathematical study of functional connectivity and the use of machine learning tools for brain activity analysis. He is part of the scientific board of the French neuroscience institute. He is the principal investigator of the Parietal team (INRIA Saclay-Île-de-France) situated within the Neurospin research center (CEA, DSV, I2BM) at Saclay, France.

**Gabriel Antoniu** received his Ph.D. degree in Computer Science in 2001 from ENS Lyon and his Habilitation for Research Supervision (HDR) from ENS Cachan in 2009. His research interests include: grid and cloud storage, large-scale distributed data management and sharing, data consistency models and protocols, grid and peer-to-peer systems. In the area of distributed data storage, he leads the MapReduce ANR project (2010- 2013) in partnership with Argonne National Nab (USA), the University of Illinois at Urbana Champaign (USA), and the INRIA-UIUC Joint Lab for Petascale Computing and IBM France. He is a Research Scientist at INRIA and leader of the KerData research team at INRIA Rennes-Bretagne Atlantique.

In this project, researchers of the Parietal and KerData INRIA teams will jointly address this computational problem using cloud computing techniques on Microsoft Windows Azure cloud computing environment. The two teams bring their complementary expertise: KERDATA (Rennes) in the area of scalable cloud data management and PARIETAL (Saclay) in the field of neuroimaging and genetics data analysis. The Map-Reduce programming model has recently arisen as a very effective approach to develop high-performance applications over very large distributed systems such as grids and now clouds. KerData has recently proposed a set of algorithms for data management, combining versioning with decentralized metadata management to support scalable, efficient, fine-grain access to massive, distributed Binary Large OBjects (BLOBs) under heavy concurrency. The project investigates the benefits of integrating BlobSeer with Microsoft Windows Azure storage services and aims to evaluate the impact of using BlobSeer on Azure with large-scale application experiments such as the genetics-neuroimaging data comparisons addressed by Parietal.

## CLOUD COMPUTING RESOURCES

The A-Brain project was launched as part of a new Microsoft initiative from the eXtreme Computing Group at Microsoft Research Redmond called "Cloud Research Engagement Initiative". The goal is to give European scientists a common set of tools, applications, and data sets they can share with the scientific community.

Under a set of new agreements with major research institutions across the world, Microsoft will provide researchers with access to its Windows Azure cloud services and will also will help researchers integrate cloud technology into their work by giving them access to a team of Microsoft cloud specialists.

In this context, Microsoft expands its partnership with INRIA and will provide the Microsoft Research-INRIA Joint Centre three years of free Windows Azure usage, which delivers on-demand computing and storage to host, scale, and manage web applications on the Internet through Microsoft data centers. Researchers will also be provided expertise in research, science, and cloud computing through a collaboration with the European Microsoft Innovation Centre (EMIC) based in Aachen. ■

### EVENTS, WORKSHOPS, CONFERENCES, SEMINARS

We held a kickoff meeting at INRIA on January 19, 2011 with participation of Pierre Couzy Microsoft France, Goetz Brasche EMIC Aachen and Fabrizio Gagliardi Microsoft Research Cambridge.

The two next publications are good references for the work pursued in A-Brain.

# PUBLICATIONS & TALKS

### JOURNAL PAPERS AND BOOK CHAPTERS

[354] BOGDAN NICOLAE, GABRIEL ANTONIU, LUC BOUGÉ, DIANA MOISE, AND ALEXANDRA CARPEN-AMARIE. BlobSeer: Next Generation Data Management for Large Scale Infrastructures. Journal of Parallel and Distributed Computing, 71(2):168–184, February 2011.

### CONFERENCE AND WORKSHOP PAPERS

[355] JEAN-BAPTISTE POLINE, CHRISTOPHE LALANNE, ARTHUR TENENHAUS, EDOUARD DUCHESNAY, BERTRAND THIRION, AND VINCENT FROUIN. Imaging genetics: bio-informatics and bio-statistics challenges. In Yves Lechevallier and Gilbert Saporta, editors, 19th International Conference on Computational Statistics Proceedings of COMPSTAT'2010, Paris France, 08 2010.

# ANNEX - RESEARCH STAFF

## Figure 1: List of PHD students part-time or full-time at the Joint Centre.

| PROJECT | LAST NAME | FIRST NAME | COUNTRY | CURRENT AFFILIATION | NEXT AFFILIATION |
|---|---|---|---|---|---|
| Adaptative Search | ARBALAEZ | Alexandro | Colombia | Université Paris 12 | *N/A* |
| | FIALHO | Alvaro | Brazil | Université Paris 11 | *N/A* |
| Dynamic Dictionary of Mathematical Functions | BENOIT | Alexandre | France | Ecole Polytechnique | *N/A* |
| | MEZZAROBBA | Marc | France | Ecole Polytechnique | *N/A* |
| | CHEN | Shaoshi | China | École Polytechnique & Chinese Academy of Sciences | *N/A* |
| Mathematical Components | CANO | Guillaume | France | Université Nice Sophia Antipolis | *N/A* |
| | GARILLOT | Francois | France | Ecole Normale Supérieure de Paris | *N/A* |
| | PASCA | Iona | Romania | Université Nice Sophia Antipolis | *N/A* |
| | OULD BIHA | Sidi | Tunisia | Université Nice Sophia Antipolis | *N/A* |
| | SPIVAK | Arnaud | France | Ecole Polytechnique | *N/A* |
| | TASSI | Enrico | Italy | University of Bologna | *Researcher, University of Bologna* |
| | ZUMKELLER | Roland | Germany | Ecole Polytechnique | *Post doc at University of Pittsburg (w/ Thomas Hales)* |
| ReActivity | MASSON | Nicolas | France | Université Paris 11 | *N/A* |
| | TABARD | Aurelien | France | Université Paris 11 | *N/A* |
| | HENRY | Nathalie | France | Université Paris 11 | *Full time researcher at Microsoft Research Redmond (VIBE Group)* |
| Secure Distributed Computations and their Proofs | CADE | David | France | Ecole Normale Supérieure de Paris | *N/A* |
| | DENIELOU | Pierre-Malo | France | Ecole Normale Supérieure de Cachan | *N/A* |
| | GUTS | Nataliya | Ukrain | Université Paris 6 | *University of Maryland* |
| | PAIOLO | Miriam | | Ecole Normale Supérieure de Paris | |
| | PLANUL | Jérémy | France | Ecole Normale Supérieure de Lyon | *N/A* |
| | ZANELLA | Santiago | Argentina | Université Nice Sophia Antipolis | *Postdoc, IMDEA Software, Madrid* |
| | ZHENGQIN | Luo | China | Université Nice Sophia Antipolis | |
| TLA+ | VANZETTO | Hernan Pablo | Italy | Université Nancy | |
| | ZAMBROVSKY | Simon | Germany | Hamburg University | *N/A* |
| Image and Video Mining for Science and Humanities | GAIDON | Adrien | France | INP Grenoble | *N/A* |
| | HARCHAOUI | Warith | France | Supelec | |
| | WHYTE | Oliver | Grear Britain | University of Oxford | *N/A* |

Candidate  Defended

## Figure 2: List of Post Doc students part-time or full-time at the Joint Centre

| PROJECT | LAST NAME | FIRST NAME | COUNTRY | CURRENT AFFILIATION | NEXT AFFILIATION |
|---|---|---|---|---|---|
| Adaptative Search | HANSEN | Niklaus | Germany | Université Paris 11 | *N/A* |
| | JABBOUR | Saïd | Morocco | Université d'Artois | *Université d'Artois* |
| Dynamic Dictionary of Mathematical Functions | DARRASSE | Alexis | Greece | Université de Paris 6 | *LIP6 (UPMC)* |
| | GERHOLD | Stefan | Austria | Vienna University of Technology | *N/A* |
| | KOUTSCHAN | Christoph | Austria | Tulane University (New Orleans, USA) Université de Linz | |
| | STAN | Flavia | Romania | Tulane University (New Orleans, USA) Université de Linz | |
| Mathematical Components | LEROUX | Stéphane | France | Ecole Normale supérieure de Lyon | *Ecole Polytechnique* |
| | MELQUIOND | Guillaume | France | Ecole Normale Supérieure de Lyon | *Full time researcher at INRIA* |
| | MAHBOUBI | Assia | France | Université Nice-Sophia Antipolis | *Full time researcher at INRIA* |
| | O CONNOR | Russel | Canada | McMaster University | |
| | TASSI | Enrico | Italia | University of Bologna | |
| ReActivity | BOUKHELIFA | Nadia | France | University of Leeds (UK), School of Computing. | |
| | CHEVALIER | Fanny | France | Université Paris 11 | *OCAD, Totonto* |
| | ELMQVIST | Niklas | Sweden | Chalmers University, Göteborg | *Assistant professor at Purdue University* |
| | LETONDAL | Catherine | France | Institut Pasteur (on sabbatical leave) | *Institut Pasteur (on sabbatical leave)* |
| | LI | Xiujun | China | Nanyang Technological University, Singapore | *N/A* |
| | LICCARDI | Ilaria | Italia | University of Southampton (England) | *Post doc INRIA Saclay Ile de France* |
| | MOSCOVICH | Tomer | USA | University of Toronto | *LAB126* |
| | TSANDILAS | Theophanis | Greece | University of Toronto | *N/A* |
| Secure Distributed Computations and their Proofs | CORIN | Ricardo | Argentina | University of Twente | *Universidad Nacional de Cordoba, Argentina.* |
| | LE GUERNIC | Gurvan | France | Université de Rennes/Kansas University | *KTH Stockholm* |
| | PIRONTI | Alfredo | Italia | Politecnico di Torino | |
| | REZK | Tamara | Argentina | INRIA Sophia Antipolis | *Full time researcher at INRIA* |
| | STRUB | Pierre-Yves | France | Ecole Doctorale Polytechnique | |
| | ZALINESCU | Eugen | Romania | Université de Nancy | *ETH, Zurich* |
| TLA+ | CHAUDHURI | Kaustuv | India | CMU University | *N/A* |
| | COUSINEAU | Denis | France | Ecole Polytechnique | |
| | KUPPE | Markus | Allemagne | Universität Hamburg | |
| | RICKETTS | Daniel | USA | Brown University | *N/A* |
| | TRISTAN | Jean-Baptiste | France | Université Paris 7 | *N/A* |
| Image and Video Mining for Science and Humanities | CHERMIAVSKY | Neva | USA | University of Washington-Seattle | *N/A* |
| | GORTHI | Rama Krishna | India | Indian Institute of Technology (IIT), Madras | *N/A* |
| | RUSSELL | Bryan | USA | Institut technologique de Cambridge - Massachusset USA | *N/A* |

Current Left

## Figure 3: List of INRIA Researchers contributing to the Joint Centre

| PROJECT | STATUS | LAST NAME | FIRST NAME | AFFILIATION |
|---|---|---|---|---|
| Adaptative Search | **Team leader** | SCHOENAUER | Marc | INRIA Saclay-Île-de-France |
| | **Researcher** | AUGER | Anne | INRIA Saclay-Île-de-France |
| Dynamic Dictionary of Mathematical Functions | **Team leader** | SALVY | Bruno | INRIA Paris-Rocquencourt |
| | **Researcher** | BOSTAN | Alin | INRIA Paris-Rocquencourt |
| | **Researcher** | CHYZAC | Frédéric | INRIA Paris-Rocquencourt |
| Management | **Director** | HUET | Gérard | INRIA Paris-Rocquencourt |
| | **Director** | LEVY | Jean-Jacques | INRIA Paris-Rocquencourt |
| | **System Engineer** | ROUSSE | Guillaume | INRIA Saclay-Île-de-France |
| Mathematical Components | **Researcher** | BARRAS | Bruno | INRIA Saclay-Île-de-France |
| | **Researcher** | BERTOT | Yves | INRIA Sophia Antipolis-Méditerranée |
| | **Researcher** | MAHBOUBI | Assia | INRIA Saclay-Île-de-France |
| | **Researcher** | RIDEAU | Laurence | INRIA Sophia Antipolis-Méditerranée |
| | **Researcher** | THERY | Laurent | INRIA Sophia Antipolis-Méditerranée |
| | **Researcher** | WERNER | Benjamin | INRIA Saclay-Île-de-France |
| ReActivity | **Team leader** | FEKETE | Jean-Daniel | INRIA Saclay-Île-de-France |
| | **Team leader** | MACKAY | Wendy | INRIA Saclay-Île-de-France |
| | **Researcher** | DRAGICEVIC | Pierre | INRIA Saclay-Île-de-France |
| | **Researcher** | PIETRIGA | Emmanuel | INRIA Saclay-Île-de-France |
| Secure Distributed Computations and their Proofs | **Researcher** | BARTHE | Gilles | INRIA Sophia Antipolis-Méditerranée |
| | **Researcher** | BHARGAVAN | Karthik | INRIA Paris-Rocquencourt |
| | **Researcher** | CORIN | Ricardo | INRIA Paris-Rocquencourt |
| | **Researcher** | GREGOIRE | Benjamin | INRIA Sophia Antipolis-Méditerranée |
| | **Researcher** | LEIFER | James | INRIA Paris-Rocquencourt |
| | **Researcher** | REZK | Tamara | INRIA Sophia Antipolis-Méditerranée |
| | **Researcher** | ZAPPA NARDELLI | Francesco | INRIA Paris-Rocquencourt |
| TLA+ | **Team leader** | DOLIGEZ | Damien | INRIA Paris-Rocquencourt |
| | **Researcher** | MERZ | Stephan | INRIA Lorraine |
| Image and Video Mining for Science and Humanities | **Team leader** | PONCE | Jean | INRIA Paris-Rocquencourt |
| | **Researcher** | HARCHAOUI | Zaid | INRIA Grenoble-Rhône-Alpes |
| | **Researcher** | JEGOU | Hervé | INRIA Grenoble-Rhône-Alpes |
| | **Researcher** | LAPTEV | Ivan | INRIA Rennes Bretagne Atlantique |
| | **Researcher** | MEMIN | Etienne | INRIA Rennes Bretagne Atlantique |
| | **Researcher** | PEREZ | Patrick | INRIA Rennes Bretagne Atlantique |
| | **Researcher** | SCHMID | Cordelia | INRIA Grenoble-Rhône-Alpes |
| A-Brain | **Team leader** | THIRION | Bertrand | INRIA Saclay-Île-de-France |
| | **Team leader** | ANTONIU | Gabriel | INRIA Rennes Bretagne Atlantique |

## Figure 4: Researchers from other public institutions

| PROJECT | STATUS | LAST NAME | FIRST NAME | AFFILIATION |
|---|---|---|---|---|
| Adaptative Search | **Researcher** | SEBAG | Michèle | CNRS |
| ReActivity | **Researcher** | BEAUDOUIN-LAFON | Michel | Université Paris 11 |
| | **Researcher** | CHAPUIS | Olivier | CNRS |
| Scientific Images and Video Mining | **Team leader** | PONCE | Jean | Ecole Normale Supérieure de Paris |
| | **Researcher** | DESSALES | Hélène | Ecole Normale Supérieure de Paris |
| | **Researcher** | LABORELLI | Louis | Institut National de l'Audiovisuel (INA) |
| Secure Distributed Computations and their Proofs | **Researcher** | BLANCHET | Bruno | Ecole Normale Supérieure de Paris |

## Figure 5: List of Microsoft Researchers contributing to the Joint Centre

| PROJECT | STATUS | LAST NAME | FIRST NAME | AFFILIATION |
|---|---|---|---|---|
| Adaptative Search | **Team leader** | HAMADI | Youssef | Microsoft Research Cambridge |
| | **Research Software** | BORDEAUX | Lucas | Microsoft Research Cambridge |
| Management | **Deputy Director** | XECH | Pierre-Louis | Microsoft France |
| Mathematical Components | **Team leader** | GONTHIER | Georges | Microsoft Research Cambridge |
| ReActivity | **Researcher** | BONGSHIN | Lee | Microsoft Research Redmond |
| | **Researcher** | CZERWINSKI | Mary | Microsoft Research Redmond |
| | **Researcher** | FISHER | Danyel | Microsoft Research Redmond |
| | **Researcher** | HENRY-RICHE | Nathalie | Microsoft Research Redmond |
| | **Researcher** | MEYERS | Brian | Microsoft Research Redmond |
| | **Researcher** | SMITH | Greg | Microsoft Research Redmond |
| | **Researcher** | ROBERTSON | Georges | Microsoft Research Redmond |
| Secure Distributed Computations and their Proofs | **Team leader** | FOURNET | Cédric | Microsoft Research Cambridge |
| | **Principal Researcher** | GORDON | Andrew | Microsoft Research Cambridge |
| | **Researcher** | BHARGAVAN | Kartik | Microsoft Research Cambridge |
| TLA+ | **Researcher** | LAMPORT | Leslie | Microsoft Research Silicon Valley |
| Image and Video Mining for Science and Humanities | **Researcher** | BLAKE | Andrew | Microsoft Research Cambridge |
| | **Researcher** | SZELISKI | Rick | Microsoft Research Redmond |

## Figure 6: List of Visitors

| PROJECT | LAST NAME | FIRST NAME | AFFILIATION | YEAR |
|---|---|---|---|---|
| Mathematical Components | ASPERTI | Andrea | University of Bologna, Italy | 2008 |
| | AVIGAD | Jéremy | CMU University, USA | 2009 |
| Secure Distributed Computations and their Proofs | LANEVE | Cosimo | University of Bologna, Italy | Regular visitor |
| | ADAO | Pedro | Instituto de Telecomunicações Lisboa, Portugal | Regular visitor |
| | GUNTER | Carl | University of Illinois. | 2010 |
| ReActivity | KLEMMER | Scott | Stanford University, USA | 2008 |
| | FRY | Jeremy | University of Southampton, UK | 2008 |
| | SCHRAEFEL | M.C | University of Southampton, UK | 2008 |
| | ANDRE | Paul | University of Southampton, UK | 2008 |
| | WILSON | Max | University of Southampton, UK | 2008 |
| | KARGER | David | MIT, USA | 2008 |
| | VAN KLEEK | Max | MIT, USA | 2008 |
| | HOLLAN | James | University California San Diego, USA | 2008 |
| | CANGIANO | Gaston | University California San Diego, USA | 2008 |
| | FOUSE | Adam | University California San Diego, USA | 2008 |
| | MALONEY | Chris | Brown University, USA | 2008 |
| | ISENBERG | Petra | University of Calgary, Canada | 2008 |
| | BEZERIANOS | Anastasia | NICTA, Australia | 2008 |
| | CARPENDALE | Sheelagh | University of Calgary, Canada | 2010 |
| Image and Video Mining for Science and Humanities | DURAND | Fredo | MIT, USA | 2009-2010 |
| | NEVATIA | Ram | University of Southern California, USA | 2009-2010 |
| | PAJDA | Tomas | Czech Technical University in Prague | 2009-2010 |
| | EFROS | Alexei | Carnegie Mellon University, USA | 2010 |

MICROSOFT RESEARCH
INRIA
JOINT
CENTRE

**The Microsoft Research-INRIA Joint Centre is located at**
**Parc Orsay Université, 28, rue Jean Rostand, 91893 Orsay Cedex, France.**

The Joint Centre is headed by Jean-Jacques Lévy from INRIA

**Governance is supervised by a management committee whose members are for INRIA:**
Nozha Boujemaa, *Director of INRIA Saclay-Ile de France*,
Claude Kirchner, *CEO for Science and Technology*,
Pascal Guitton, *Director of Research*,
Bruno Sportisse, *Director of Technology Transfer and Innovation*.

**For Microsoft, they are:**
Andrew Blake, *Managing Director of Microsoft Research Cambridge*,
Fabrizio Galgliardi, *External Research Director EMEA of Microsoft Research* ,
Bernard Ourghanlian, *CTO of Microsoft France*.